UNIVERSITÀ DEGLI STUDI DI PARMA

Dipartimento di Fisica e Scienze della Terra

Dottorato di Ricerca in Fisica

XXIX ciclo

STRUCTURAL AND MOLECULAR DETERMINANTS AFFECTING THE INTERACTION OF RETINOL WITH HUMAN CRBPs

Coordinatore:
Chiar.mo Prof. CRISTIANO VIAPPIANI

Tutor:
Chiar.ma Prof.ssa EUGENIA POLVERINI

Dottoranda:
Dott.ssa ILARIA MENOZZI

2014-2016

"Se sapessimo esattamente quel che stiamo facendo, non si chiamerebbe ricerca."

Albert Einstein

Index:

INTRODUCTION

1.1 Vitamin A and its derivatives

Vitamin A is a term reserved to describe every compound possessing the biological activity of retinol. The term "retinoids" was designated by IUPAC-IUB in 1982 to include compounds consisting of four isoprenoid units joined in a head-to-tail manner; all retinoids may be formally derived from a monocyclic parent compound containing five carbon-carbon double bonds and a functional terminal group at the end of the acylic portion (Rune Blomhoff and Blomhoff 2006).

The parent retinoid, all-*trans* retinol, is a primary alcohol with a molecular weight of 286 Da. In most animal tissues, the predominant retinoid is retinyl palmitate, but other fatty acid esters, such as retinyl oleate and stearate are also found. Most of these derivatives are present in the all-*trans* conformation but the 11-cis retinal (the aldehyde form) occurs in the retina of the eyes and several acidic forms, such as *trans* and 13-*cis* retinoic acid, are metabolites of retinol found in many tissues (Joseph L. Napoli 1999).

All-trans retinol and its derivatives are highly unstable in the presence of oxidants and light, which lead to their oxidative degradation or isomerization. No animal species have the capability for *de novo* vitamin A synthesis, but plants and microorganisms can synthetize carotenoids. These latter compounds constitute a large group of pigments common in nature; many of these can be absorbed, stored and cleaved to form biologically active molecules, in animals. Alternatively, vitamin A can be absorbed from diet by eating tissues from animals that already have converted the provitamin A carotenoids into retinoids (generally in the form of retinyl esters) (Joseph L. Napoli 1999).

1.2 Carotenoids and Retinoid absorptions

The uptake of carotenoids by enterocytes, in the small intestine, occurs by passive diffusion. Thus, in this tissue, the β-carotene is converted to retinoids by β,β-carotene-15,15'-monooxygenase, an enzyme that makes a central cleavage at the 15-15' carbon double bond of the β-carotene arising two molecules of retinal (Fig. 1.1) (Moore 1930) (Olson and Hayaishi 1965). An additional mechanism of carotene cleavage has also been identified: an asymmetric cleavage made by β,β carotene 9',10'-dioxygenase (Kiefer et al. 2001) leads to the formation of two molecules of β-apocarotenal with different chain length. The longest chain can subsequently be enzymatically

shortened, causing the formation of retinoic acid or retinal (Fig. 1.1) (Glover and Redfearn 1954) (J L Napoli and Race 1988). Retinal formed in the intestine is subsequently reduced to retinol by retinal reductase (Li and Tso 2003).

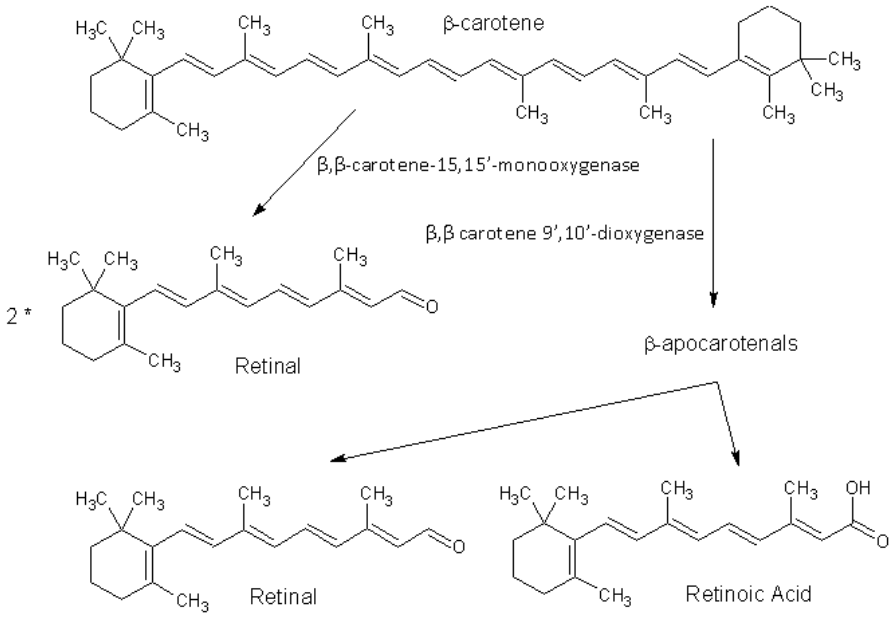


Figure 1.1 Scheme of β-carotene cleavage.

Dietary retinyl esters are enzymatically converted to retinol in the intestinal lumen prior to uptake in the enterocytes (Fig. 1.2). The un-esterified retinol is taken up by enterocytes and bound to a specific binding protein, the cellular retinol-binding protein II (CRBPII) (Crow and Ong 1985). The major part of the retinol is re-esterified after absorption in the enterocyte adding long chain fatty acid. Retinol bound to CRBP II is easily esterified by lecitine:retinol acyl transferase (LRAT) (Jamison, Newcomer, and Ong 1994); the major part of formed retinil ester is then incorporated into chylomicrons (R Blomhoff et al. 1982). These are aggregates of triacylglycerol and phospholipid molecules packed together with carotenoids, retinyl esters, retinol, and few species of apolipoproteins that are secreted from the enterocytes into the intestinal lymph and are moved into the general circulation to reach the parenchymal cells of the liver (Fig. 1.2).

In hepatocytes, the retinyl esters are hydrolyzed, and a large portion of free retinol is transferred to stellate cells of the liver, for storage. Other unesterified retinol may associate with the plasma retinol binding protein (RBP) and this complex is translocated from the endoplasmic reticulum to the Golgi apparatus and secreted into the plasma (Ronne et al. 1983). RBP seems to be essential for the cellular uptake of retinol in the retina, especially when the availability of dietary Vitamin A

is marginal. Some studies indicate that, in most of the other tissues, RBP is not essential for the uptake of plasma retinol, even if the mechanism for this non-RBP mediated cellular uptake of retinol has not been well characterized (Rune Blomhoff and Blomhoff 2006).

A number of other retinoids are present in the plasma at nanomolar concentrations (about 5-10nM) and are carried by albumin, while carotenoids are soluble fats and can be transported in plasma by lipoprotein and then transformed inside the target tissues (Wyss, Bucheli, and Hartenbach 1998).

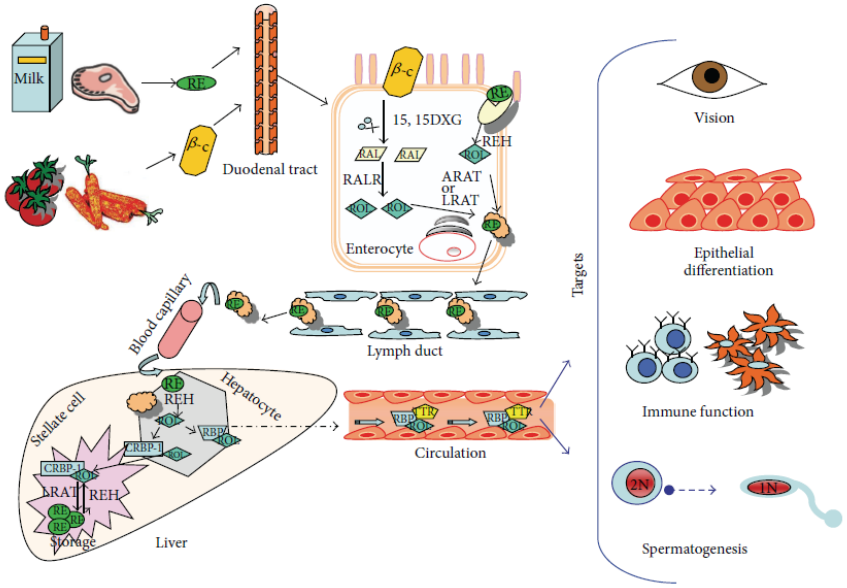
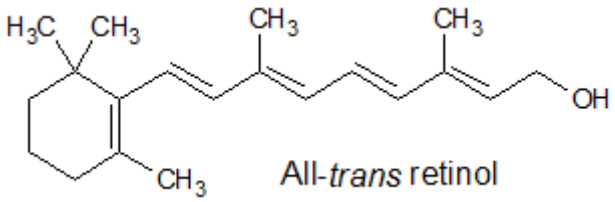


Figure 1.2 Absorption, transport and distribution of dietary retinoids. (RAL = retinaldheyde; ROL = retinol; RE = retinyl esters) (Doldo et al. 2015).

1.3 Retinoid functions

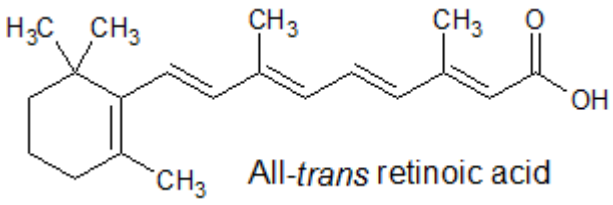
1.3.1 Retinol



It is generally accepted that active retinoid metabolites are synthetized into the target cells. The major source for this synthesis is the all-*trans* retinol taken up from plasma (in higher vertebrates),

but cellular uptake of lipoproteins containing other retinoids (such as retinyl esters, caroteinoids, all-*trans* retinoic acid) or retinoids stored in lipid droplets in the target cells, contribute to the synthesis of active molecules.

1.3.2 Retinoic acid



All-*trans* retinoic acid

All-*trans* retinoic acid is a major active cellular retinoid metabolite, one of the first morphogens identified. It is a low molecular weight (300 Da) lipophilic molecule, but is also partly soluble in extra- and intracellular fluids. Retinoic acid is a rapidly diffusing signaling molecule that can specify cell identities and controls gene expression through activation of specific nuclear receptors (Retinoic Acid Receptors, RARs). All-*trans* retinoic acid also modulates the activity of protein Kinase C (PKC), a protein that regulates fundamental cellular functions including proliferation, differentiation, tumorigenesis and apoptosis (Joseph L. Napoli 1999).

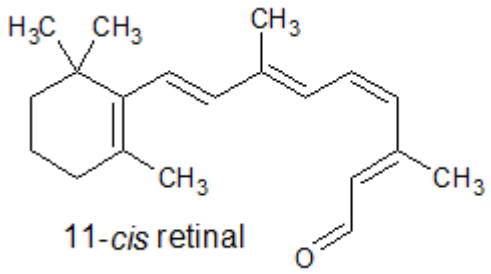
Retinoic acid synthesis from all-trans retinol occurs in a two-steps reaction: the oxidation of retinol to retinal and the oxidation of retinal to retinoic acid.

The first reaction is catalyzed by two different enzymes: one cytosolic and one membrane-bound type. The Alcohol Dehydrogenases (ADH 1,2,4) are involved in oxidation of the free all-*trans* retinol to all-*trans* retinal, but do not seem to convert retinol bound to CRBP (Joseph L. Napoli 1999). Oxidation of retinol to retinal appears also to be catalyzed by members of the Short-Chain Dehydrogenase/Reductase (SDR) family of microsomal enzymes. The SDRs utilize the retinol bound to CRBP as substrate (Everts, Sundberg, and Ong 2005).

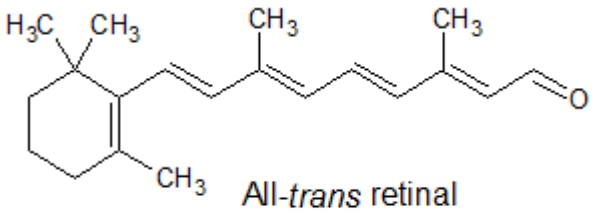
The conversion of retinal to retinoic acid is accomplished by a class of enzymes named retinal dehydrogenase (RALDH also called ALDH) and differently expressed in organs. One of these, RALDH4, largely expressed in mouse liver and kidney, has preference for 9-cis-retinal instead of the all-*trans* form and it was suggested that it is involved in the biosynthesis of 9-cis-retinoic acid,

which is able to bind retinoid receptors (RXR) activating the synthesis of a number RXR controlled genes (Joseph L. Napoli 1999).

1.3.3 Retinal



11-*cis* retinal



All-*trans* retinal

Photosensitivity is based on the isomerization of retinal. This pigment is responsible for absorption of light and consists of a retinal prosthetic group covalently liked to proteins called opsins. The reaction that follows to the light absorption always involved *cis* to *trans* or *trans* to *cis* isomerization.

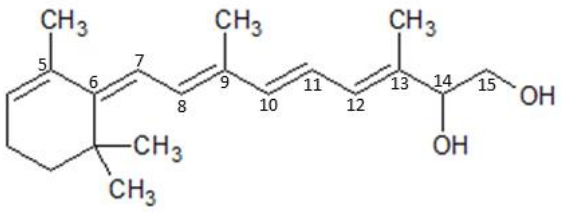
Vertebrate phototransduction is initiated by a photochemical reaction where 11-*cis*-retinal, bound to opsin, is isomerized to all-*trans* retinal, producing changes in the opsin conformation. Rhodopsin is the best characterized vertebrate visual pigment, in which one molecule of 11-cis retinal is attached to Lys53 by a Shiff base linkage. When exposed to light, 11-*cis* retinal isomerizes to all-*trans* retinal; the different geometry of retinal induces a conformational change in rhodopsin, which leads to a sequence of events which in turn generate a nervous impulse (Palczewski et al. 1994).

1.3.4 Dihydroxy-retinol

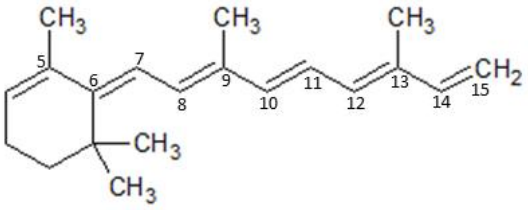
Another intracellular retinol derivative was found in numerous cells line: the 13,14-dihydroxy-retinol (DHR). DHR is converted to 14-hydroxy-retro-retinol (14HRR) by mild acid treatment, but not by cells; therefore DHR is not a biosynthetic intermediate in the conversion of retinol to

14HRR. Although it is linked to cell proliferation, its biological role remains to be determined (F Derguini et al. 1995).

1.3.5 Retro-retinoids



all-*trans*-14-hydroxy-4,14-*retro*-retinol (14HRR)



all-trans-anhydroretinol (AR)

Most retinoids contain five double C-C bonds located between atoms 5-6, 7-8, 9-10, 11-12 and 13-14. A shift in the "standard" position of these double bonds generates biologically active retinoids called retro-retinoids.

Whereas retinoic acids control nuclear events, a second class of retinol metabolites, exemplified by 14-hydroxy-retro-retinol (14HRR), operates primarily in the cytoplasm. They act as regulatory cofactors for cell survival/cell death decisions. In accordance with these biological aspects, it was demonstrated that these retinoids bind protein kinase C (PKC) alpha with nanomolar affinity and markedly enhance the activation of PKC alpha and the entire downstream MAP kinase pathway by reactive oxygen species (Imam et al. 2001); the activation of transcription factors, operated by the MAP kinase pathway, promotes the transcription of specific genes.

Every cell line is able to synthetized these retinoids, in particular lymphocyte B, that require 14HRR to mediate growth control (Buck et al. 1991). This compound is the first described bioactive retro-retinoid and was found to be the intracellular mediator of cell proliferation. It sustains the growth of B cells and the activation of T cells at 10-30 fold lower concentration than retinol. The anhydroretinol (AR) is the second retro-retinoid identified in human B cells. AR functions as a reversible inhibitor of retinol and probably binds the same target protein as 14HRR. AR and 14HRR constitute the first naturally occurring agonist/antagonist pair of retinoids to be described (F. Derguini et al. 1994).

Other retro-retinoids were identified but their role remain, at date, unknown: all-trans-4,14-retro-retinol, all-trans-14-hydroxyl-4,14-retro-deoxyretinol, all-trans-4-hydroxy-anhydroretinol.

1.4 Retinoid-binding protein

The lability of Vitamin A, due to its hydrophobic nature and to the presence of a number of double bonds, is prevented by the presence of specific carrier proteins which act as chaperons and participate to metabolism regulation, signaling and transport of retinoids. The *stimulated by* retinoic acid 6 receptors (STRA6) mediate vitamin A uptake from the plasma retinol-binding protein (RBP) to the target cells as shown in figure 1.3. Within the cell, retinol and its derivatives are solubilized by intracellular retinoid-binding proteins: cellular retinol-binding protein (CRBP) is selective for retinol, cellular retinoic-acid binding protein (CRABP) for the retinoic acid transport and cellular retinal-binding protein (CRALBP) is necessary for 11-cis-retinal carriage.

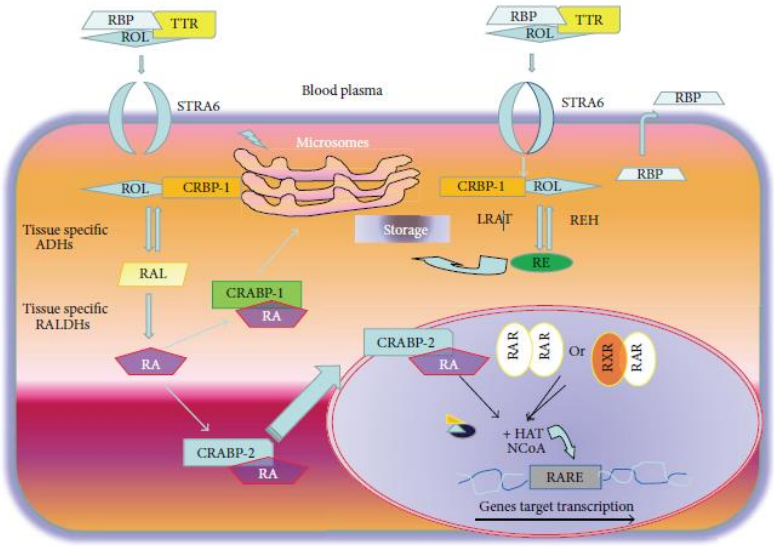


Figure 1.3 Intracellular retinoid pathway. (Doldo et al. 2015)

1.4.1 Retinol-binding protein

The transport of retinol in the bloodstream is guaranteed by the presence of the retinol-binding protein, expressed primarily in the liver. This 21 kDa protein acts as retinol carrier, preserving it to degradation in the plasma environment. Holo-RBP is present in the blood of higher vertebrates, bound, for its major part, to another plasma protein: the transthyretin (TTR) (Episkopou et al.

1993). This 1:2 (TTR:RBP) molar complex increases the stability of retinol-RBP complex and preserves the elimination of the relatively small RBP through kidney glomeruli (Zanotti et al. 2008).

RBP belongs to lipocalin superfamily and shares high structural conservation with the other members of the family. The name lipocalins derives from the fact that their structures resemble a calyx and they bind lipophilic compounds. In figure 1.4 is shown the RBP 3D structure; this monomeric protein is made by eight antiparallel β-strands (A-H) arranged in a barrel and an alpha helix that flaks the barrel; a short 3_{10} helix is present at the C terminus. The cavity formed by the β-barrel can host one molecule of retinol deeply inserted with its cyclohexene ring in the bottom of the pocket and the polyene chain completely buried. The hydroxyl group of the retinol is nearly at the protein surface, as is shown in the right picture of figure 1.4 (Newcomer and Ong 2000).

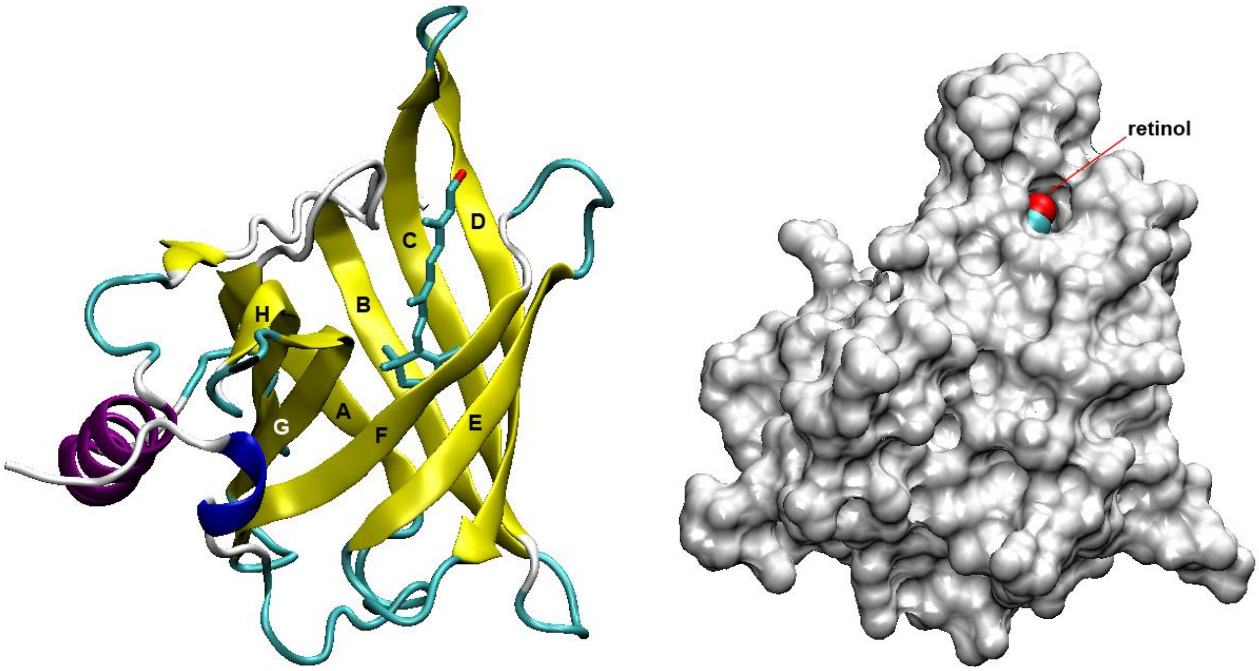


Figure 1.4 The picture on the left shows the cartoon representation of hRBP (PDB ID 1RBP), colored by secondary structure: eight β-stands (A-H) and one α-helix define the typical lipocalin structure. On the right the molecular surface of hRBP is shown to indicate that the hydrophobic part of retinol is shielded from the surrounding environment with the exception of its polar ending part which is solvent accessible.

Crystallographic analysis of human and bovine holo-RBP indicate that the wall of the cavity around the ligand is substantially apolar; no charged residues are present into the binding pocket, with the exception of the residues Arg121 and Asp102 localized in the upper part of the cavity and

exposed to the solvent. Three disulfide bonds (Cys4-Cys160, Cys29-Cys120 and Cys70-Cys174) are present in plasma RBP, with structural role.

Crystallographic comparisons between the apo and holo form of human and bovine RBP were performed to highlight the differences between the bound and unbound ligand states (Zanotti and Berni 2004). The major differences seem to concern the region encompassing amino acid 33-37 (shown in figure 1.5): in the apo-protein, Phe36 changes its orientation, occupying some of the space left from the retinol and blocking the access to the cavity; in addition, the side chain of Leu35 moves out of the binding site pointing towards the exterior of the protein.

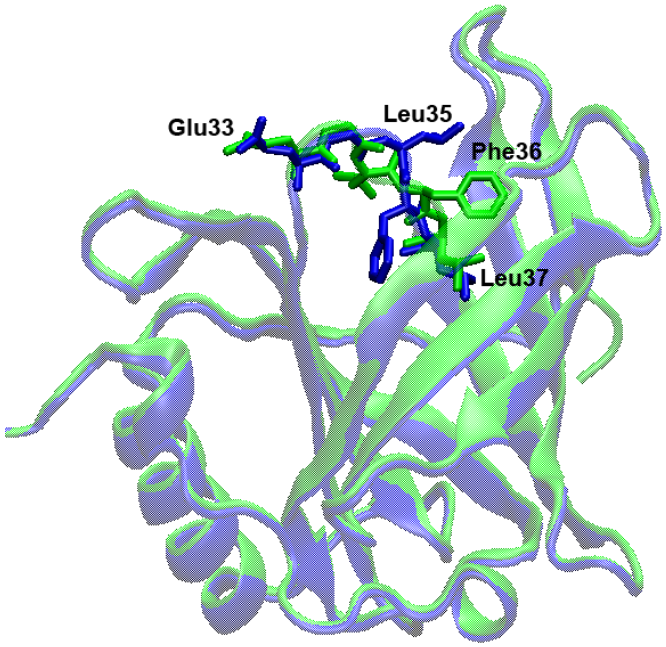


Figure 1.5 Superposition of bovine apo-RBP (PDB ID 1HBQ, green) and bovine holo-RBP (PDB ID 1HBP, blue). In sticks are shown residues 33-37, which change their position in the bound and unbound ligand states.

However the conformational changes associated with the apo-holo transition in RBP is rather limited and confined to one of the entrance loop (A-B). These differences do not give any information regarding the access of the vitamin to the binding cavity, even if the open end of the calyx is most likely to be the portal involved in retinol entering and exiting.

1.4.2 Cellular retinoic acid binding-proteins

Two proteins with high affinity for all-*trans*-retinoic acid have been identified: CRABP-1 and CRABP-2 and they show 74% amino acid identity in human. CRABP-1 binds retinoic acid with somewhat higher affinity than CRABP-2; in mouse the dissociation constant are 0.06 nM and 0.13 nM for type 1 and 2, respectively (Dong et al. 1999). These two carriers display different patterns of expression: CRABP-1 is expressed almost ubiquitously, whereas CRABP-2 is expressed only in the skin, uterus, ovary and the choroid plexus (Noy 2000). In addition, to solubilizing and protecting retinoic acid in the cellular environment, CRABPs are responsible for transport of RA between different cellular compartments to regulate its bioactivity. CRABP-2 has been suggested to deliver retinoic acid to the nucleus, where RA binds to and activates RAR (Schug et al. 2007).

RA, delivered from the cytoplasm to the nucleus by its interaction with the fatty acid-binding protein 5 (FABP5), can also activate the peroxisome proliferator-activated receptor β/δ (PPARβ/δ) (Schug et al. 2007). RAR activation results in growth inhibition while PPARβ/δ activation promotes cell survival. The fate of RA action is determined by expression level of these two transporters (CRABP-2 and FABP5) (Kono and Arai 2015).

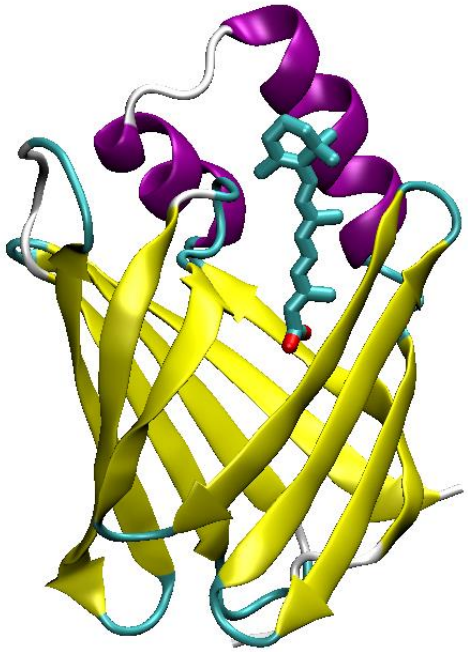


Figure 1.6 Ribbon diagram of wild-type CRABP-2, colored by secondary structure, bound to all-*trans*-retinoic acid (colored by atom type).

1.4.3 Cellular retinaldehyde-binding proteins

Cellular retinaldheyde-binding protein is a 36 kDa soluble retinoid carrier that was isolated from the cytoplasm of bovine retina (Futterman, Saari, and Blair 1977). CRALBP belongs to the Sec14 protein family, which consist of soluble proteins with a lipid-binding domain. CRALBP has high affinity to 11-*cis*-retinal, with a Kd of ~ 20 nM. A crystal structural study also reveals the high stereoselectivity of CRALBP towards 11-*cis*-retinoids (He, Lobsiger, and Stocker 2009). CRALBP is abundantly expressed in RPE and Muller cells (Bunt Milam and Saari 1983), two cell types implicated in rode and cone visual cycles.

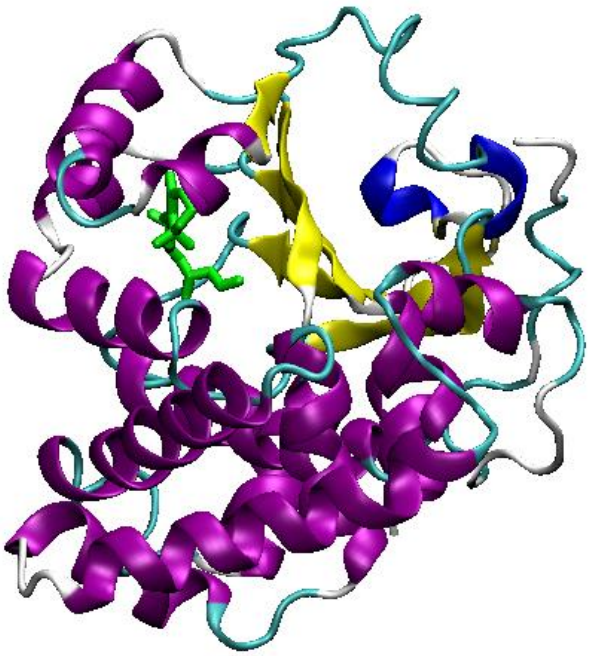


Figure 1.7 Ribbon diagram of wild-type CRALBP, colored by secondary structure, bound to 11-*cis*-retinal (green).

1.4.4 Cellular retinol-binding proteins

Cellular retinol binding proteins are cytoplasmic carriers belonging to the iLBP family, which also includes CRABPs and fatty acid binding proteins (FABPs). All members of the iLBP family share a very similar 3D structure, but transport different types of hydrophobic ligands.

Four isoforms of CRBP (1, 2, 3 and 4) have been reported in human, but only three exist in mice. Mouse CRBP3 appears to be phylogenetically similar to human CRBP4. CRBP1 is the most widespread retinol-binding protein: in particular it is abundant in liver, kidney, lung and

reproductive organs and it is also expressed in the choroid plexus, in the brain, and in pigment epithelial cells in the eyes (Eriksson et al. 1984) (Palczewski et al. 1994). The expression of CRBP2 is restricted to the small intestine (Schaefer et al. 1989) while human CRBP3 is abundant in kidney and liver and human CRBP4 is expressed mainly in kidney, heart, and colon.

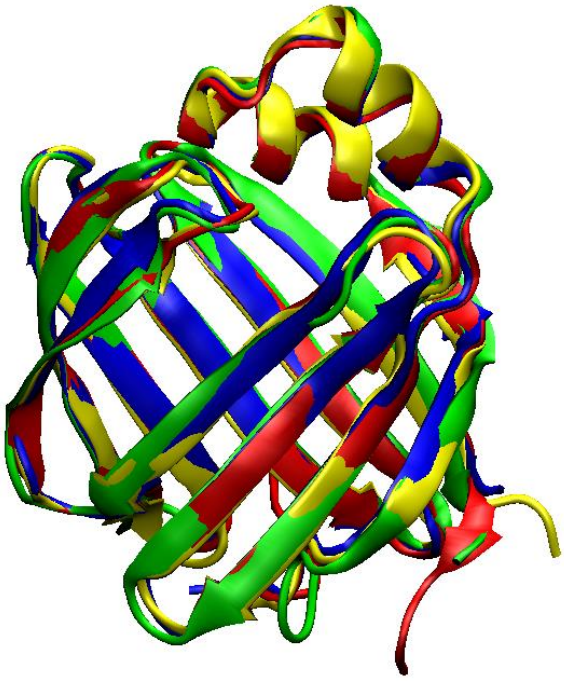


Figure 1.8 Superimposition of rat CRBP1 (yellow), human CRBP2 (red), human CRBP3 (green), human CRBP4 (blue) shown in ribbon representation: the structure is highly conserved.

CRBPs, especially CRBP1, sequester retinoids from the cellular environment, keeping the concentration of free retinoids very low and protecting the relatively labile retinoids from non-specific interactions that might degrade them (Joseph L. Napoli 1999). In addition to these roles, there is evidence that CRBPs have role in the regulation of retinol and retinal metabolism. CRBP1 seems to facilitate the synthesis of retinyl esters by selectively presenting its ligand to lecitine:retinol acyl transferase (LRAT) (Yost, Harrison, and Ross 1988) (Posch et al. 1991), possibly via a direct protein-protein interaction (Jiang and Napoli 2012). Also the RA synthesis seems to involve CRBP1, which may deliver retinol to retinol dehydrogenase (RDH), a microsomal enzyme that converts retinol to retinal (Posch et al. 1991), subsequently converted to retinoic acid by retinal dehydrogenase. Also CRBP2 participates in retinoid metabolism: in the intestine, it binds retinal produced for the b-carotene cleavage.

The role of CRBP 3 and 4 is not still clarified and despite they are classified as retinol-binding proteins on the basis of their sequences, their affinity for retinol is very low.

Despite the high structural conservation (Fig. 1.8) of the four CRBPs, they exhibit low sequence identity (Fig. 1.9) and a different affinity for retinol.

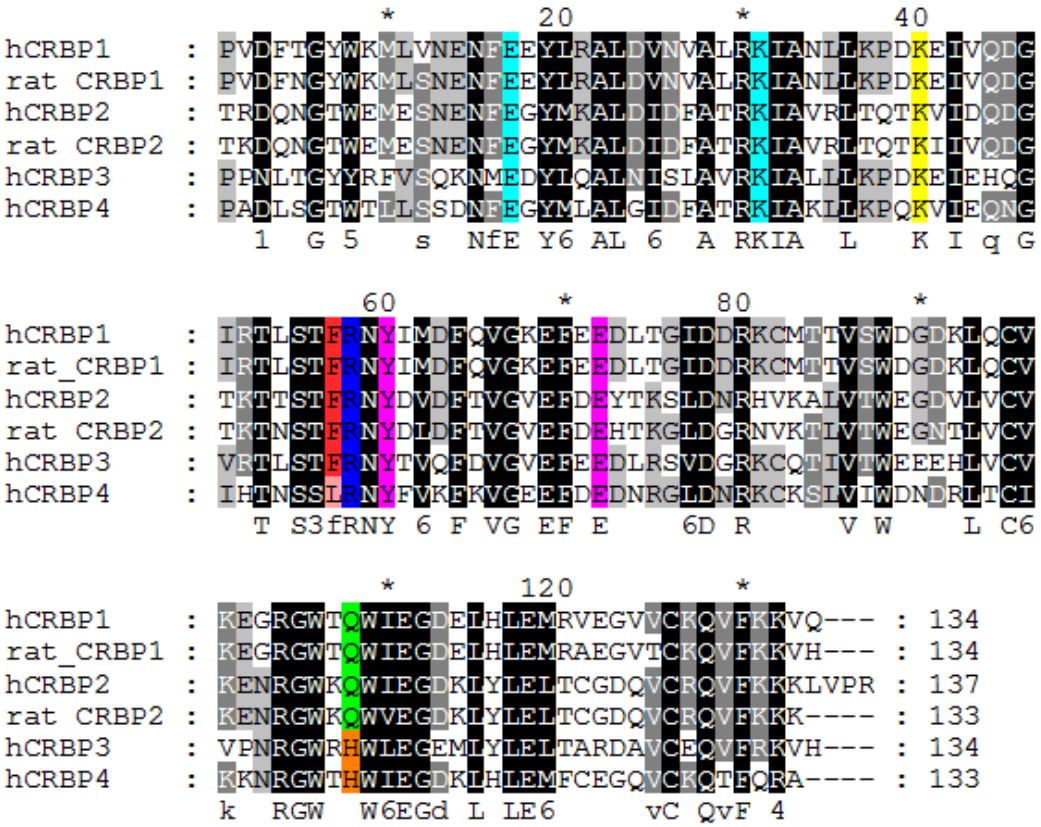


Figure 1.9 Multiple alignments of the amino acid sequences of the four human CRBP isoforms and rat CRBP1 and 2. Black letters indicate high residues conservation. In the text we used the numeration reported for rat CRBP1 as reference. The coloring changes from black to white with the increase of the diversity between residues (black: highly conserved residues). We have colored in yellow the conserved Lys40; in green and orange the binding key residue at position 108; in cyan the conserved residues Glu17 and Lys31; in pink the conserved residues Lys60 and Glu72; in blue the conserved R58; in red residues at position 57. UniProt accession numbers are: P02696 and P06786 for rat CRBP1 and 2 respectively; P09455 and P50120 for human (h) CRBP1 and 2, respectively; P82980 and Q96R05, for hCRBP3 and 4, respectively.

As members of the iLBP family, CRBPs are characterized by a global fold consisting in a flattened β-barrel formed by two orthogonal β-sheets, which in turn are composed by a total of 10 β-strands (A-J); one end of the barrel is partially obstructed by the N-terminus of the protein, often assuming a short 3_{10} helix structure, while the other end is covered by a helix-turn-helix domain (alpha-helix I and II) (Fig 1.10). The formation of a salt bridge (Glu 17 – Lys 31) (Fig 1.10) could help to stabilize the loop between the two helices. No disulfide bridges are present. Between the strand D and E there is the so called “gap region”: a gap in the β-barrel structure due to a bulge at the highly

conserved Gly67 residue. This gap seems not to be the entry point for the ligand, because several side chain of adjacent residues fill the gap, blocking the access to the cavity (Franzoni et al. 2002). Furthermore, the hydrogen bond between Glu72 on the β-strand E and Tyr60 on the β-strand D (Fig 1.10), conserved in the four CRBP isoforms, probably stabilizes the gap region. The function of this feature is unknown but typical of the iLBP family (Young et al. 1994); however, between the various members of this family there are clear distinctions in structural stabilities, in ligand binding properties (Zimmerman, van Moerkerk, and Veerkamp 2001) and in their backbone dynamics (Hanhoff, Lücke, and Spener 2002). According with many studies regarding the dynamics of CRBPs, it was proposed that the entry site of retinol is located between α-helix II, the subsequent linker to β-strand B, the CD and EF loops (Banaszak et al. 1994).

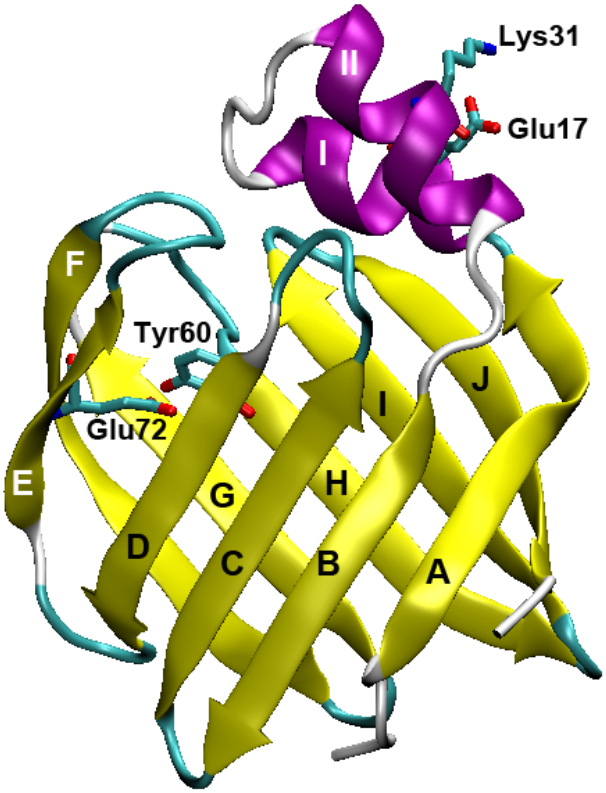


Figure 1.10 Cartoon representation of a generic CRBP, colored by secondary structure. In stick, colored by atom types, are shown Glu72 and Tyr60, conserved in all human CRBPs.

1.4.4.1 Cellular retinol binding protein 1

The first member of the family of intracellular retinoid-binding protein to be discovered was CRBP1 in 1973 (Bashor, Toft, and Chytil 1973). The potential importance of this protein was

suggested by its wide distribution, with the highest abundance in the liver, kidney, lung and retinal pigment epithelium cells of the eye (Eriksson et al. 1984).

CRBP1 is a 134 residues cytosolic protein, structurally formed by 10 beta-strands organized in a beta-barrel capped with two alpha helices, as the other members of the iLBP family. The identity of its endogenous ligand, all-*trans*-retinol, was established in 1982 by high-pressure liquid chromatographic analyses, after extraction of the ligand from CRBP purified from bovine retina (Saari, Methods Enymol, 81:819-826, 1982). The Kd value of the CRBP1-retinol complex was not so easy to determine because the affinity for this ligand is very high, but by direct and competitive ligand-binding assays it was estimated a Kd lower than 0.1 nM (Malpeli et al. 1995). The formation of the retinol-CRBP1 complex is clearly detectable by using a spectrophotometer: the characteristic spectrum has a peak at the maximum wavelength of 348-350 nm and two lateral shoulders as shown in figure 1.11. Furthermore, CRBP1 is able to bind 13-*cis*-retinol, 3-dehydroretinol and 9-*cis*-retinal with high affinity, but is not able to bind retinoic-acid and 11-cis-retinol (MacDonald and Ong 1987).

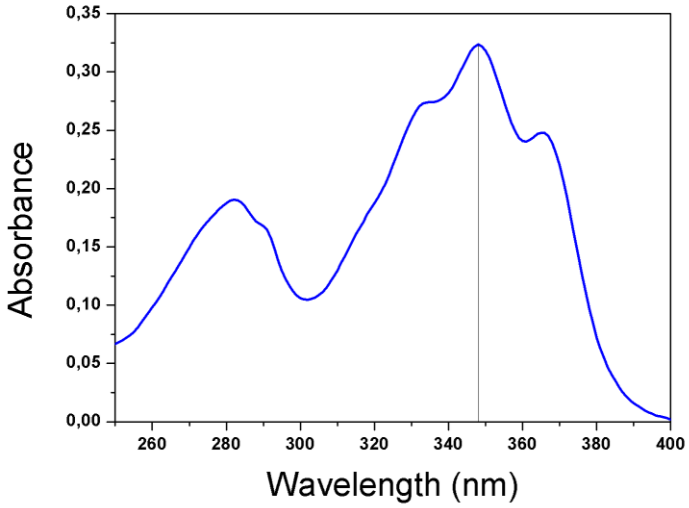
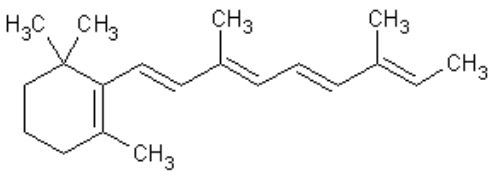
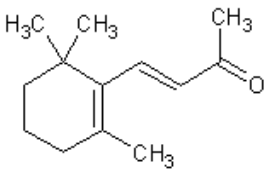


Figure 1.11 Absorption spectrum of the retinol-hCRBP1 complex.

Axerophtene, a retinol analog bearing a hydrogen atom in place of the hydroxyl end group, and β-ionone, in complex with CRBP1 exhibit rather low binding affinity; this suggests that the hydroxyl group and the isoprene tail contribute to the retinol-binding affinity and specificity (Malpeli et al. 1995).



axetophthene



β–ionone

CRBP1-retinol complex is the most studied because retinol is considered the endogenous ligand of isoforms 1 and 2. A recent study performed by Silvaroli and co-workers reports the crystal structure of the human holo-CRBP1 (Silvaroli et al. 2016) in complex with all-trans retinol, confirming what was already known about CRBP1-retinol complex of other organisms: the retinol is deeply buried inside the binding cavity and anchored to the key residues Glutamine 108 and Lysine 40. As shown in figure 1.12, the hydroxyl group of retinol acts as a hydrogen donor to the carbonyl group of the Gln108 side chain, the NH2 group of which is located close to the center of the Phe4 ring, probably forming an amino-aromatic interaction that stabilizes it (Dougherty 2013). The role of Lys40 is to be clarified, because it can act as donor of hydrogen bonds with the hydroxyl group of retinol and also with Threonine 53 and Glutamine 128 (in the latter case, mediated by a water molecule), but an interaction between the positive charge of the Lysine with the conjugated π electron field of the isoprene tail is also possible (Cowan, Newcomer, and Jones 1993) (Franzoni et al. 2002). However, no mutant forms concerning the key residues were available till now.

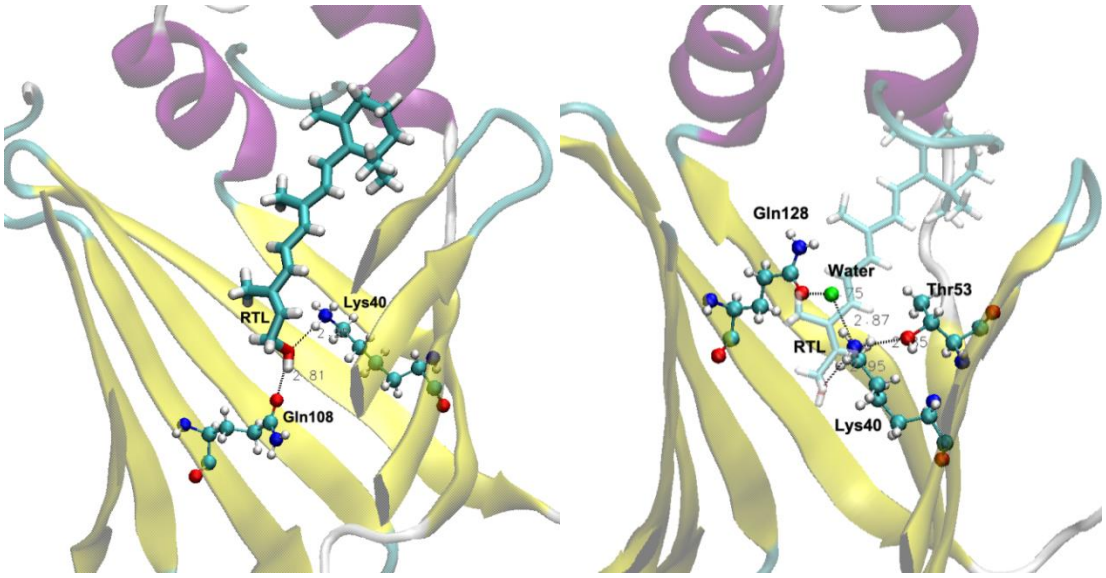


Figure 1.12 Ribbon representation of hCRBP1 (PDB ID=5H8T), colored by secondary structure, bound to all-trans-retinol (stick representation); the hydrogen bond network in holo-CRBP1 are shown with a ball and stick representation: on the left the interaction between retinol (RTL) and the two key residues; on the right the three possible hydrogen bonds between Lys40 and Thr53, RTL and Gln108.

In CRBP1, the “helical cap” (the helix-tour-helix motif typical of CRBPs) seems to be involved in the interaction with the retinol hydrogenases (RDHs) enzyme. In a study realized by Nalpoli and Penzes, with the aim to demonstrate that these region is important for protein-protein interaction, the binding of various mutants of CRBP1 (L29A, I32E, L35A, L35R, L36A, F57A, R58A, R58E) with retinol was also evaluated. This study has revealed that L35 seems to be important for the up-take/release of retinol, while F57, a suspected key residue in controlling ligand entrance/exit, seems not to be relevant, because F57A mutant had a Kd similar to the WT (Penzes and Napoli 1999).

1.4.4.2 Cellular retinol-binding protein 2

Cellular retinol-binding protein 2 (CRBP2), the second major vitamin A transporter, is clearly involved in the metabolism and transport of intracellular retinol. Differently from CRBP1, that is widely expressed in many tissues, CRBP2 is more prevalent in the small intestine (Crow and Ong 1985) suggesting its involvement in the regulation of the intestinal retinoid metabolism; it has also been reported that CRBP2 facilitates the reduction of retinal to retinol (Kakkad and Ong 1988) and subsequently to retinyl ester (MacDonald and Ong 1988). CRBP2 binds its endogenous ligand, all-*trans* retinol, with a Kd approximately of 10 nM (100 times higher than CRBP1) (Li et al. 1991). The formation of the retinol-CRBP2 complex is easily detectable by absorption spectrum: similarly to CRBP1 the addition of retinol to CRBP2 causes the formation of a peak with a maximum wavelength of 348nm. The crystal structures of human CRBP2, both in complex with *all-trans*-retinol and in the apo form, were available since 2007 (PDB ID 2RCT and 2RCQ respectively) (Tarter et al. 2007), but a most recent work by Nossoni and co-worker (Nossoni et al. 2014) presents the crystal structure of CRBP2-retinol complex at a better resolution (PDB ID 4QYN) and, for the first time, has resolved the crystal structure of CRBP2 bound to *all-trans*-retinal (PDB ID 4QYP). Even in the case of CRBP2, the main residues responsible to the ligand binding are Gln 108 and Lys 40. These two residues are important for both retinol and retinal binding, but interact with these ligands in a different way. However, CRBP1 has a low affinity for *all-trans*-retinal (~90-100 nM) while CRBP2, that shown a 56% amino acid sequence identity with the isoform 1, binds it with a Kd lower than 50 nM (Levin et al. 1988).

As shown in Figure 1.13 the retinol hydroxyl group can form two direct hydrogen bonds with Gln 108 and Lys 40, as in CRBP1-retinol complex.

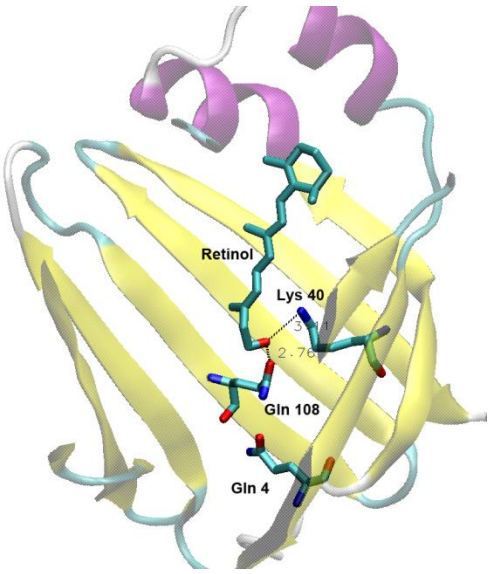


Figure 1.13 Ribbon representation of hCRBP2 (PDB ID=4QYN), colored by secondary structure, bound to the all-trans-retinol (stick representation); the hydrogen bond network between retinol hydroxyl group and the key residues is shown with a ball and stick representation.

Despite the conserved interactions between retinol and CRBP1 and 2, the aforementioned affinity of the two isoforms for this ligand is different and the reason is still not clear. It was suggested that in CRBP1 the presence of a phenylalanine residues at position 4 holds the side chain of Gln108 in a favorable position to interact with the retinol hydroxyl group, serving as H-bond acceptor. Because in CRBP2 the amino acid at position 4 is a glutamine, Gln108 can orient its side chain as either a hydrogen bond donor or acceptor. This would explain why CRBP2 binds the aldehyde and the alcohol with similar affinity. Nevertheless, F4Q-typeI and Q4F-typeII mutants show an affinity for retinol similar to WT types (Jamison et al. 1995).

In the case of retinal-CRBP2 complex, the crystal structure (PDB ID 4QYP) reveals that the position of the retinal, apart from the oxygen, is almost perfectly superimposable with that of retinol in retinol-CRBP2 complex structure. The carbonyl group of the retinal makes a water-mediated interaction with Gln108 and Lys40 (Fig. 1.14) and this is a unique feature for CRBP2 (Nossoni et al. 2014). The side chain of Gln108 is fixed by a small network of water molecules and residues (Thr1, Asp9, and Gln4).

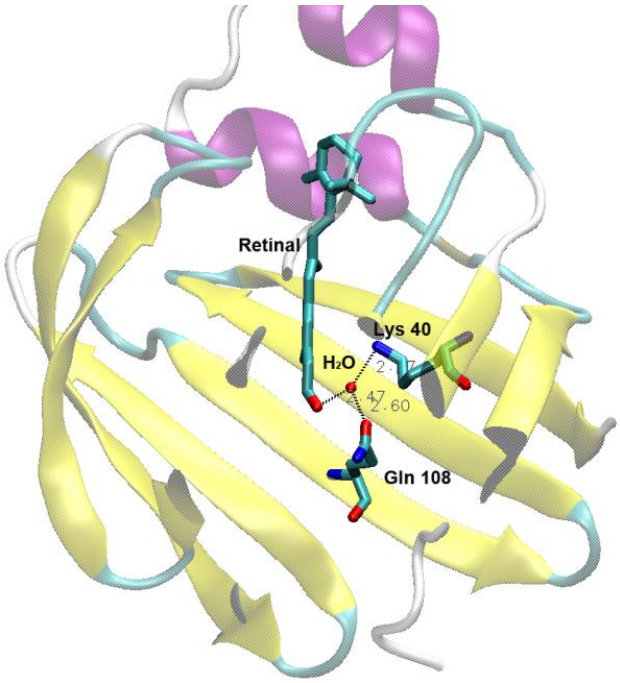


Figure 1.14 Ribbon representation of hCRBP2 (PDB ID=4QYP), colored by secondary structure, bound to all-trans-retinal (stick representation); the hydrogen bond network between retinal carbonyl group and the key residues is mediate by a water molecules and shown in a ball and stick representation.

The study of the CRBP2 cavity, has suggested that the identity of the amino acid at position 51 may be a deciding factor for the retinal positioning in CRBP1 and 2: retinal cannot bind to CRBP1 in the same way that it does in type 2 due to a possible steric clash between Isoleucine 51 presents in CRBP1 and retinal carbonyl group. However, T51I-CRBP2 mutant shown the same affinity for *all-trans*-retinol and *all-trans*-retinal as compared with the WT form (Nossoni et al. 2014). In conclusion remain to be clarifying the real cause of different affinity for retinoid in CRBP1 and 2.

1.4.4.3 Cellular retinol-binding protein 3

A putative new human CRBP was identified in 2000 by Folli and co-workers (C Folli et al. 2001) on the basis of an N-terminal amino acid sequence deriving from an unidentified protein copurified with CRBP1 from bovine kidney. This putative CRBP, designate CRBP3, that has an estimated molecular mass of 15,9 KDa, shares a considerable sequence identity with iLBPs. It exhibits a remarkably higher sequence identity with CRBPs (55.6% and 49,6% identity with human CRBP I and II, respectively) than with the other iLBPs thus far characterized (less than 40% identity). These data had suggested that CRBP3 presumably belongs to the CRBP family. After recombinant

expression in E.coli and purification of isoform 3, various retinoids (all-*trans*-retinol, 9-*cis*-retinol, 13-cisretinol, all-*trans*-retinaldehyde, all-*trans*-retinoic acid, and all-*trans*-3,4-didehydroretinol) were tested in their interaction with this protein. The spectra obtained with the major part of the retinoids added to CRBP3 are very close to those of the unbound compounds, and, as expected, they disappeared gradually after prolonged incubation. The addition of all-*trans*-3,4-didehydroretinol to CRBP3 seems to produce an absorption spectrum closely resembling the spectrum of the same retinoid added to CRBP1, but probably the bound is weak. Finally, the addition of all-trans-retinol to CRBP3 shows a spectrum similar to those of retinol bound to CRBP1 and 2, whilst less intense (figure 1.15). Spectrofluorometric titration of human CRBP3 with retinol had permitted to calculate a Kd of approximately 60 nM.

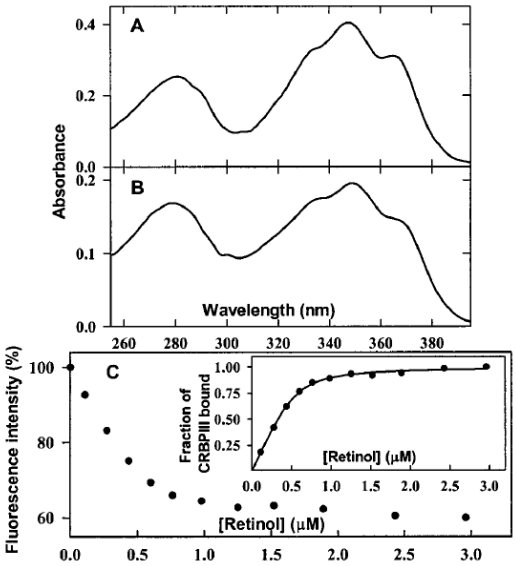


Figure 1.15 The absorption spectra of (A) the retinol–CRBP1 and (B) the retinol–CRBP3 complexes were recorded after the addition of a slight molar excess of all-trans-retinol to recombinant rat apo-CRBP1 and human apo-CRBP3. (C) Spectrofluorometric titration of human CRBP3 with retinol. Intrinsic protein fluorescence (%) of apo-CRBP3 is plotted as a function of retinol concentration.

This similarity indicates that the ligand interacts with CRBP3 in a specific manner and that such interaction is similar to that established between retinol and CRBP1. In the case of the retinol–CRBP3 complex, however, the A350/A280 ratio is lower than that of the retinol–CRBP1 complex, and a tentative of co-crystallization of the retinol with CRBP3 has led to the obtaining of an apo-CRBP3 crystal structure (pdb ID=1GGL, (C Folli et al. 2001)) (Fig. 1.16), that shows a very high structural conservation if compared with other CRBPs (Fig. 1.8).

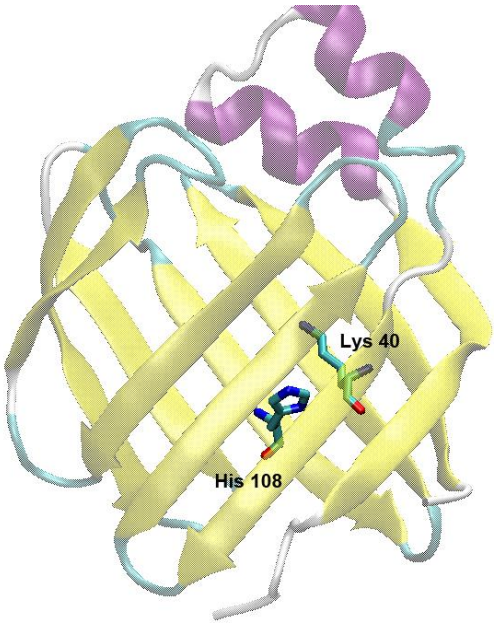


Figure 1.16 Ribbon representation of apo-hCRBP3 (pdb ID=1GGL), colored by secondary structure. The two presumable key residues, His108 and Lys40, are represented in stick and colored by atom types.

This unexpected result has made it impossible to define how retinol interacts with the binding cavity of the isoform 3. As we can observe in figure 1.9, the multiple alignment of the four human CRBPs reports a difference in the two key residues: a substitution of Gln108 with a histidine residue is present only in CRBP3 and 4. It is possible that this replacement can modify the affinity of these carriers for retinol.

1.4.4.4 Cellular retinol-binding protein 4

Two years later the identification of CRBP3, another human member of CRBPs was characterized (Claudia Folli et al. 2002). Specifically, this protein, designated human CRBP4, shares a higher sequence identity with human CRBP1 (57%) and 2 (58%) than with human CRBP3 (49%). A very high sequence identity (89%) is shared with murine CRBP3, clearly indicating that human CRBP4 is most likely the human ortholog of murine CRBP3. The obtaining of a recombinant CRBP4, expressed in E.coli and purified, has permitted to test its interaction with retinol. The addition of all-trans-retinol to human apo-CRBP4 produces an absorption spectrum clearly distinct from those of both free retinol and retinol bound to previously characterized CRBP types, with absorbance intensity, in the 320-340 nm range, lower than that of the protein at 280 nm (Fig. 1.17).

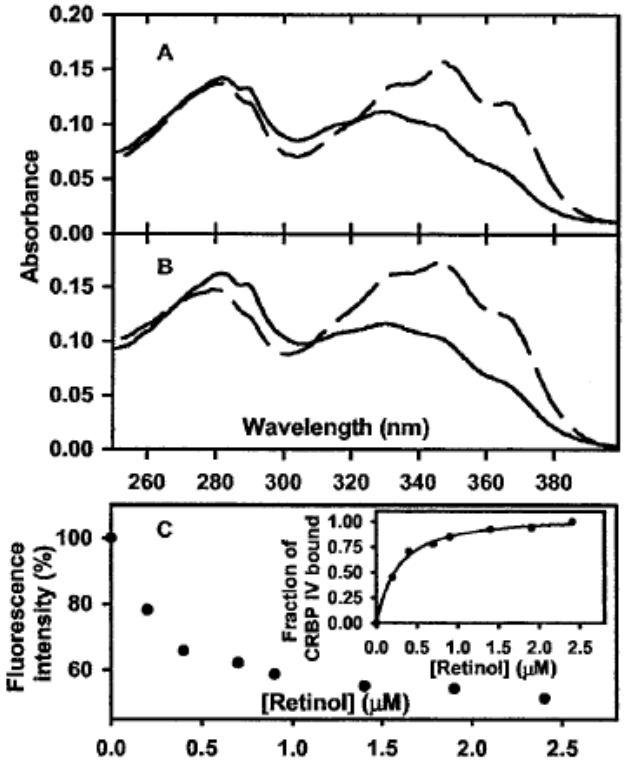


Figure 1.17 Retinol binding to human apo-CRBP4. (A) Absorption spectra of the complexes of retinol with wild-type human CRBP4 (solid line) and human CRBP1 (dashed line). (B) Absorption spectra of the complexes of retinol with human H108Q-CRBP4 mutant (solid line) and human Q108H-CRBP1 mutant (dashed line). (C) Intrinsic protein fluorescence (percentage) of apo-CRBP4 is plotted as a function of retinol concentration. Inset: fitting of the experimentally determined fractional saturation to a theoretical curve, giving an apparent Kd of 200 nM.

Similar to CRBP3, the isoform 4 contains the substitution of a glutamine to a histidine residue in position 108, as compared with CRBP1 and 2 (Fig 1.18). Although this replacement is the same present in CRBP3, the retinol-CRBP4 complex shows a different absorption spectrum, and the fact that the intensity decreases quickly in time seems to suggest a very low affinity and perhaps a different mode of binding. Unfortunately, the crystallization of CRBP4 in the presence of a molar excess of all-trans-retinol led to an apo-CRBP4 3D structure (pdb ID = 1LPJ) (Claudia Folli et al. 2002). Spectrofluorometric titration of human CRBP4 with retinol permitted to calculate a Kd of approximately 200 nM (Fig 1.17C).

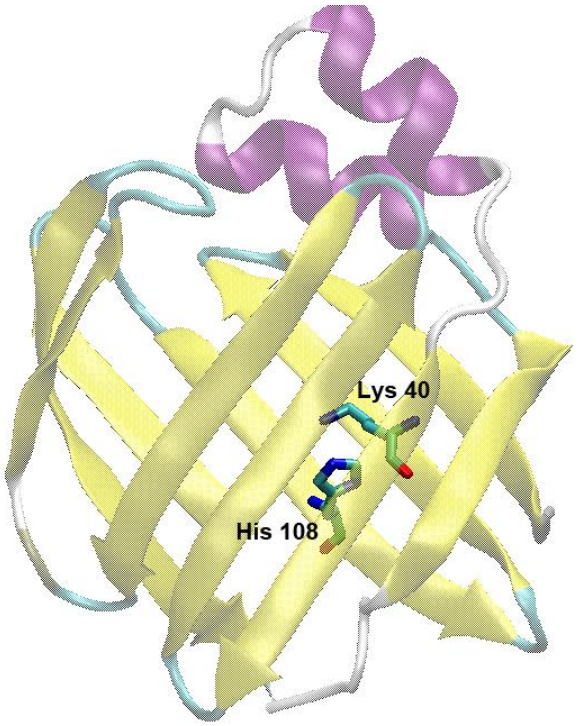


Figure 1.18 Ribbon representation of apo-hCRBP4 (PDB ID=1LPJ), colored by secondary structure. The two putative key residues, His108 and Lys40, are represented in stick and colored by atom types.

With the aim to ascribe the contribution of this different key residue to the difference in the absorption spectrum, two mutants were obtained: H108Q-hCRBP4 and Q108H-CRBP1, trying to restore, for the CRBP4 mutant, the typical spectrum seen for CRBP1 and 2, and, for CRBP1 mutant, the one of CRBP4. As inferable by figure 1.17B, which reports the absorption spectra of the mutants, the interactions of all-trans-retinol with mutant forms seem to be highly similar to those present in the corresponding wild-type holo-proteins and it was concluded that as-yet-unidentified factors, other than the presence of a His residue at position 108, are responsible for the unusual spectrum of holo-CRBP4.

1.5 Ligand exchange in intracellular lipid binding proteins

Structural studies have shown that iLBP binding site is not accessible from the protein bulk and a conformational change is needed for the entry of the ligand. Although a broad literature is available on the comparison between apo- and holo-form of a number of iLBP members (Winter, Bratt, and Banaszak 1993) (Calderone et al. 2002) (Silvaroli et al. 2016) (Xu, Bernlohr, and

Banaszak 1993) (Vaezeslami et al. 2006), at date an obvious path for ligand entry is not defined, due to the very few differences between the ligand bound and unbound form of the proteins. Nevertheless, it was suggested a putative “portal region” formed by the alpha helix II, the linker between this helix and the beta strand B, and the CD and EF loops. This hypothesis was proposed when a little opening was found on the surface of the intestinal FABP (Sacchettini, Gordon, and Banaszak 1989). Other evidences, in agreement with this hypothesis, were supplied by crystallographic studies on other FABPs, in which a little part of the ligand stays exposed to the solvent (Hodsdon and Cistola 1997). NMR studies on rat CRBP1 (Franzoni et al. 2002) and other FABPs (Hodsdon and Cistola 1997) (Cai et al. 2012) confirmed the structural similarity of the apo- and holo forms also in solution and suggested the important role of the portal region dynamics in the ligand binding mechanism. A number of data are available about the different flexibility of the portal region in various iLBPs, ranging from the relatively rigid ileal FABP and CRBP1 and 2 to the more flexible myelin, adipocyte, and liver FABPs, to almost completely disordered intestinal FABP. The different dynamic behavior may be the cause of various binding preferences or of different binding mechanisms, specific for each protein (Ragona et al. 2014).

1.6 NMR studies on rat CRBP1

The crystal structures of most iLBPs reveal only minimal differences between the apo- and the holo-form. NMR data confirm the structural similarities and suggest that a crucial role in ligand binding is likely to be played by the higher conformational flexibility of the portal region in the apo-form. An accurate NMR study on the backbone dynamics of rat CRBP1, was carried out by Franzoni and co-workers with the aim to gain new insight about the ligand up-take in rat CRBP (Franzoni et al. 2002). The comparison of the apo- and holo-forms of rat CRBP1 in solution reveals high similar structures. The tertiary structure of these two forms are almost superimposable and the region that shows more extensive backbone variations involves the N-terminus, the loop between the two alpha helices, the entire alpha helix II, the subsequent linker to βB and almost all the loops connecting the β-strands. The size of the gap between βD and βE, obtained measuring the distance between residues Phe64 and Phe70 (located at the base of CD and EF loops), is the same in apo- and holo-CRBP1. In the holo-protein, the β-ionone ring of retinol is located in a hydrophobic pocket defined by residues belonging to alpha helix II and loops CD and EF; in the

apo-form, despite these regions are less defined, the aforementioned hydrophobic residues adopt a conformation similar to the holo-form, with the exception of Arg58 and Ile77. In several conformers of the apo-protein, the mutual disposition of the two latter side chains leaves a small aperture, but the space isn't larger enough to permit the passage of retinol (Franzoni et al. 2002).

In vivo, ligand up-take occurs in proximity to the membrane. In fact, it is known that close to a biomembrane, both local pH and dielectric constant are lower than in the bulk solution; the micro-environment might assist a transient opening of the more flexible structure at the portal region. However, *in vitro*, holo-CRBP is easily obtained by mixing the apo-protein with retinol in the solution. Likely, a spontaneous up-take is always possible, while, when the protein is near the membrane, the entry of the ligand may be facilitated by the environmental conditions.

It remains to be understood how the ligand enters into the binding cavity and how, after intracellular transport, it is released to the membrane-bound enzymes that catalyze its conversion in retinaldehyde (Lapshina et al. 2003) (Posch et al. 1991). Another study realized by Mittag and co-workers, that reports the NMR line shape changes analysis caused by the interaction of rat CRBP1 with retinol, suggests that the initial contact of retinol with the apo-CRBP1 leads to structural changes in the portal region of the protein, allowing specific and high affinity binding, after a number of non-specific associations of the ligand with the surface (Mittag et al. 2006).

1.7 NMR studies on rat CRBP2 in comparison with CRBP1

NMR studies on CRBP2, performed in 1999 (Lu et al. 1999), suggest that, in solution, apo-CRBP2 undergoes conformational changes on the microseconds to milliseconds time-scale, which result in increased access to the binding cavity. A small opening of the portal region was observed, which could potentially accommodate the initial entry of the polar end of the retinol molecule. They have found that in the solution structure of apo-CRBP2, the binding cavity was more accessible to ligand entry as compared with the crystal structure. One of the major differences noticed was the reorientation of the Phe57 side-chain, even if it was not sufficient to accommodate the ring moiety of the retinol molecule. Nevertheless, the localized backbone disorder observed in the segment 27-37 and at βE-βF loop may allow further opening of the portal to accommodate the β-ionone ring of retinol. Once the ligand has entered the binding pocket, a hydrophobic interaction

between Phe57 and the bound retinol may cause the reorientation of Phe57 side chain. Moreover, the distance between CD and EF loops in the solution structure of apo-CRBP2 is increased, if compared with both crystal structure of apo and holo-forms. The NMR data about the solution structure of apo-CRBP2, show that the segment 27-34 does not retain a helical conformation, while in the holo-CRBP2 the same segment is a helix and all the structure is more rigid (Lu et al. 1999).

In 2010, Franzoni and co-workers investigated the conformational features of the rat CRBP2 ligand entry portal under a variety of solution conditions (Franzoni et al. 2010). Contrary to the previous work (Lu et al. 1999), the NMR data obtained on rat apo-CRBP2 showed that the segment 27-34 does adopt a helical conformation, in agreement with the crystal structures of rat and human CRBP2. Hence, the previously suggested structural implication for the mechanism of retinol binding to CRBP2, need to be reconsidered. Franzoni and co-workers concluded that in solution the global fold of the portal region in apo-CRBP2 is very similar to that of apo-CRBP1. Furthermore, the differences in local flexibility between the apo-forms of the two primary cellular retinol carriers, are not restricted to helix II, but occur also in helix I and in β-strands A to F. However, the alpha helix II represents the secondary structure element with the highest flexibility in the holo-form of both proteins, that could be an important feature to assist a transient conformational change that opens the portal during ligand release (Franzoni et al. 2010).

AIM OF THE STUDY

2. Aim of the Study

Three aspects were mainly faced in this work: to clarify the differences between CRBPs, to give insights on the real role of the binding residues, in particular of CRBP1, and to explore the ligand up-take mechanism of the four isoforms. To this aims, we have obtained and characterized some mutants of CRBP1 in which residues at the two key positions 40 and 108 were replaced with a leucine. Then, by means of molecular dynamics simulations techniques, we studied the protein flexibility of the four isoforms and the ligand entrance mechanisms, suggesting models of up-take that we are experimentally validating by the characterization of selected mutants.

EXPERIMENTAL TECHNIQUES

3.1 Site-directed mutagenesis

In vitro site-directed mutagenesis is an invaluable technique for characterizing the dynamic, complex relationships between protein structure and function, for studying gene expression elements, and for carrying out vectors' modification.

To obtain hCRBP1 mutants we have used the Quick Change site-direction mutation methods (Fig. 3.1) based on the use of a Pfu (high fidelity) DNA polymerase together with a pair of mutagenic primers. The oligonucleotide primers, each complementary to opposite strands of the vector and containing the mutation, are extended during the PCR (Polymerase Chain Reaction). Extension of the oligonucleotide primers generates a mutated plasmid containing staggered nicks. At the end of the reaction, a lot of mutant vectors and a little part of the parental methylated DNA (used as a template) are both present. To eliminate the WT parental DNA, the sample is treated with the DpnI endonuclease, that specifically recognizes and digests the methylated DNA. The nicked DNA vector, containing the desired mutations, is then transformed inside XL1-Blue super-competent cells, in which the nick is repaired by the cell ligase. Plasmidic DNA of surviving cells is sequenced to verify the presence of the mutation.

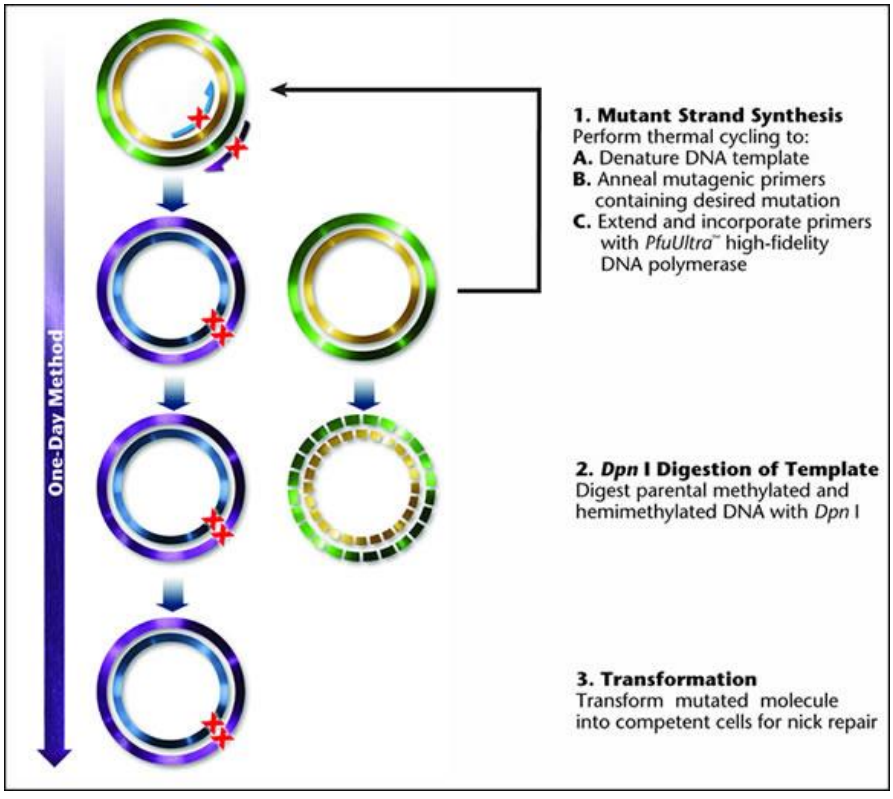


Figure 3.1 Quick Change mutagenesis method.

3.2 Expression of recombinant proteins

Escherichia coli is one of the organisms of choice for the production of recombinant proteins. Its use as a cell factory is well-established and it has become the most popular expression platform. For this reason, there are many molecular tools and protocols at hand for a high-level production of heterologous proteins, such as a vast catalog of expression plasmids, a great number of engineered strains and many cultivation strategies. In theory, the steps needed for obtaining a recombinant protein are: take your gene of interest, clone it in whatever plasmid expression vector you want, transform it inside the host of choice, induce the protein expression, and then the protein is ready for purification and characterization. However, something can go wrong, for example a slow or little growth of the host, the formation of inclusion bodies, the inactivity of protein, and much more.

The advantages of using *E. coli* as the host organism are well known:

- It has incomparable fast growth kinetics in glucose-salts media and in optimal environmental conditions its doubling time is of about 20 min.
- High cell density cultures are easily achieved.
- Rich complex media can be made from readily available and inexpensive components.
- Transformation with exogenous DNA is fast and easy.

The pET expression system is one of the most widely used systems for the cloning and in vivo expression of recombinant proteins in E. coli. In the pET system, the protein coding sequence of interest is cloned downstream of the T7 promoter; this sequence ia a strong promoter and recognizes specifically the RNA polymerase of the bacteriophage T7 (T7 RNA polymerase). To use this system is necessary to transform, with the pET vector, genetically engineered host cells, which contain, in their chromosomal DNA, the coding gene for the T7 RNA polymerase. This gene is under the control of the lac promoter, that means that the gene is translate only if an inductor - like IPTG (Isopropil-β-D-1-tiogalattopiranoside) - recognized by lac promoter is present. The T7 RNA polymerase recognizes the T7 promoter on the plasmid vector pET, and promotes the expression of the sequence of interest (Fig. 3.2). The ribosomal apparatus on the cell, provide the recombinant protein, which is released into the cytosol.

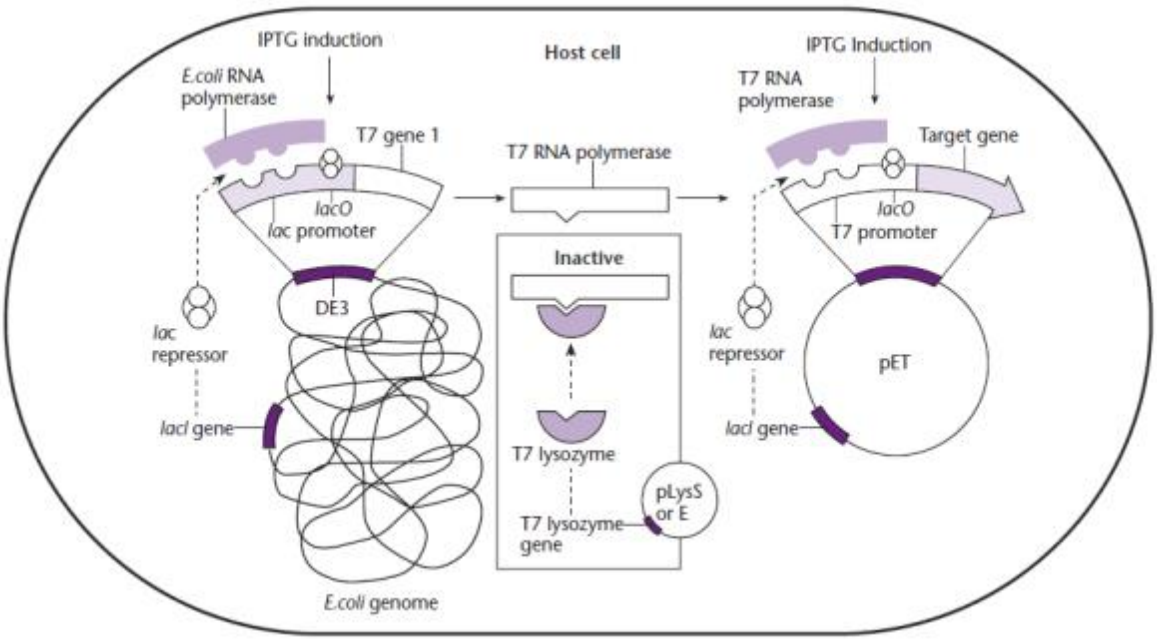


Figure 3.2 pET expression system.

3.3 Protein purification

Together with the protein of interest, a number of other E.coli proteins are expressed. To achieve a pure protein solution is necessary to have a good purification system. In our case, two chromatographic passages are sufficient to obtain a good degree of purity.

3.3.1 Ion exchange chromatography

The most popular method for the purification of proteins and other charged molecules is the ion exchange chromatography, in which negatively charged molecules are attracted to a positively charged solid support.

To optimize the binding of all charged molecules, the mobile phase is generally a solution with low to medium conductivity (low to medium salt concentration). The adsorption of the molecules to the solid support is driven by the ionic interaction between the oppositely charged ionic groups in the sample molecule and in the functional ligand on the support. The strength of the interaction is determined by the number and location of the charges. By increasing the salt concentration (generally by using a linear salt gradient) the molecules with the weakest ionic interactions start to elute first from the column. Molecules that have a stronger ionic interaction require a higher salt

concentration and elute later (Fig. 3.3). As a rule, the pH of the mobile phase buffer must be between the pI (isoelectric point) of the charged molecule and the pKa of the charged group on the solid support.

As in most of the chromatographic methods, a protein sample in a low ionic force solution is injected into the column, where it will be strongly retained (Tris HCl 30mM pH= 7.5 in our case). A gradient of linearly increasing salt concentration is then applied to elute the sample components from the column. To purify hCRBPs we used the automate Fast Liquid Chromatographic System AKTA (Amersham Pharmacia Biotech), using as solid support a positively charged resin. Proteins separated during the chromatography are collected in different tubes and their absorbance at 280nm is analyzed; then, by means of a polyacrylamide gel, the fractions containing the protein of interest are identified.

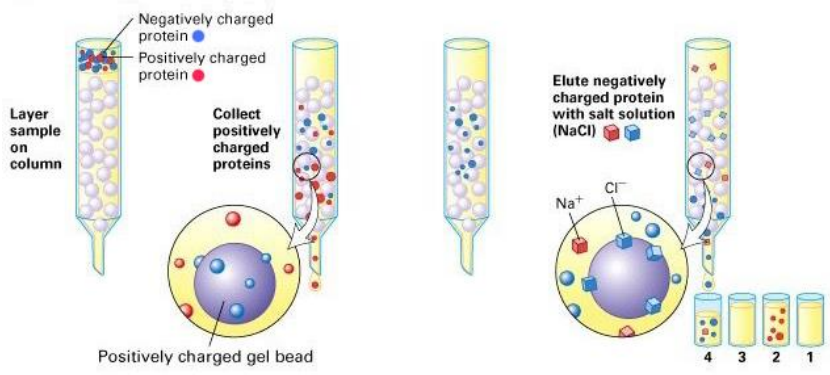


Figure 3.3 Ion-exchange chromatography mechanism.

3.3.2 Size exclusion chromatography

Size exclusion chromatography separates molecules based on their size by filtration through a gel. The gel consists of spherical beads containing pores of a specific size distribution. Separation occurs when molecules of different sizes are included or excluded by the pores within the matrix. Small molecules diffuse into the pores and their flow through the column is retarded according to their size, while large molecules do not enter the pores and are eluted in the column's void volume. Consequently, molecules separate based on their size as they pass through the column, where they are eluted in a decreasing molecular weight order (fig 3.4). To purify hCRBPs (~16KDa) the Ultrogel® Aca 54 matrix, that have a fractionation range of 5000-70000 Da was used.

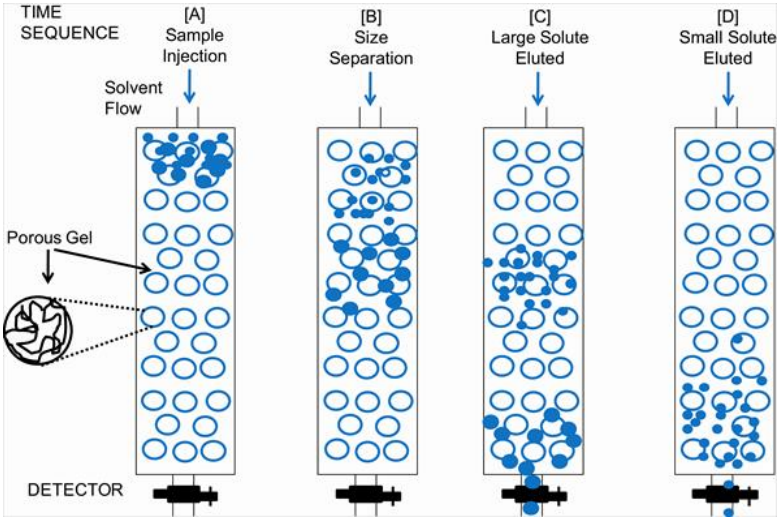


Figure 3.4 Size exclusion chromatography mechanism.

3.4 SDS-PAGE

SDS-PAGE, acronym for sodium dodecyl sulfate (SDS) polyacrylamide gel electrophoresis, is the most widely used technique to separate proteins from multi-protein samples. In the presence of an electric field, proteins migrate towards the anode inside the poly-acrylamide gel under denaturing conditions. In SDS-PAGE, a denaturant action by both heat and the detergent SDS, which also covers all the proteins with a negative charge, determine that the electrophoretic mobility of a protein is only affected by its molecular weight in the porous acrylamide gel. At the end of the run, gel is stained in a solution containing Comassie Brilliant Blue, that binds the positively charged groups of the protein. A destaining solution is then applied to the gel to remove the dye aspecifically bound: only the protein bands remain colored. A molecular weight marker is generally loaded as reference to identify a protein of a known size (Fig. 3.5).

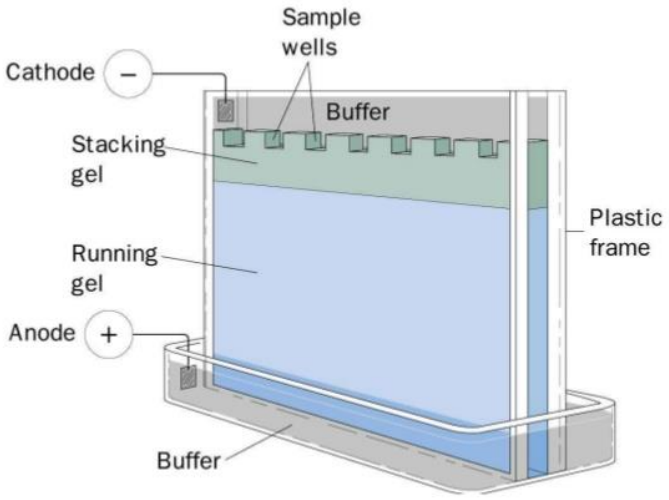


Figure 3.5 The typical apparatus used for an SDS-PAGE.

3.5 Ligand binding assays

3.5.1 Absorption and fluorescence spectra of Retinoid-CRBP complex

To establish whether a retinoid interacts specifically with the CRBP binding site, absorption and fluorescence spectra of the retinoid-CRBP complex can be analyzed.

As described in the introduction (Par. 1.4.4), retinol bound to CRBP shows a characteristic absorption spectrum, with a maximum at 348-350nm that is clearly distinct from the absorption spectrum of the free retinol, which exhibit a maximum at 323nm. This shift is a clear signal of the complex formation.

Due to fluorescence energy transfer between the protein and the ligand, to investigate the binding of retinoid to CRBPs fluorescence titration can be carried out by following either the quenching of protein fluorescence or the enhancement of the bound retinoid fluorescence. It is more convenient to monitor the ligand interaction following the quenching of intrinsic protein fluorescence, because the titration method based on the enhancement of retinoid fluorescence is remarkably less sensitive and therefore less suitable for an accurate analysis of binding proprieties.

The fluorescence energy transfer phenomenon in CRBPs is due to the presence of Trp106 into the protein binding cavity (Fig. 3.6). When the ligand is bound, the exciting of the protein at the tryptophan maximum absorption wavelength, 280nm, gives a fluorescence emission at 345nm, at

which retinol absorbs; if retinol is bound into the cavity, part of the emission energy is transferred from Trp106 to retinoid accommodated into the binding cavity.

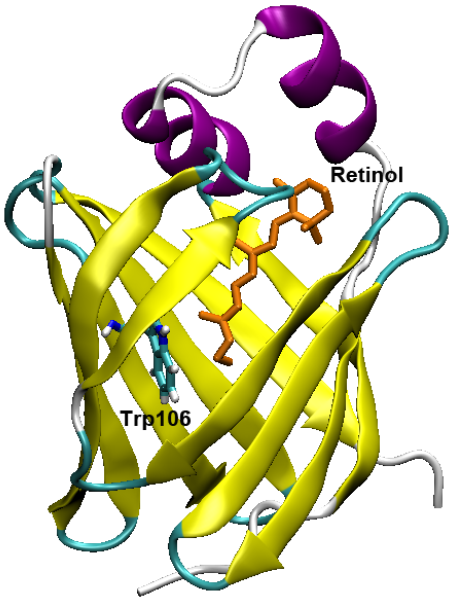


Figure 3.6 Cartoon representation of a generic retinol-CRBP complex. Trp106 (colored by atom type) in proximity of retinol (orange) are shown in stick.

3.5.2 Analysis of Fluorescence Titration data

CRBPs possess one single binding site for retinoids. According to equilibrium expression:

$$CRBP + RTL \leftrightarrow CRBP \bullet RTL \quad (1)$$

the mass low equation for the binding of retinol (RTL) is:

$$K_d = [CRBP][RTL]/[CRBP \bullet RTL] \quad (2)$$

where K_d is the dissociation constant, $[CRBP]$ and $[RTL]$ are the concentrations of free protein and free ligand in solution, respectively, and $[CRBP \bullet RTL]$ is the concentration of the CRBP-ligand complex. $[CRBP]$ and $[RTL]$ present in solution can be expressed as:

$$[CRBP] = [CRBP]_0 - [CRBP \bullet RTL] \quad (3)$$

$$[RTL] = [RTL]_0 - [CRBP \bullet RTL] \quad (4)$$

where $[CRBP]_0$ and $[RTL]_0$ are the total protein and ligand concentrations, respectively. The fraction of CRBP bound to RTL is:

$$\alpha = [CRBP \bullet RTL]/[CRBP]_0 \qquad (5)$$

By substituting (3) and (4) in equation (2) and introducing α as defined by (5), the following equation can be obtained from equation (2):

$$\alpha^2[CRBP]_0 - \alpha([RTL]_0 + [CRBP]_0 + K_d) + [RTL]_0 = 0$$

By taking the negative square root, the fraction of protein bound can be expressed as:

$$\alpha = \frac{[RTL]_0 + [CRBP]_0 + K_d - \sqrt{([RTL]_0 + [CRBP]_0 + K_d)^2 - 4[RTL]_0[CRBP]_0}}{2[CRBP]_0} \qquad (7)$$

Recording the fluorescence intensity at different RTL concentration, the values of α can be calculated using the relationship:

$$\alpha = (F_0 - F)/(F_0 - F_\infty)$$

where F_0 and F_∞ represent the two limit values of fluorescence intensity in the absence and in the presence of saturating RTL concentration, respectively, that can be extrapolated from the titration curve, and F represents the fluorescence intensity at a certain concentration of RTL. The experimentally determined values of α are fitted with equation (7) by means of non-linear least squares regression.

Equations (6) and (7) may be written in the form of equations (9) and (10):

$$\alpha^2 n[CRBP]_0 - \alpha([RTL]_0 + n[CRBP]_0 + K_d) + [RTL]_0 = 0 \qquad (9)$$

$$\alpha = \frac{[RTL]_0 + n[CRBP]_0 + K_d - \sqrt{([RTL]_0 + n[CRBP]_0 + K_d)^2 - 4n[RTL]_0[CRBP]_0}}{2n[CRBP]_0} \qquad (10)$$

to include n, the number independent binding sites present on the protein molecule. Thus, the fitting of the experimentally determined values of α with equation (10) allows the determination of both ligand-binding affinity and stoichiometry (Malpeli, Stoppini, Zapponi, Folli, & Berni, 1995).

To evaluate binding affinities by means of fluorescence, the protein concentration in the system should not be substantially greater than the dissociation constant.

3.6 Crystallization method

The goal of crystallization is to produce a well-ordered crystal that is lacking in contaminants, while it is still large enough to provide a diffraction pattern when exposed to X-rays. This diffraction pattern can, then, be analyzed to discern the protein's tertiary structure. Protein crystallization is inherently difficult because of the fragile nature of protein crystals. Proteins have irregularly shaped surfaces, which results in the formation of large channels within any protein crystal. Therefore, the noncovalent bonds that hold together the lattice must often be formed by several layers of solvent molecules. In addition, to overcoming the intrinsic fragility of protein crystals, a number of environmental factors must also be overcome. Due to the molecular differences between individual proteins, conditions unique to each protein must be obtained for a successful crystallization.

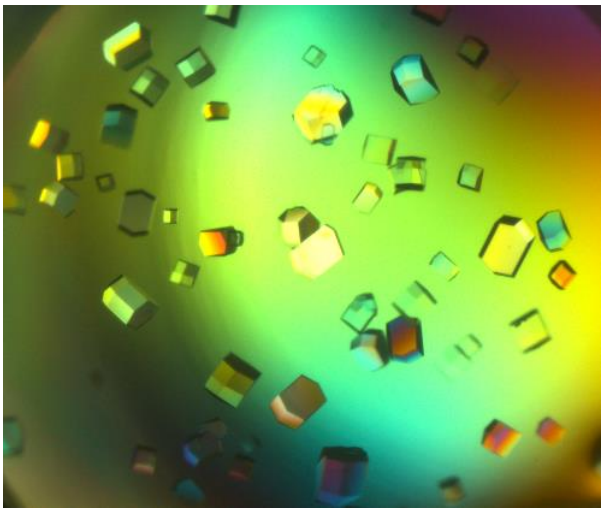


Figure 3.7 Protein crystals.

3.6.1 The hanging drop vapor diffusion crystallization

The hanging drop vapor diffusion technique (Fig. 3.8) is a popular method for the crystallization of macromolecules. The principle of vapor diffusion is straightforward. A drop composed of a mixture of sample and reagent is placed in vapor equilibrium with a liquid reservoir of reagent. Typically, the drop contains a lower concentration of reagent than the reservoir. To achieve equilibrium, water vapor leaves the drop and eventually ends up in the reservoir. As water leaves the drop, the sample undergoes an increase in relative supersaturation. Both the sample and reagent increase in concentration as water leaves the drop for the reservoir. Equilibration is reached when the reagent concentration in the drop is approximately the same as that in the reservoir.

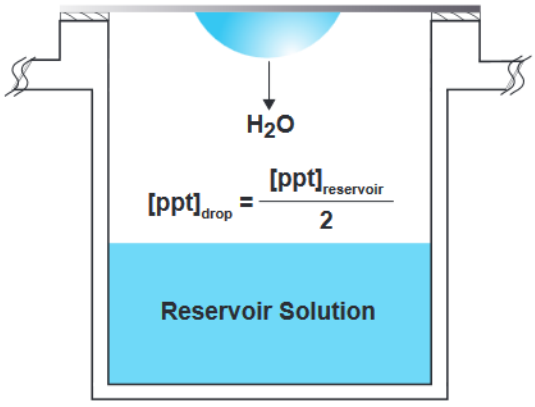


Figure 3.8 Process of vapor diffusion.

The hanging drop crystallization method presents several advantages:

- it can be cost effective;
- sample and reagents are in contact with a siliconized glass surface, ensuring the easy access to crystals;
- we can perform multiple drops (experiments) with a single reservoir.

3.7 Mass Spectrometry

Mass spectrometry is a sensitive technique used to detect, identify and quantify molecules, basing on their mass to charge ratio (*m/z*). All mass spectrometers have an ion source, a mass analyzer and an ion detector. The nature of these components varies depending on the type of mass

spectrometer, the type of data required, and the physical properties of the sample. Samples are loaded into the mass spectrometer in liquid, gas or dried form and then vaporized and ionized by the ion source. The charge that these molecules receive allows the mass spectrometer to accelerate the ions throughout the system. Basically, the ions pass first through a velocity selector, in which an electric field *E* is orthogonal to a magnetic field *B* and to the velocity of particles. In this condition, the velocity depends from the intensity of the two fields: $v = E/B$. Subsequently, a magnetic field again orthogonal to the selected velocity deflects the paths of individual ions. The radius of curvature of the deflected path is related to the mass-to-charge ratio of the particle by the equation $r = mv/zB$, where *v* is the selected velocity of the particle and *B* is the intensity of the magnetic field.

Commonly used mass analyzers include time-of-flight [TOF] methods, orbitraps, quadrupoles and ion traps, and each type has specific characteristics. Mass analyzers can be used to separate all analytes in a sample for global analysis.

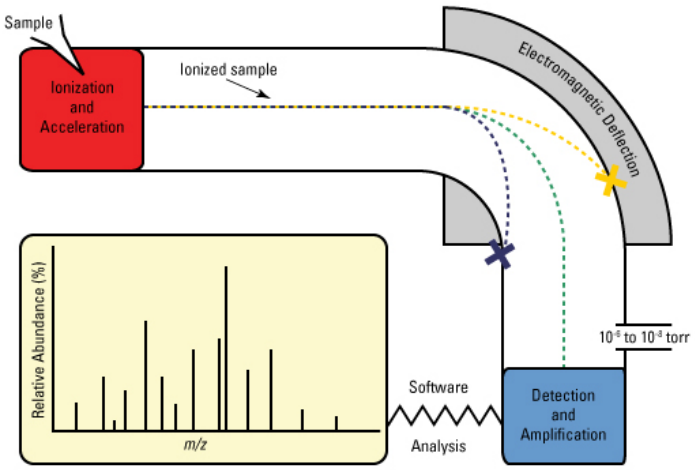


Figure 3.9 Mass Spectrometry scheme.

Different ionization sources permit diverse analyses; in our case a Matrix-assisted laser desorption/ionization coupled with a Time-of-flight analyzer and an Electrospray source coupled with the Orbitrap analyzer, available to the CIM (Centro Interdipartimentale Misure) of University of Parma, were used to determine the mass and the empirical formula of an unknown ligand bound to CRBP4.

3.7.1 MALDI-TOF

Matrix-assisted laser desorption/ionization (MALDI) mass spectrometry has become a widespread analytical tool for peptides, proteins and most other biomolecules. In MALDI analysis, the analyte is first dried with a large molar excess of a matrix compound, usually an organic acid (in our case α-Cyano-4-hydroxycinnamic acid = HCCA); then, a laser radiation incident on this analyte-matrix mixture causes the vaporization of the matrix which carries the analyte with it (Fig. 3.10). The matrix therefore plays a key role by strongly absorbing the laser light energy that causes first its vaporization and then the vaporization of analyte by the release of the excess energy. The matrix also serves as a proton donor and acceptor, ionizing the analyte with positively or negatively charge. The efficient and directed energy transfer during a matrix-assisted laser-induced desorption event provides high ions amount of the intact analyte and allows for the measurement of compounds with high accuracy and sub-picomole sensitivity.

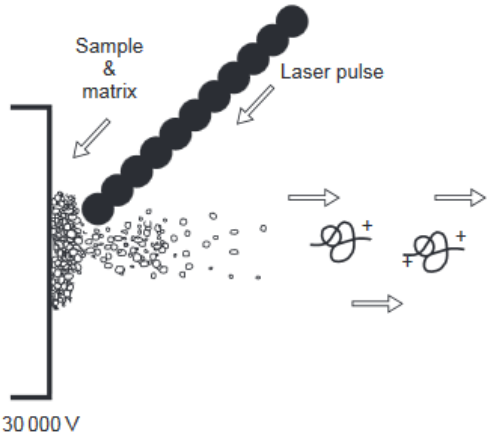


Figure 3.10 MALDI technique.

The MALDI system is coupled with the linear time-of-flight (TOF) mass analyzer, based on the acceleration of a set of ions towards a detector. Because the ions have the same kinetic energy, but a different mass, the ions reach the detector at different times. The smaller ions reach the detector first because of their greater velocity while the larger ions take longer owing to their larger mass. Hence, the analyzer is called TOF because the mass is determined from the ions' time of flight. The arrival time at the detector is dependent upon the mass, the charge, and the kinetic energy of the ion.

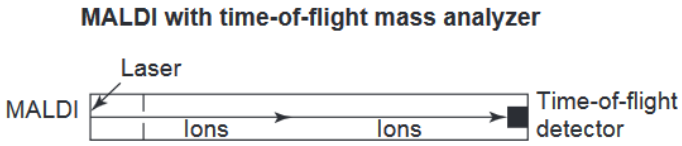


Figure 3.11 MALDI with TOF analyzer.

This accurate technique can be extremely useful for protein identification and characterization. For example, a protein can often be unambiguously identified by the accurate mass analysis of its constituent peptides. In our case the MALDI analysis was used to characterize the compound fortuitously bound to CRBP4.

3.7.2 Electrospray-Orbitrap

The Electrospray ionization (ESI) mass spectrometry uses electrical energy to assist the transfer of ions from solution into the gaseous phase before they are subjected to a mass spectrometric analysis. Ionic species in solution can thus be analyzed by ESI-MS with increased sensitivity. Neutral compounds can also be converted to ionic form in solution or in gaseous phase and hence can be studied by ESI-MS. The transfer of ionic species from solution into the gas phase by ESI involves three steps: (1) dispersal of a fine spray of charge droplets from a capillary subjected to a high voltage, followed by (2) solvent evaporation and (3) ion ejection from the highly charged droplets (Fig. 3.12).

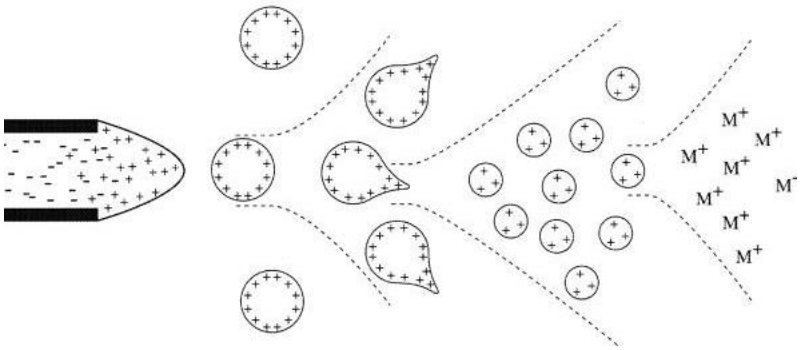


Figure 3.12 Mechanism of electrospray ionization.

The application of a gas (e.g. nitrogen), which shears around the eluted sample solution, enhances the sample flow rate. The charged droplets are continuously reduced in size by evaporation of the solvent, leading to an increase of surface charge density and a decrease of the droplet radius. Finally, the electric field strength within the charged droplet reaches a critical point at which it is kinetically and energetically possible for ions at the surface of the droplets to be ejected into the gaseous phase. The emitted ions are sampled by a sampling skimmer cone and are then accelerated into the mass analyzer for subsequent analysis of molecular mass and measurement of ion intensity.

In our case the ESI source was coupled with the Orbitrap mass analyzer, an ion trap mass analyzer consisting of an outer barrel-like electrode and a coaxial inner spindle-like electrode that traps ions in an orbital motion around the spindle. In an orbitrap, ions are trapped because their electrostatic attraction to the inner electrode is balanced by their inertia. Thus, ions cycle around the inner electrode on elliptical trajectories. In addition, the ions also move back and forth along the axis of the central electrode thus their trajectories in space resemble helices. The ions' motion is completely independent not only of motion around the inner electrode but also of all initial parameters of ions except their mass-to-charge ratios m/z.

COMPUTATIONAL TECHNIQUES

4.1 The aims of Molecular dynamics

One of the principal tools in the theoretical study of biological molecules is the method of molecular dynamics simulations (MD). This computational technique calculates the time dependent behavior of a molecular system, providing detailed information on the fluctuations and conformational changes of proteins and nucleic acids. This technique permits, by means of computer simulations, to understand the proprieties of biological macromolecules, in terms of their structure and dynamics, and the microscopic interactions between them. Molecular dynamics is of help to conventional experiments, acting in synergy with them, and giving information that cannot be achieved by experimental techniques. Computer simulations act as a "bridge" between microscopic length, time scale and energy and the macroscopic world of the laboratory.

In this work MD simulations have been useful to investigate the structural determinants and the solvent behavior of different proteins, studying their flexibility and their ligand up-take mechanism. The mechanism of access and release of one or more molecule to a protein is a process very difficult to observe with conventional experiments - neither NMR spectroscopy - because the time of these events is too short. MD simulations permit to suggest, at atomic interaction level, a model for the ligand up-take that subsequently can be validated experimentally.

The limits of this computational approach regard mainly the complexity of the systems, in spite of the great progress in terms of computing power, and the use of empirical potential function approximations. In addition, biological molecules exhibit a wide range of time scale over which specific processes occur; for example:

- Local Motions (0.01 to 5 Å, 10 fs to 100 ms)
  - Atomic fluctuations
  - Sidechain Motions
  - Loop Motions
- Rigid Body Motions (1 to 10Å, 1ns to 1s)
  - Helix Motions
  - Domain Motions (hinge bending)
  - Subunit motions

- Large-Scale Motions (> 5Å, 100ns to 10^4 s)
  - Helix coil transitions
  - Dissociation/Association
  - Folding and Unfolding

However to limit the computational load is useful an accurate knowledge of the system to reduce the number of degrees of freedom, accelerating the calculation; for examples it is possible constrain part of the protein or limit the motion to the only side chains.

4.2 Molecular dynamics simulations

Molecular dynamics simulations consist of numerical step-by-step solution of the classical equations of motion, which for a simple atomic system may be written

$$\mathbf{F}_i = \mathbf{a}_i m \qquad \mathbf{F}_i = -\frac{dE}{d\boldsymbol{r}_i} = -\nabla_i E = m_i \frac{d^2\boldsymbol{r}_i}{dt^2}$$

In this way we are able to calculate the resultant force $\mathbf{F}_i$ acting on each atom, knowing the potential energy (E). Newton's equations of motion can then relate the gradient of the potential energy to the changes in position as a function of time, which can be calculated by a double integration. Therefore, to calculate a trajectory, we only need the initial positions of the atoms, available by the experimentally known structures; an initial distribution of velocities, assigned randomly from a Maxwell-Boltzmann distribution at a given temperature; and the accelerations, which are determined by the gradient of the potential energy function.

4.2.1 Some useful basic principles of statistical mechanics

Molecular dynamics simulations generate information at the microscopic level, as atomic positions and velocities. The conversion of this microscopic information to macroscopic observables such as pressure, energy, heat capacities, etc., requires statistical mechanics, that provides the rigorous mathematical expressions that relate macroscopic properties to the distribution and motion of the atoms and molecules of our N-body system. The goal is to understand and to predict macroscopic phenomena from the properties of individual molecules or atoms making up the system. In order

to connect the macroscopic system to the microscopic system, time independent statistical averages are often introduced. We first recall a few definitions.

The *thermodynamic state* of a system is usually defined by a small set of parameters, for example, the temperature, T, the pressure, P, and the number of particles, N. Other thermodynamic properties may be derived from the equations of state and other fundamental thermodynamic equations.

The *mechanical* or *microscopic state* of a system is defined by the atomic positions, **r**, and momenta, **p**; these can also be considered as coordinates in a multidimensional space called *phase space*. For a system of N particles, this space has 6N dimensions. A single point in phase space, denoted by G, describes the state of the system. An *ensemble* is a collection of points in phase space (i.e. a collection of systems which have different microscopic states) all satisfying the conditions of a particular thermodynamic state. From this point of view, a molecular dynamics simulation generates a sequence of points in phase space as a function of time; these points belong to the same *ensemble*, and they correspond to the different conformations of the system, i.e. the particle positions, and their respective momenta.

There exist different ensembles depending on the thermodynamics state, in particular:

- *Microcanonical ensemble* (NVE): the thermodynamic state is characterized by a fixed number of atoms, N, a fixed volume, V, and a fixed energy, E. It corresponds to an isolated system.
- *Canonical Ensemble* (NVT): the thermodynamic state is characterized by a fixed number of atoms, N, a fixed volume, V, and a fixed temperature, T.
- *Isobaric-Isothermal Ensemble* (NPT): this ensemble is characterized by a fixed number of atoms, N, a fixed pressure, P, and a fixed temperature, T.
- *Grand canonical Ensemble* (μVT): the thermodynamic state for this ensemble is characterized by a fixed chemical potential, μ, a fixed volume, V, and a fixed temperature, T.

An experiment is usually made on a macroscopic sample that contains an extremely large number of atoms or molecules sampling an enormous number of conformations. In statistical mechanics, averages corresponding to experimental observables are defined in terms of ensemble averages; an ensemble average is average taken over a large number of replicas of the system considered simultaneously. The ensemble average is given by

$$\langle A \rangle_{ensemble\ (p,r)} = \iint dp\ dr\ A_{(p,r)}\ \rho_{(p,r)}$$

Where $A_{(p,\ r)}$ is the experimental observable expressed in the function of **p** and **r** of the system and the integration is over all the possible **p** and **r** variables. ρ(p, r) is the probability density that the system is in the microstate characterized by those **p** and **r** variables.

A trajectory calculated in a molecular dynamics simulation performed under defined macroscopical (thermodynamical) conditions, i.e. in a chosen statistical ensemble, represents the evolution in time of a microscopic state, which is the distribution in a temporal sequence of the points that constitute the ensemble. Each microscopic state corresponds to a conformation, with defined positions and momenta. All the conformations belonging to the same ensemble (= the trajectory) correspond to the same constant macroscopic variable. Therefore, to calculate an ensemble average, the molecular dynamics simulations must pass through all possible states corresponding to the particular thermodynamic constraints. Another way, as done in an MD simulation, is to determine a time average of A. The experimental observables are therefore the spatial average, while by molecular dynamics simulation time averages can be calculated.

This leads us to one of the most fundamental axioms of statistical mechanics, the ergodic hypothesis, which states that the time average equals the ensemble average:

$$\langle A \rangle_{ensemble} = \langle A \rangle_{time}$$

In theory, if the system is allowed to evolve in time indefinitely, it will pass through all possible states. Therefore, a molecular dynamics simulation should generate enough representative conformations such that this equality is satisfied. If this is the case, experimentally relevant information concerning structural, dynamic and thermodynamic properties may then be calculated. Because the simulations are of fixed duration, one must be certain to sample a sufficient amount of phase space.

4.2.2 The potential energy function

To each conformation of a system is associated a potential energy value.

The potential energy function is formed by the sum of two terms: non-bonded interactions energy ($E_{non\text{-}bonded}$) and bonding potential (E_{bonded}).

$$E(\{r\} = E_{non-bonded} + E_{bonded}$$

The part of the potential energy $E_{non\text{-}bonded}$ represents non-bonded interaction between atoms due substantially to *van der Waals* and electrostatic interactions. The first interaction type is described by the Lennard-Jones (LJ) potential:

$$E_{vdW} = \sum_{nonbonded\ pairs} 4_{\varepsilon_{ij}} \left[\left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}} \right)^{6} \right]$$

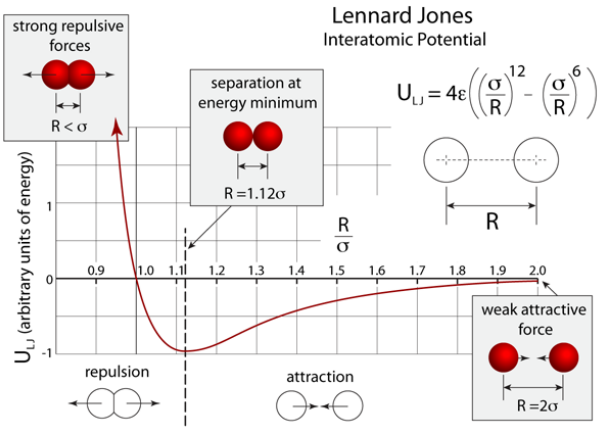


Figure 4.1 Lennard-Jones potential pair potential showing the r^{-12}, r^{-6} contribution.

The second interaction type is described by the Coulomb potential, useful when net charges are present.

$$E_{electrostatic} = \sum_{nonbonded\ pairs} \frac{q_i q_j}{4\pi\varepsilon_0 \varepsilon r_{ij}}$$

Some non-bonded interactions are eliminated with the aim to make the calculation faster. The LJ potential is by default shifted to zero by subtracting the value at the cut-off; while the Culomb potential is cut-off by the Particle Mesh Ewald (PME) methods which take advantage of the Fast Fourier Transform, an algorithm optimized to calculate in a fast way the Fourier Transform and its inverse transform. This method replaces the summation in the real space with an equivalent Fourier summation which converged faster than in the real space. In addition, in the Fourier space

the charge is considered as a density of charge and not as a single point charge, resolving in this way the problem of an irregular distribution of charge.

For molecules is needed to consider the intramolecular bonding interaction, defined by the bonding potential (E_{bonded}). The simplest molecular model includes terms of the following kind:

$$E_{bonded} = \sum_{1,2\ pairs} K_b(r - r_0)^2 + \sum_{angles} K_\theta(\theta - \theta_0)^2$$
$$+ \sum_{dihedrals} K_\emptyset[1 + \cos(n\emptyset - \delta)] + \sum_{\text{impropers}} \mathrm{K}_\varphi(\varphi - \varphi_0)^2$$

The first terms is relative to the bond stretching, the second regard the bending, both armonic potentials, and to follow are reported the torsions along the bond (dihedral angles and improper angles); these energy are due to the geometry of the molecule that we can see in figure 4.2.

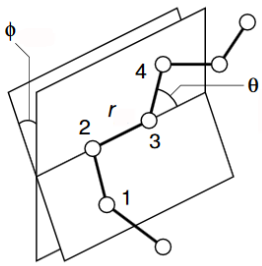


Figure 4.2 Geometry of a simple chain molecule, illustrating the definition of interatomic distance r, bend angle θ and torsion angle ϕ.

The potential energy equation contains numerical parameters which are used to define different atoms relative to their particular bonding situation.

4.2.3 The force field and the integration algorithm

The computational complexity is simplified by empirical potential, named *force-field*, less computational-consuming which permit to unload the calculation using some approximations. A force filed includes the potential energy function and a set of parameters derived from physical or

chemical experiments, quantum mechanics calculation, or both. For this study we have used the Gromos ffG53a6 *force-field* (Oostenbrink, Villa, Mark, & Van Gunsteren, 2004).

Using GROMACS, the *leap-frog* integrator algorithm for the integration of the equation of motion is proposed by default. Using this algorithm, the velocities (v) are not calculated in the same time of the positions (x) as is clearly shown in figure 4.3.

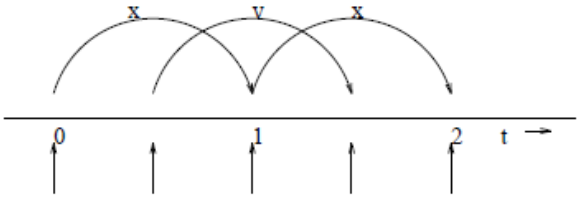


Figure 4.3 The Leap-Frog integration method.

The *leap-frog* algorithm uses the following relationships:

(a) $$\boldsymbol{r}(t+\delta t) = \boldsymbol{r}(t) + \delta t \boldsymbol{v}\left(t+\frac{1}{2}\delta t\right)$$

(b) $$\boldsymbol{v}\left(t+\frac{1}{2}\delta t\right) = \boldsymbol{v}\left(t-\frac{1}{2}\delta t\right) + \delta t \boldsymbol{a}(t)$$

To implement the *leap-frog* algorithm, the velocities $\boldsymbol{v}\left(t+\frac{1}{2}t\right)$ are first calculated from the velocities at time $\left(t-\frac{1}{2}\delta t\right)$ and the acceleration time t. The positions $\boldsymbol{r}(t+\delta t)$ are then deduced from the velocities just calculated together with the position at time $\boldsymbol{r}(t)$ using equation (a). The velocities at time t can be calculated from:

(c) $$\boldsymbol{v}(t) = \frac{1}{2}\left[\boldsymbol{v}\left(t+\frac{1}{2}\delta t\right) + \boldsymbol{v}\left(t-\frac{1}{2}\delta t\right)\right]$$

The velocities thus *'leap-frog'* over the positions to give their values at $t+\frac{1}{2}\delta t$. The positions then leap over the velocities to give their new values at $t+\delta t$, ready for the velocities at $t+\frac{3}{2}\delta t$, and so on.

4.2.4 The solvation and the periodic boundary conditions

In all biological systems, water is essential and influences their structure, dynamics and thermodynamics. One of the most important effects is the shielding of electrostatic interactions, but it is also important for the formation of hydrogen bonds between water molecules and polar solute atoms. For all these reasons is essential to well consider the presence of the water in MD simulations. In our work, we have chosen to use explicit water molecules, surrounding our solute macromolecule by a water box of defined dimensions. This method considerably increases the calculation load but is the most reliable and in our case is essential to simulate the up-take mechanism, because some interaction of ligand or protein with water could be of fundamental importance. Periodic boundary conditions (PBC) are used in molecular dynamics simulations to minimize edge effects caused by finite size, and make the system more like an infinite one. The using of periodic boundary conditions implies that particles to be simulated are enclosed in a space-filling box, which is surrounded by translated copies of itself (in three dimensions, each box would have 26 nearest neighbours). Thus there are no boundaries of the system, and the particles in the box experience forces as if they were in bulk fluid. All the image particles move solidary with their original particle from the simulated box, therefore only the central box is effectively simulated. When a particle enters or leaves the simulation region, an image particle leaves or enters this region, such that the number of particles from the simulation region is always conserved. Particle *i,* in the simulation box, appears to be interacting not only with the other particles in the box, but also with their images, increase enormously the interacting pairs. In this case is used both a potential with a finite range (figure 4.4) - by which the interaction of two particles separated by a distance exceeding a cut-off distance can be ignored - and the minimum image convention, by which each atom “sees” at most just one image of every other atom in the system. The energy (and force) is calculated with the closest atom or image. When a cutoff is employed, the interactions between all pair of atoms that are further apart than the cutoff value are set to zero, taking into account the closest image. When periodic boundary conditions are being used, the cutoff should not be so large that a particle sees its own image or the same molecule twice. This has the effect to limiting the cutoff radius, that must not exceed half the shortest box vector.

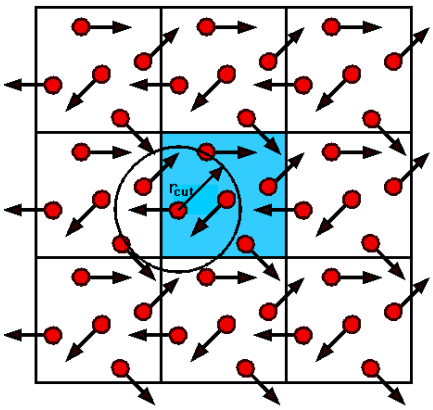


Figure 4.4 Periodic boundary conditions in two dimensions. In the figure r_{cut} is the cutoff radius that is normally applied when calculating the force between two atoms.

4.2.5 Potential function approximations

The empirical potential function has several limitations, which result in inaccuracies in the calculated potential energy.

One limitation is due to the fixed set of atom types employed when determining the parameters for the force field. Atom types are used to define an atom in a particular bonding situation; for example, an aliphatic carbon atom in a sp3 bonding situation has different properties than a carbon atom found in an aromatic ring. Instead of presenting each atom in the molecule as a unique one described by a unique set of parameters, there is a certain clustering in order to minimize the number of atom types. This can lead to type-specific errors. The properties of certain atoms, like aliphatic carbon or hydrogen atoms, are less sensitive to their surroundings and a single set of parameters may work quite well, while other atoms like oxygen and nitrogen are much more influenced by their neighboring atoms. These atoms require more types and parameters to account for the different bonding environments.

An approximation introduced to decrease the computational demand is the pair-wise additive approximation: the interaction energy between one atom and the rest of the system is calculated as a sum of pair-wise interactions; therefore the pair of atoms does not see the other atoms in the system. The simultaneous interaction between three or more atoms is not calculated, so certain polarization effects are not explicitly included in the force field. This can lead to subtle differences between calculated and experimental results.

Another important point to take into account is that the potential energy function does not include entropic effects. Thus, a minimum value of E calculated as a sum of potential functions does not necessarily correspond to the equilibrium, or the most probable state. Because of the fact that experiments are generally carried out under isothermal-isobaric conditions (constant pressure, constant system size and constant temperature) the equilibrium state corresponds to the minimum of Gibb's Free Energy, G. A molecular dynamics simulation preferentially samples the lower-energy regions of phase space (that are the most probable), while a ergodic trajectory would visit also the high energy regions, that will never be adequately sampled. Therefore, the results for the free energy and other entropic properties will be inaccurate.

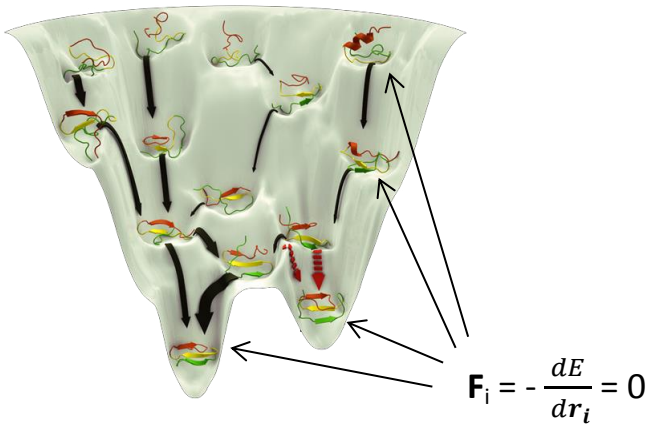
4.3 Energy minimization

The energy minimization is a necessary operation before performing a molecular dynamics simulation, with the aim to find an energetically stable state. This is important to avoid structural distortions in the molecule due to the presence of unbalanced forces. Energy minimization is a numerical procedure for finding a minimum on the potential energy surface starting from a higher energy structure. Given a potential energy function, is possible to find a minimum as the point in which the derivative of the potential energy is zero; as consequence of this, also the resultant force on each atom will be zero.

During energy minimization, the geometry is changed in a stepwise fashion so that the energy of the molecule is reduced step-by-step; after a number of steps, a local or global minimum on the potential energy surface is reached.

Several algorithm are used to minimize the energy, some of them are useful when the starting structure has a high energy and is far from the minimum (e.g. simplex, steeepest discent...) other of them are used to optimize structures near the minimum (e.g. conjugate gradients, Newton-Raphson...). Usually it is better to use more than one minimization algorithm.

Figure 4.5 Energy graph for a protein.



4.4 Standard simulation protocol in short

1. Choose the starting coordinates: from a crystal or NMR structure, from a model.
2. Energy minimization of the protein.
3. Water box generation.
4. Ions addition to neutralize the total charge of the system, and to simulate the presence of a ionic force (in general Na^+ and Cl^- ions are added).
5. Energy minimization of the protein-solvent system.
6. *Position restrained* molecular dynamics, to relax the solvent around the protein.
7. Full motion dynamics: all the atoms of the system are free to move. This step can be separated in two phase. a) Equilibration phase: after initial changes, the system usually reaches an equilibrium state, i. e. the values of a set of monitored properties (energy, temperature, pressure, and also structural properties) become stable. Sometimes this is preceded by a heating phase, in which the temperature is slowly increased to the desired value, periodically assigning new velocities. b) Production phase: in this phase, by averaging over an equilibrium trajectory, many macroscopic properties can be calculated.
8. At regular intervals the configuration of the system (the atomic coordinates) is output to a file, which can be analyzed. Proprieties not calculated during the simulation are determined and the configurations are examined.

MATERIALS AND METHODS

5.1 Molecular dynamics simulations

All MD simulations were prepared and run by GROMACS 4.5.5 software (Van Der Spoel et al., 2005), using the *force-filed* Gromos G53a6 (Oostenbrink, Villa, Mark, & Van Gunsteren, 2004) and employing the intensive numerical computing service of the Department of Physics (grid-ui.fis.unipr.it).

5.1.1 Studies on the differences between the four isoforms

We have performed two different kinds of simulations: on the apo-protein and in presence of the ligand. The simulation parameters for retinol were already contained in the Gromos G53a6 force field.

We used as starting coordinates the pdb structures of apo-CRBPs available on Protein Data Bank (PDB ID: 2RCQ for hCRBP2, 1GGL for hCRBP3, 1LPJ for hCRBP4), with the exception of hCRBP1, for which we used the coordinate obtained in our laboratory by X-ray crystal diffraction. For each structure we manually removed water and all other molecules from the pdb file. For hCRBP3, the pdb structure file contains two monomers and we considered only the A chain.

For the simulations in the presence of retinol, we used the same apo-protein starting structures to which we added one or more retinol molecules, by means of the VMD graphical software. Initially, we tried to add four retinol molecules to hCRBP2, in agreement whit a work of Tsfadia and co-workers, in which 4 palmitate anions were added to the toad liver basic Fatty acid-binding protein to facilitate the encounter between the substrate and the protein (Tsfadia et al., 2007). We positioned three retinol molecules close to the anti-portal region (Levin, Nachliel, Gutman, & Tsfadia, 2009) and one near the portal region (Fig 5.1). However, since the presence of more than one retinol molecule didn't seem to condition the entrance of the ligand, we then performed all the subsequent simulations keeping only one retinol molecule. Considering the lipophilic nature of the ligand, we placed it near to hydrophobic protein regions, as shown in figure 5.2 A, B and C for hCRBP1, 3 and 4, respectively.

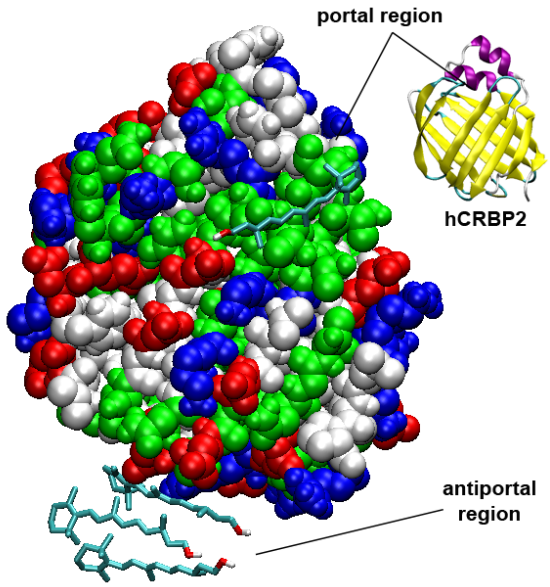


Figure 5.1 Van der Waals representation of hCRBP2, colored by residue type (polar residues in green, positively in blue, negatively in red, hydrophobic in white), in the orientation shown in the little upper right figure, where hCRBP2 is represented in cartoon and colored by secondary structure. The four molecules of retinol are represented in sticks and colored by atom type: three molecules were positioned at the anti-portal region and one molecule at the portal region.

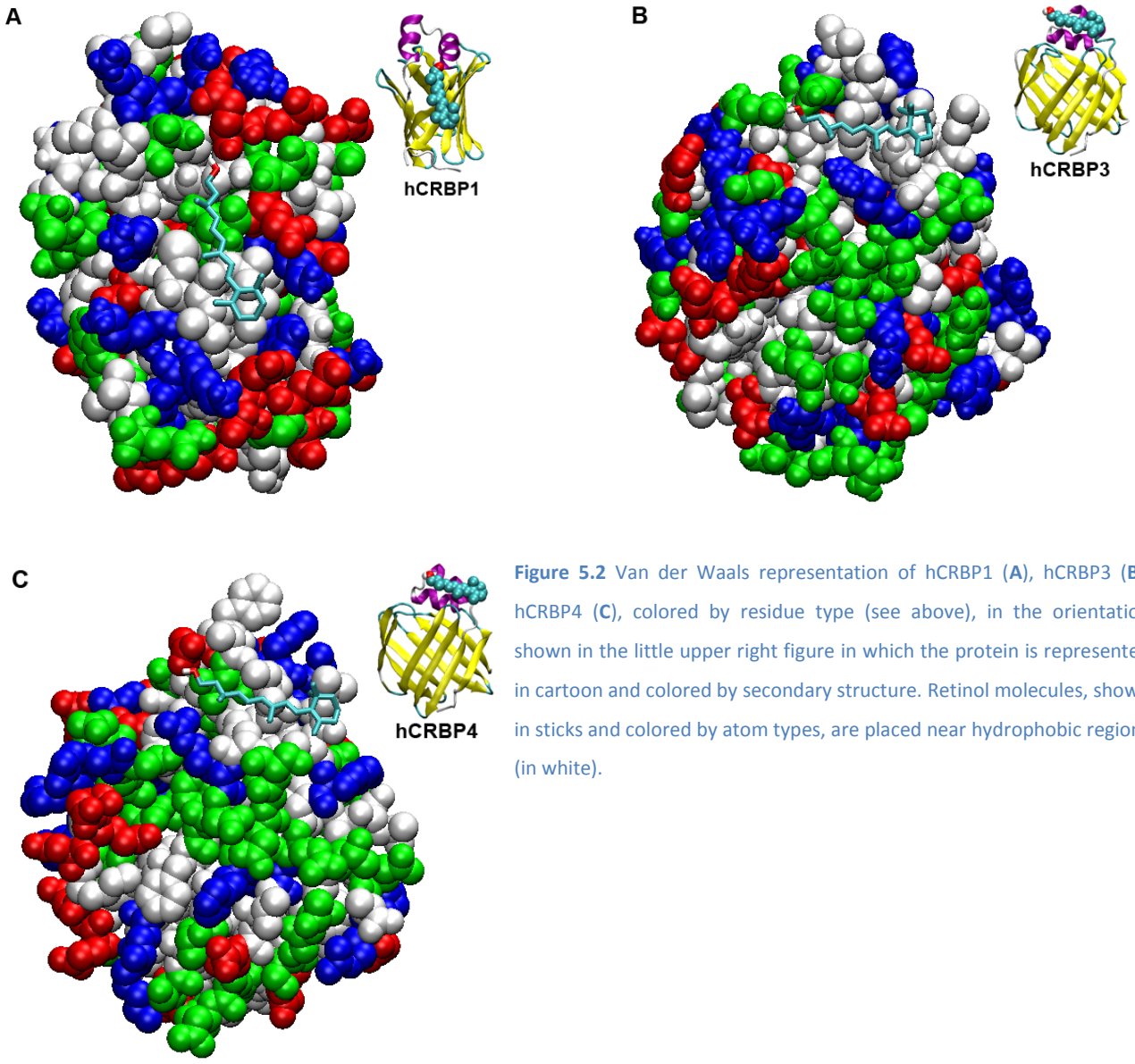


Figure 5.2 Van der Waals representation of hCRBP1 (**A**), hCRBP3 (**B**), hCRBP4 (**C**), colored by residue type (see above), in the orientation shown in the little upper right figure in which the protein is represented in cartoon and colored by secondary structure. Retinol molecules, shown in sticks and colored by atom types, are placed near hydrophobic regions (in white).

Since cellular retinol binding proteins are cytoplasmic proteins we performed all simulation embedding proteins in water solvent, adding enough Na^+ Cl^- ions to neutralize the system and to reach a physiological ions concentration of 0.1 M. Using GROMACS subroutines, for each kind of simulation (both in the presence and in the absence of ligand) we inserted the protein in a water box, keeping a water layer 1 nm thick around the macromolecule. During the simulations, the periodic boundary conditions were applied to the system.

For each system, internal constraints were relaxed by an energy minimization of the whole system using the *steepest descent* algorithm. Then, a short position restrained MD 50 ps long, was performed to relax solvent molecules around the protein.

For the apo-proteins simulations, a full MD run was carried out for 100 ns at 300K and 1 atm with an integration step of 2 fs.

For the simulations in the presence of ligand, as a first attempt we added only one retinol molecule to the hCRBP2, positioning the ligand in five different starting points and performing five MD simulations 200 ns long, at room temperature and 1 atm. In all simulations the ligand could not enter correctly or completely into the protein. Arising from FABP MD simulation previously mentioned (Levin et al., 2009), we have performed other six simulations adding four retinol molecules to hCRBP2 to facilitate the encounter between ligand and protein; each simulation was carried out for 40 ns, 300K and 1 atm. Summing to previously attempts, it corresponds to a total of 1240 ns of simulation. Neither in this case, we obtained a realistic ligand entry. For this reason, we decided to increase the temperature to 350K with the aim to supply sufficient energy to the system to overcome possible potential barriers which could trap the ligand in non-specific positions (Hu & Liu, 2014). Several MD simulations, for each isoform, were performed at 350 K and 1 atm for the duration of 40 ns using only one retinol molecule and as starting coordinates those shown in figures 5.1 (with the retinol molecule positioned at the portal region) and 5.2.

5.1.2 Study of the effect of the His-Tag on the protein conformation

Molecular dynamics simulations were performed by adding a His_6-Tag to the WT apo CRBP1 crystal structure, by means of the Swiss-PDB Viewer program (Guex & Peitsch, 1997).

To better reproduce the acidic crystallization conditions (pH 5.5) of HT human CRBP1(Silvaroli et al., 2016), the pKa values of ionizable residues were calculated by means of the program PROPKA, embedded in the software package PDB2PQR (Dolinsky, Nielsen, McCammon, & Baker, 2004)). Based on PROPKA results, a first MD simulation was carried out on the His-tagged protein, keeping all His residues in a protonated form. Subsequently, to discriminate the effects of the charges of protonated His-residues, a second MD simulation was performed with all His residues in their neutral form. Moreover, the MD simulation performed on WT apo-CRBP1, previously described, was used as control.

These MD simulations were carried out in the same solvation conditions previously described, applying the pbc. The whole system was subjected to an energy minimization and then a position restrained MD, lasting 50 ps, was performed to relax the solvent around the protein. Finally a 100 ns long full MD run was carried out, at 300 K and 1 atm.

5.1.3 Programs used for the data analysis

- Gromacs subroutines
- Swiss-PdbViewer (Guex & Peitsch, 1997)
- RasMol (Sayle & Milner-White, 1995)
- Origin 8.0 (Microcal Software, Inc)
- FirstGlance in Jmol (Jmol: an open-source Java viewer for chemical structures in 3D. http://www.jmol.org/)
- VMD (Humphrey, Dalke, & Schulten, 1996)

5.1.4 RMSD calculation

The Root Mean Square Deviation (RMSD) is commonly used to measure the similarity between two superimposed structures and, in a trajectory, find timepoints when conformation changes. The RMSD value is calculated by:

$$RMSD = \sqrt{\frac{1}{n}\sum_{i=1}^{n} d_i^2}$$

Where the averaging is performed on the *n* pairs of equivalent atoms and d_i is the distance between the two atoms in the i^{th} pair. We had used the RMSD calculation tool of the VMD software to compute it.

5.1.5 RMSF calculation

The Root Mean Square Fluctuation (RMSF) is a measure of average atomic mobility, and it is useful to obtain information on local structural flexibility. The RMSF value is calculated by:

$$\mathrm{RMSF} = \frac{1}{T}\sum_{T_i=1}^{T} \left(x_i(t_j) - \tilde{x}_i\right)^2$$

Where T is the total time (=number of frame) to be averaged, $\tilde{x}_i$ is the average position of each particle, x_i the position of each particle at the time t. To calculate the RMSF we had used the g_rmsf GROMACS tool, which computes the root mean square fluctuation of atomic positions in the trajectory, after fitting to a reference average structure. We have considered for the average structure calculation, only the last part of the simulation in which the proteins were stable.

5.2 Recombinant proteins

5.2.1 Culture medium and antibiotics

Luria broth (LB)

0.5% (w/v) Yeast extract

1% (w/v) Tryptone

1% (w/v) Sodium Chloride

Re-suspend all components in water and sterilize at 120°C for 20 min. (pH= 7.9)

LB Agar

Add Agar at 1.5% (w/v), to the LB medium. Sterilize at 12°C for 20min.

SOC medium

LB medium + 10 mM $MgCl_2$ + 20 mM glucose

Antibiotics

Ampicillin 100 µg/ml

Chloramphenicol 34 µg/ml

Kanamycin Sulfate 50 µg/ml

Tetracycline 10 µg/ml

All antibiotics were purchased from Sigma-Aldrich.

5.2.2 Cells lines

XL1-Blue (XL1B) Competent Cells were used for cloning passages. These *Escherichia coli* cells have the maximum transformation efficiency, optimal for cloning and plasmid propagation. XL1B cells are Tetracycline resistant and allow blue-white color screening for recombinant plasmids.

BL21cp Competent Cells were used for protein expression. These chemically competent E*scherichia coli* cells are suitable for transformation and protein expression. BL21 cells contain a plasmid with the Chloramphenicol resistance and more genes copies which codify for rare tRNAs.

The efficient production of recombinant proteins is often prevented by the deficiency of some tRNAs that, on the contrary, are abundant in the organism from which the protein derives. Using BL21 cells, the problem of rare tRNAs is solved.

5.2.3 Plasmid vector

We had used two types of plasmid vectors pET-28b and pET-11b (Novagen) containing the sequences of the protein of interest. Both vectors are members of the pET series, in which the cloning gene is under control of the T7 RNA polymerase promoter, regulated by the lac operator. Their genomic maps are shown in figure 5.3 and 5.4.

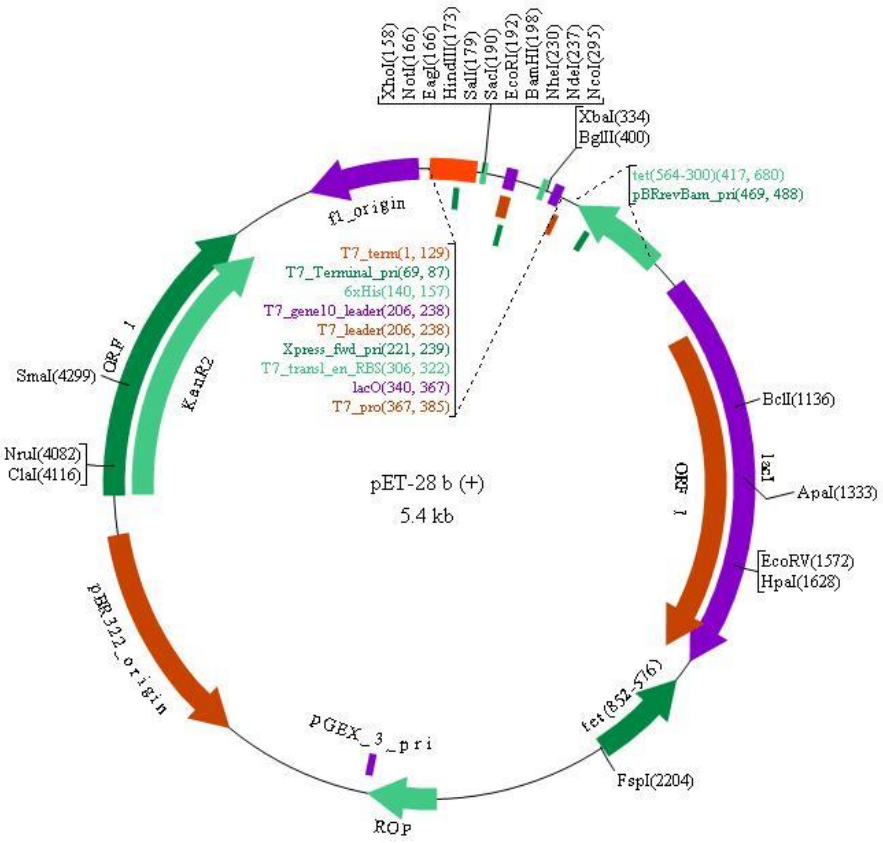


Figure 5.3 pET28b genome map.

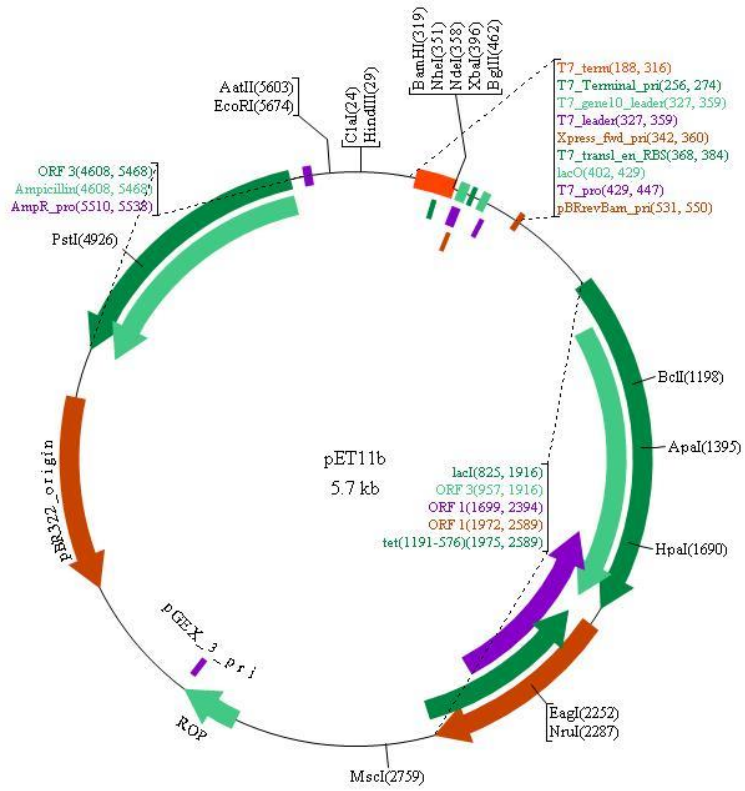


Figure 5.4 pET11b genome map.

5.2.4 Mutants

The cloning operations had already been made, thus the sequence of the WT hCRBP1 was included into the plasmid vectors pET28b and ready to the mutagenesis.

5.2.4.1 Primers design

As described in the experimental techniques section (Parag. 3.1) the achievement of hCRBP1 mutants needs the design of pairs of degenerated primers. All oligomers were purchased from BMR Genomics and their sequences are listed below:

<u>hCRBP1 Q108L</u>

Up-stream: 5’ GTGGCTGGACC<u>CTG</u>TGGATCGAGGG 3’

Down-stream: 5’ CCCTCGATCCA<u>CAG</u>GGTCCAGCCAC 3’

To obtain this mutant, the WT hCRBP1 sequence was used as template.

hCRBP1 K40L and hCRBP1 K40L/Q108L

Up-stream: 5' CTTGCTGAAGCCAGACCTAGAGATCGTGCAGGAC 3'

Down-stream: 5' GTCCTGCACGATCTCTAGGTCTGGCTTCAGCAAG 3'

To obtain the hCRBP1 K40L mutant, the WT hCRBP1 sequence was used as a template.

To obtain the hCRBP1 K40L/Q108L mutant, the hCRBP1 Q108L mutant sequence was used as a template.

hCRBP1 Y60F

Up-stream: 5' TGAGCACTTTTAGGAACTTCATCATGGACTTCCAGGT 3'

Down-stream: 5' ACCTGGAAGTCCATGATGAAGTTCCTAAAAGTGCTCA 3'

To obtain this mutant, the WT hCRBP1 sequence was used as a template.

hCRBP1 R58L

Up-stream: 5' CGCACGCTGAGCACTTTTCTGAACTACATCATGGACTTC 3'

Down-stream: 5' GAAGTCCATGATGTAGTTCAGAAAAGTGCTCAGCGTGCG 3'

To obtain this mutant, the WT hCRBP1 sequence was used as a template.

hCRBP1 R58L/Y60F

Up-stream: 5' CGCACGCTGAGCACTTTTCTGAACTTCATCATGGACTTC 3'

Down-stream: 5' GAAGTCCATGATGAAGTTCAGAAAAGTGCTCAGCGTGCG 3'

To obtain this mutant, the hCRBP1 Y60F mutant sequence was used as a template.

hCRBP1 E72L

Up-stream: 5' GTTGGGAAGGAGTTTGAGCTGGATCTGACAGGCATAGA 3'

Down-stream: 5' TCTATGCCTGTCAGATCCAGCTCAAACTCCTTCCCAAC 3'

To obtain this mutant, the WT hCRBP1 sequence was used as a template.

5.2.4.2 PCR reaction

A PCR reaction was accomplished to create mutant plasmids, preparing the sample reactions (final volume 50uL) as indicate below:

- 5 µl of 10X Pfu reaction buffer
- 30 ng of dsDNA template
- 125 ng of upstream oligonucleotide
- 125 ng of downstream oligonucleotide
- 1 µl of dNTP mix containing each dNTP at 12.5 mM concentration
- Add water to a final volume of 50 µl.

Finally, 1 µl of Pfu Turbo DNA polymerase (2.5 U/µl, Stratagene) was added to the reaction.

The PCR reaction was, then, carried out following the cycling parameters listed below and using a Thermocycler:

1. 95°C for 30 seconds
2. 95°C for 30 seconds
3. 55°C for 1 minute
4. 68°C for 12 minutes

Steps 2, 3 and 4 were repeated for 12 cycles.

At the end of the reaction, the sample was placed in ice for 2 minutes to cool the reaction.
To eliminate the parental DNA, 1µl of DpnI restriction enzyme (10 U/µl, Stratagene) was added to the amplification reaction and the sample was incubated at 37°C for 2 hours. After an over-night precipitation at -20°C, the DNA was recovered and used to transform E. coli XL1B cells by electroporation (BioRad MicroPulser). Some colonies were chosen and their plasmid DNA was sequenced to verify the occurrence of the desire mutation. Finally, E. coli BL21cp cells were transformed with mutant plasmids, in order to proceed with the expression and purification of mutant proteins.

5.2.5 Expression of WT hCRBP1 and its mutant forms

5.2.5.1 Cells grown and protein expression

One of the several colonies obtained after the transformation, and containing the WT or the mutant plasmids, was inoculated into 10 mL of LB medium, adding Chloramphenicol and Kanamycin at the concentrations previously defined. The culture was grown over-night at 37°C, and 5 mL of it were inoculated in 200 mL of LB medium in presence of the above mentioned antibiotics. The culture was grown at 37°C, under agitation, since an absorbance value (OD) between 0.6 and 0.8 (non-induced) at 600 nm was reached. Then a control sample was put aside and the expression of the recombinant protein was induced by adding IPTG (isopropyl-1-thio-β-D-galactopyranoside) in the flask at the final concentration of 0.5 mM. The protein expression was carried on over-night at 4°C, under agitation. After this time, the OD at 600 nm should be doubled if compared with the OD of non-induced culture. 1 mL of induced culture was put aside as control sample. To verify the correct expression accuracy of the protein of interest, the two control samples (non-induced and induced) were loaded on a polyacrylamide gel using denaturant conditions (SDS-PAGE).

5.2.5.2 Cells lysis

The recombinant proteins expressed in E. coli cells, for less than they have a secretory sequence, will be released into the cytosol. To recover the recombinant protein, we carried on the cells lysis, using the Misonix Sonicator 3000 (Cole-Parmer®), after the re-suspension of the cells in the lysis buffer (Na phosphate 50 mM, NaCl 0.3 M, ,10% glycerol (pH=8.0)) and the addition of protease inhibitors (leupeptin, pepstatin, phenylmethylsulfonyl fluoride (PMSF)). At the end of the lysis, the sample was centrifuged to separate the pellets from the supernatant; for both phases, a control sample was isolated and then loaded on a polyacrylamide gel for the SDS-PAGE analysis.

5.2.6 Expression of WT hCRBP3 and WT hCRBP4

The cloning operations had already been made, thus the sequences of hCRBP3 and 4 were already included into the plasmid vectors pET11b and were ready to be used for the transformation of BL21cp cells and to proceed with the protein expression.
For the expression of these two isoforms we followed the same strategy used for hCRBP1 and its mutants with the only exception of the antibiotic type; the sequences coding for hCRBP3 and 4 are inserted into the pET11b plasmid vector, which requires ampicillin for the selection. Thus, during the preparation of the medium for cells culture, ampicillin and chloramphenicol were added.

5.2.7 Protein purification

For each type of CRBP (hCRBP1, 3, 4 and hCRBP1 mutants) two chromatographic steps were enough to obtain a sufficient degree of protein purity (protein/contaminants ration higher than 80%). At first, the supernatant sample derived from the lysis, was dialyzed in Tris HCl 30 mM pH=7,5 (buffer A) and loaded on a pre-equilibrated anion exchange chromatographic column, in the same buffer A; in this way we could elute positively charged proteins in the flow through and separate negatively charged proteins after the application of an increasing salt gradient (buffer A: Tris HCl 30 mM pH 7,5; buffer B: Tris HCl 30 mM + NaCl 0.5 M pH=7,5). Fractions that showed significant 280 nm (higher than 0.1 OD) absorbance values, and a maximum in the 270-280 nm range, were analyzed by SDS-PAGE, and fractions containing the protein of interest were collected and the total volume was reduced to apply it to a size exclusion column. In this passage, the proteins were separated on the basis of their native dimension. Significant fractions were collected and concentrated again.

5.2.8 Protein quantification

Pure proteins were quantified using the Cary UV-Visible spectrophotometer measuring the absorbance at 280 nm. Using ProtParam (Gasteiger et al., 2005), a tool available on the web

(http://web.expasy.org/protparam/), we had calculated the extinction coefficient (ε) of each protein of interest; these values are summarized in table 1.

Table 1 Extinction coefficient value.

Protein	E ($M^{-1}cm^{-1}$)	Abs corresponding to the concentration of 1 g/L
hCRBP1 WT	26595	1.67
hCRBP1 Q108L	26595	1.679
hCRBP1 K40L	26595	1.679
hCRBP1 Q108L/K40L	26595	1.681
hCRBP1 Y60F	24980	1.580
hCRBP1 E72L	26470	1.670
hCRBP1 R58L	26470	1.675
hCRBP1 R58L/Y60F	24980	1.582
hCRBP3	23950	1.505
hCRBP4	24980	1.608

Knowing the ε value and the absorbance of the sample at 280 nm is possible to calculate the concentration by the Lambert-Beer equation, considering an optical length of 1 cm:

$$A = log\left(\frac{I_0}{I}\right) = \varepsilon cd$$

where:

c = sample concentration (M)

d = optical length (cm)

I_0 = incident light

I = light after it passes through the sample

5.3 Binding assays performed by absorption spectroscopy

The protein absorption spectrum was recorded in the 250-400 nm range, after the addition of the protein in a concentration range of 7-9 µM (corresponding to an absorbance of 0.2-0.4) directly into a spectrophotometric cuvette containing Tris HCl 30 mM + NaCl 0.150 M (pH=7,5). After the addition of a slight molar excess of ligand (2-2.5 molar equivalents) to the protein sample, it was maintained in the dark for 10 minutes and then centrifuged; thus, the spectrum of the complex was recorded.

5.4 Fluorescence Titrations

To investigate the binding of retinoid to CRBPs, fluorescence titrations were carried out adding into the spectrofluorometric cuvette, containing apo-CRBP 0.1-0.5 µM in 500 µL of Tris HCl 30 mM + NaCl 0.150 M buffer (pH=7.5), 0.2 µL aliquots of concentrated ethanolic solutions of retinol (at a known concentration), progressively saturating the protein. The ethanol concentration in the binding assay buffer should not exceed 1-2% (v/v) in order to reduce to a minimum the possible influence of water-miscible organic solvent on retinoid-protein interaction. At the same time, the addition of small aliquots preserves the sample dilution that may cause changing in the signal. The solution was gently stirred after each addition and maintained in the dark, at 23°C±1°C, for the time needed to obtain a stable fluorescence signal.

To detect the fluorescence signal we had used an excitation wavelength of 280 nm. Each titration was repeated four times. The plotting of the intensity of protein intrinsic fluorescence as a function of retinol concentration was carried out using SigmaPlot Version 11.0.

5.5 Delipidation by lipidex resin

It is known that CRBPs and FABPs are able to bind small lipophilic E. coli molecules during their recombinant expression. With the aim to obtain an authentic apo-protein from purified recombinant WT CRBP1, the protein was delipidated by treatment with the hydrophobic resin

Hydroxyalkoxypropyl-Dextran Type IX (Sigma Aldrich), equivalent to Lipidex 1000, essentially as described in literature (Glatz & Veerkamp, 1983), and using a minicolumn:

1. 200 µL of resin were washed with 200 µL of ethanol for 3 time.
2. The column was equilibrated in the protein buffer (Tris Hcl 15 mM) pre-heated at 37°C.
3. The protein solution, previously heated at 37°C, was applied to the column and after 10 minutes of incubation at 37°C, the sample was centrifuged at 5000 rpm for 15 minutes. The eluate was recovered and incubated at 37°C.
4. Steps 1-3 were repeated for 3 times to be sure to remove any ligand bound to the protein.

The lipidex delipidation was also used to isolate the ligand bound to the protein; in this case the ligand is eluted during the wash with methanol. This fraction can be concentrated and used for the characterization of the small molecule by absorbance spectroscopy, mass spectrometry or much more.

5.6 3D structures

5.6.1 Crystallization protocol

Below are summarized steps useful for the crystallization process:

1. Apply a bed of sealant (petroleum jelly) creating a circular bead of the upper edge of the reservoir allowing a 2 mm gap to promote the escape of the air during the application of the cover slide.
2. Pipet the crystallization reagent (0,5-1 mL) into the reservoir.
3. Pipet 0,5-2 µL of sample on the cover slide and add an equal volume of crystallization reagent taking it from the reservoir.
4. Invert the cover slide so that the drop is hanging from it and position the cover slide onto the bead of grease on the relative reservoir.
5. Press the slide down onto the grease and twist the slide 45º to ensure a complete seal.

5.6.2 Crystallization conditions

Crystals of WT human apo- and holo-CRBP1 and of the mutant forms of human CRBP1 (apo- and holoCRBP1 K40L, holo-CRBP1 Q108L and holo-CRBP1 K40L/Q108L), were obtained at room temperature from a precipitant solution containing 0.2 M ammonium chloride, 0.1 M sodium acetate, pH 5,0, 20% PEG 6000 (PACT 1-8, Molecular Dimensions, UK) a condition that is similar, in regard to both pH and precipitant, to that adopted for the crystallization of HT human CRBP1 (Silvaroli et al., 2016). Crystals are grown in two days.

5.6.3 X-Ray diffraction

The X-Ray diffraction and the analyses of diffraction data were performed by Professor Zanotti and co-worker, of the Biomedical Science Department of the University of Padua. Crystals diffracted at high resolution, from 1.23 Å to 1.43 Å resolution, except for WT apo CRBP1, which diffracted to the relatively lower resolution of 1.70 Å. X-ray diffraction data were collected at the beamline PXIII (SLS synchrotron, Switzerland), ID29 or ID30B (ESRF, Grenoble, France) at 100°K, without any cryoprotectant. Data were processed by XDS (Kabsch, 2010) and scaled with Scala (Evans, 2006). All crystals belong to the P212121 space group, with unit cell dimensions reported in Table I. The VM value for all crystals is 2.00 - 2.08 Å3/Da, corresponding to a solvent content of about 38%-41% of the total volume, with one molecule in the asymmetric unit. The structure was solved by molecular replacement with software Molrep (Lebedev, Vagin, & Murshudov, 2007), using as a template the solution structure 1KGL (Franzoni et al., 2002). The structures were refined using Refmac5 (Murshudov et al., 2011) or Phenix (Adams et al., 2010) and manually rebuilt and corrected by software Coot (Emsley & Cowtan, 2004). Towards the end of the refinement, the electron density for the retinol or a different ligand bound inside the cavity was clearly visible in all cases, except for the authentic apo CRBP1 and apo CRBP1 K40L (which were devoid of any ligand), and was included with restraints produced by the server PRODRG (Schüttelkopf & Van Aalten, 2004). Solvent molecules were added automatically using the Phenix automated procedure and checked manually. Hydrogen atoms were introduced in the last cycle of refinement. The statistics of data collection and refinement are reported in Table I. The geometry of the final model was checked by Molprobity (Lovell et al., 2003).

5.7 Ligand characterization

5.7.1 Absorption analysis

As described at paragraph 5.5, a solution of holo-CRBP4 was delipidated by lipidex resin; the resin was washed with 200 µL of pure water for two times, after the passage of the protein. The ligand was then recovered by washing the resin with 100 µL of methanol. On this solution an absorption spectrum was recorded, which was compared with the spectra of other retinoids known in literature.

5.7.2 MALDI analysis: samples preparation

To performed the MALDI analysis on CRBP4 bound to the unknown ligand, a sample containing 2.28 mg/mL of holo-protein was conserved in Tris HCl 15 mM, pH=7.5. Three different conditions were tested:

1. Native condition: 3 µL of holo-CRBP4 were mixed to 3 µL of the matrix, α-Cyano-4-hydroxycinnamic acid 20 mg/mL (HCCA). 1 µL of the mixture was applied to one of the spots present on the metallic plate (Fig. 5.5). Other two spots were filled in this condition.
2. Complex in acetonitrile (1: 2 = protein: acetonitrile): 3 µL of holo-CRBP4 were mixed to 6µL of acetonitrile (denaturant) to obtain a mild denaturant condition. The mixture was left to rest 10 minutes at room temperature; 5 µL of the mixture were, then, mixed to 5 µL of matrix (HCCA). Finally, three drops were applied to the plate, each on containing of 1 µL of mix sample.
3. Denaturant condition (1: 1 = protein: acetonitrile): 5 µL of holo-CRBP4 were mixed to 5 µL of Acetonitrile. The mixture was treated with a temperature step of 100°C for 10 minutes to achieve the maximum denaturation degree. Then, 5 µL of HCCA were added to 5 µL of the mixture and three drops were applied to the plate.

In this experiment two control samples were needed:

1. A drop containing only the matrix was applied to one spot of the plate.
2. One spot contained only acetonitrile.

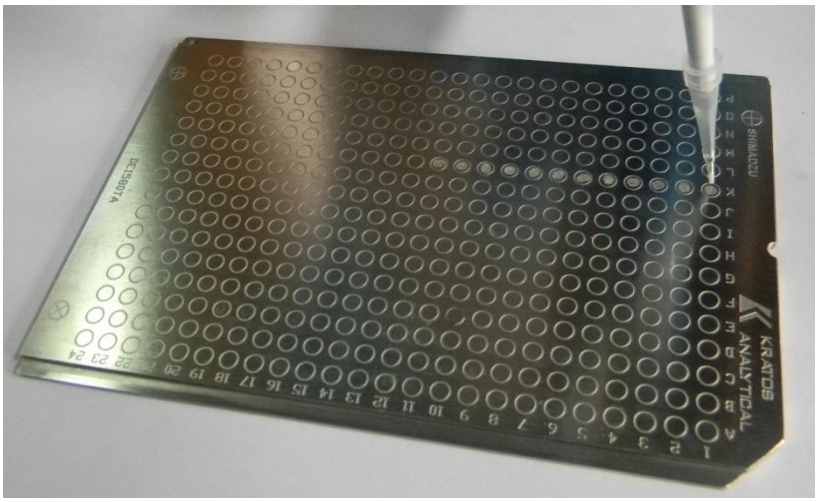


Figure 5.5 MALDI plate.

After the application of each drop, the spots were dried with a nitrogen flux. The plate was entrusted to a technician of CIM (Centro Interdipartimentale Misure) of University of Parma, who perfomed the analyses. Results were analyzed with mMass 5.5.0 software (Strohalm, Kavan, Novák, Volný, & Havlíček, 2010).

5.7.3 Mass Spectrometry analysis: sample preparation

With the aim to characterize the unknown ligand bound to CRBP4 on the basis of its molecular mass, a sample of the ligand was prepared to perform the ESI-MS analysis. Few microliters of the sample, obtained after the treatment mentioned at paragraph 5.8.1, were directly injected to the mass spectrometer (ESI-MS); a technician of the CIM center have performed the analysis.

RESULTS AND DISCUSSION

6.1 Study of the key binding residues of CRBP1

6.1.1 Proteins expression and purification

To evaluate the role and the real importance of the two key binding residues Gln108 and Lys40 in their interaction with retinol, three mutants were obtained (see Materials and Methods): Q108L/CRBP1, K40L/CRBP1 and the Q108L/K40L/CRBP1 double mutant. To verify the insertion of the mutations, DNA mutant sequences were sequenced and the proteins expression was carried out. We have also expressed the WT protein to obtain the first crystal structure of human CRBP1, and to use WT CRBP1 as a control in the binding assays. As described in Material and Methods, to verify the success of the proteins expression and to evaluate the presence of the protein of interest in the soluble part, after cell lysis, an SDS-PAGE analysis was performed and we report below the results for each expressed protein:

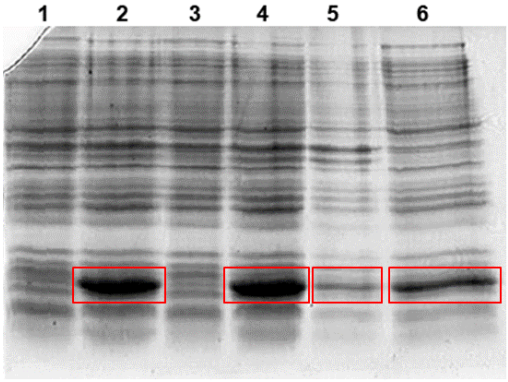


Figure 6.1 SDS-PAGE analysis of WT hCRBP1: non-induced culture (lanes 1 and 3), induced culture (lanes2 and 4), pellet (5) and soluble protein (6) after cell lysis. The WT hCRBP1 is highlight with red rectangles.

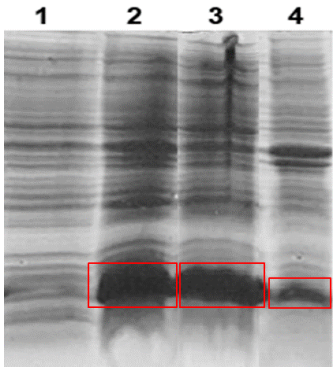


Figure 6.2 SDS-PAGE analysis of Q108L/hCRBP1: non-induced culture (lane 1), induced culture (lane 2), soluble protein (3) and pellet (4) after cell lysis. The Q108L/CRBP1 is highlight with red rectangles.

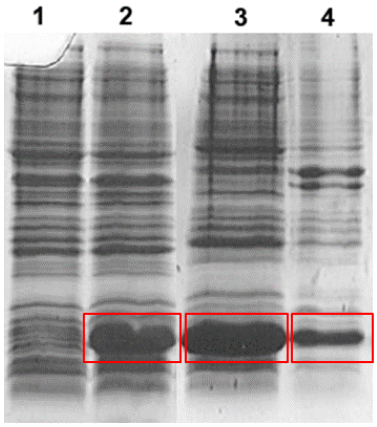


Figure 6.3 SDS-PAGE analysis of K40L/hCRBP1: non-induced culture (lane 1), induced culture (lane 2), soluble protein (3) and pellet (4) after cell lysis. The K40L/CRBP1 is highlight with red rectangles.

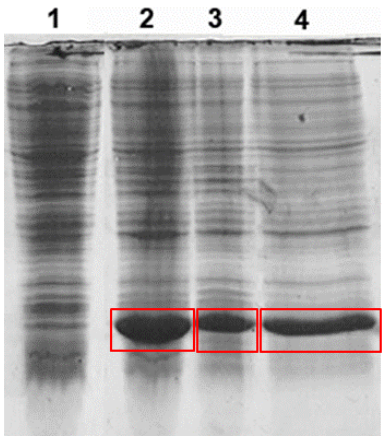


Figure 6.4 SDS-PAGE analysis of K40L/Q108L/CRBP1: non-induced culture (lane 1), induced culture (lane 2), pellet (4) and soluble protein (3) after cell lysis. The K40L/Q108L/CRBP1 is highlight with red rectangles.

The major part of the proteins of interest, for each different preparation, was soluble. Thus, we were able to proceed to the protein purification to obtain pure protein solutions. For each preparation two chromatographic passages were performed; the first was based on an anionic exchange chromatography. To evaluate the presence of proteins in the eluted fractions, we have measured the 280 nm absorbance of each fraction and those with significant OD values were analyzed by SDS-PAGE. Images of polyacrylamide gel of significant fractions are shown below for each purified protein.

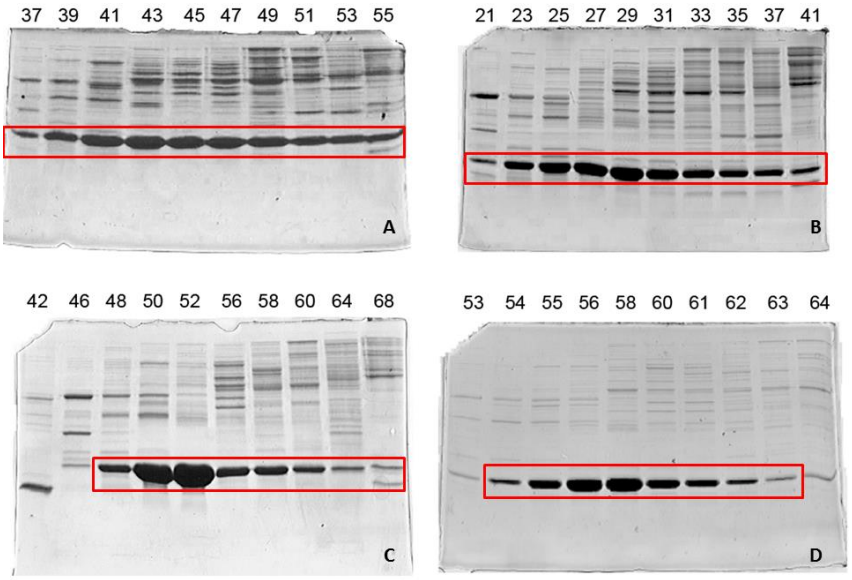


Figure 6.5 SDS-PAGE analyses of fractions with significant absorption at 280 nm, deriving from ion exchange chromatography: A) WT hCRBP1 B) Q108L/CRBP1 mutant C) K40L/CRBP1 mutant and D) K40L/Q108L/CRBP1 double mutant.

Fractions with appreciable presence of protein of interest (39 to 55 for WT, 23 to 37 for Q108L/CRBP1, 48 to 60 for K40L/CRBP1 and 54 to 62 for K40L/Q108L/CRBP1) were collected and concentrated to be processed with the second chromatographic passage: the size exclusion chromatography.

Fractions with relevant absorption at 280 nm were analyzed by SDS-PAGE; pictures of polyacrylamide gels are shown in figure 6.6.

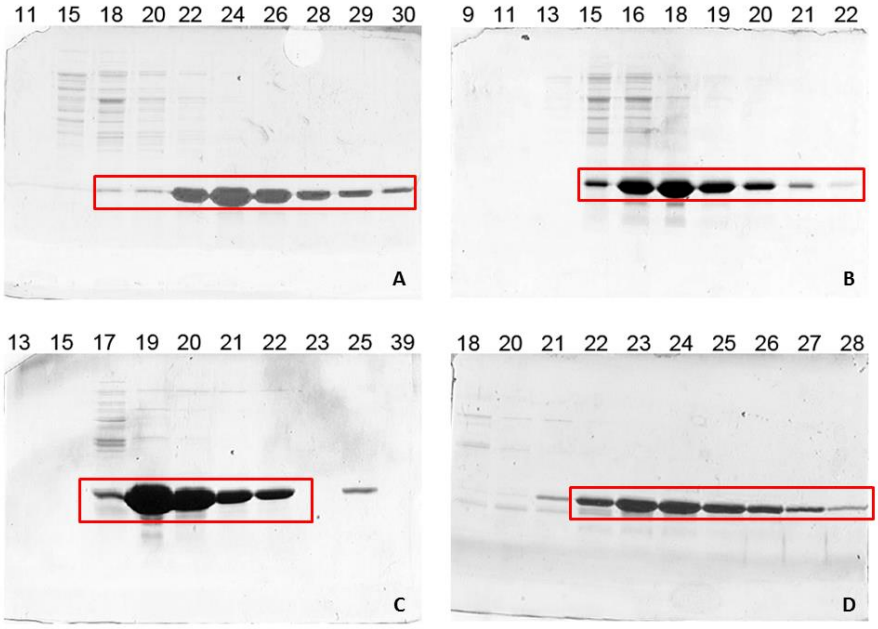


Figure 6.6 SDS-PAGE analysis of fractions with significant absorption at 280 nm, deriving from size exclusion chromatography: (**A**) WT hCRBP1 (**B**) Q108L/CRBP1 mutant (**C**) K40L/CRBP1 mutant and (**D**) K40L/Q108L/CRBP1 double mutant.

Fractions deriving from size exclusion chromatography, containing the protein of interest and showing the presence of less contaminants, were collected and concentrate (24 to 30 for WT, 18 to 21 for Q108L, 19 to 22 for K40L and 22 to 27 for K40L/Q108L). The pure proteins were obtained and quantified on the basis of their extinction coefficients (see Material and methods, paragraph 5.3). Final yields for each protein were calculated: 10 mg/mL for the WT, 12 mg/mL for the Q108L/CRBP1 mutant, 14 mg/mL for K40L/CRBP1 mutant and 8 mg/mL for K40L/Q108L/CRBP1 double mutant.

6.1.2 Binding assays for the interaction of retinol with mutant forms

The binding of retinol to WT and mutant forms of human apo CRBP1 was analyzed by monitoring typical absorption spectra of the bound ligand, which in the case of the WT protein (Fig. 6.7A) is characterized by a peak at approx. 348 nm, with two shoulders at 333 and 365 nm, respectively. In the case of the holo-Q108L/CRBP1 (Fig. 6.7B), a similar absorption spectrum is obtained, while in the case of the holo-K40L/CRBP1 (Fig. 6.7C) the absorption spectrum shows three better defined peaks at 330, 345 and 363nm, respectively. When compared to the peaks of the holo-K40L/CRBP1, three similar absorption peaks, although with lower intensity, are also visible for the double mutant holo-K40L/Q108L/CRBP1 (Fig. 6.7D). The lower absorption intensities for the latter mutant form are attributable to a relatively lower saturation by the chromophoric ligand, due to the low binding affinity. The aforementioned results suggest that the presence of Lys40 and its interaction with retinol are critical for determining the characteristic chromophoric properties of the CRBP1-bound retinol, which are likely due to the proximity of the presumably protonated side chain of Lys40 to the retinol isoprene tail (Lu et al., 2000)(Franzoni et al., 2002).

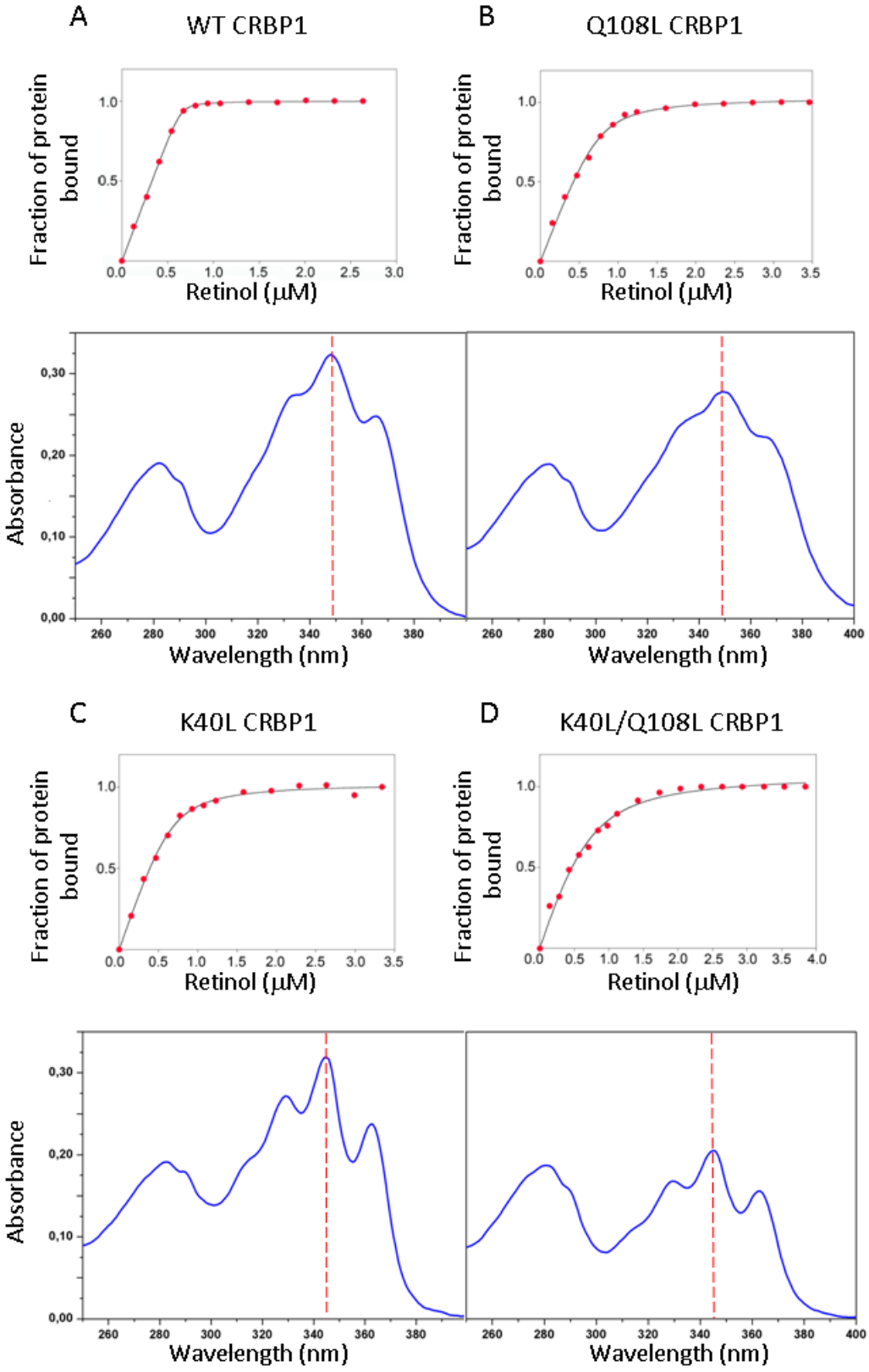


Figure 6.7 Representative titration curves for the interaction of retinol with WT CRBP1 (**A**), Q108L/CRBP1 (**B**), K40L/CRBP1 (**C**), and Q108L/K40L/CRBP1 (**D**). The estimated Kd values are 4.5 nM for WT CRBP1, 64 nM for Q108L/CRBP1, 71 nM for K40L CRBP1 and 255 nM for the double mutant Q108L/K40L/CRBP1. For each panel the absorption spectrum of the corresponding CRBP1 form in complex with retinol is also shown.

Spectrofluorometric titrations of WT and mutant forms of apo CRBP1 with retinol were carried out to determine the dissociation constant of the retinol-protein complexes, by exploiting the quenching of the intrinsic protein fluorescence, which is associated with retinol binding to the apoprotein as a result of an efficient energy transfer to the chromophoric protein-bound ligand (Fig. 5.7). Similar Kd values could be estimated for the two single mutants Q108L (65 ± 5 nM) and K40L (70 ± 5 nM), which are much higher than that determined for WT CRBP1 (4.5 ± 0.5 nM). In the case of the double mutant K40L/Q108L CRBP1, the estimated Kd value is approx. 250 nM (250 ± 30 nM), indicating that the substitution of both key residues results in a further strong decrease in binding affinity.

6.1.3 Structures of human WT apo- and holo-CRBP1

Our data on the structures of human WT CRBP1, as well as its mutant forms, were obtained before the recent publication dealing with the structure of the same human protein, which, differently from our protein, contained a His_6-Tag at its C-terminus (Silvaroli et al., 2016), and is designated as His-Tag (HT) CRBP1. Despite this difference, WT and mutant forms of CRBP1 (see below) and HT CRBP1 crystallize in similar conditions (see Materials and Methods) and crystals are essentially isomorphous (Table 1), with the exception of HT apo-CRBP1 (Silvaroli et al., 2016).

All residues present in the structures of WT CRBP1, from Pro1 to Gln134, have been included in our structural models. They are clearly visible in the electron densities, except for Pro1, which is partially flexible. The root mean square deviation (RMSD) for the superposition of equivalent Cα atoms of human WT holo-CRBP1 and human HT holo-CRBP1 (PDB ID 5H8T, (Silvaroli et al., 2016)) is 0.55 Å. The main difference between the two proteins is at the C-terminus, where the presence of the His-Tag influences the conformation of the last three amino acids. The RMSD is 0.793Å for the comparison of the structures of human WT holo-CRBP1 and rat WT holo-CRBP1 (PDB ID 1CRB, (Cowan, Newcomer, & Jones, 1993)), but the latter difference must be essentially ascribed to the lower resolution of the previous structure and possibly to different crystallization conditions. In WT holo-CRBP1, the protein structure and the mode of binding of retinol in the inner cavity is the same as that found for HT holo-CRBP1, with the retinol hydroxyl group pointing towards Gln108 and Lys40 and forming H-bonds with both Oε1 atom of Gln108(OH – Oε1 distance, 2.92Å) and Nζ of Lys40 (OH – Nζ distance, 3.22 Å), respectively (the same distances are 2.81 Å and 2.95 Å, respectively, for HT holo-CRBP1) (Figs. 6.8A and 6.11A).

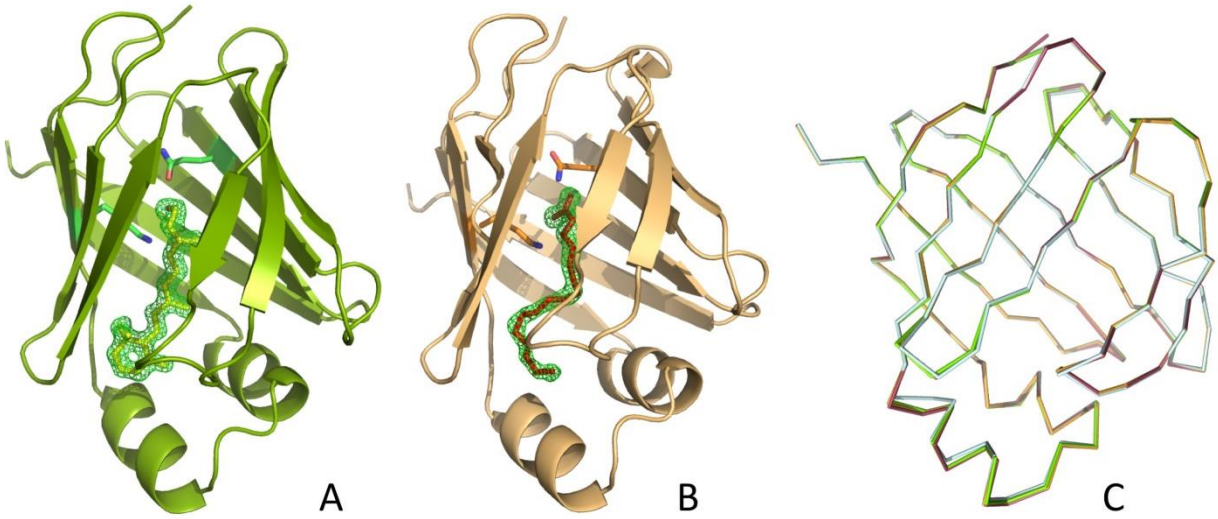


Figure 6.8 (A) Cartoon view of WT holo-CRBP1 with the retinol bound inside the inner cavity of the protein. (B) Same as (A), for CRBP1 in complex with a hypothetical fatty acid (palmitic acid). The electron density was calculated with coefficients (2Fobs – Fcalc) and contoured at 1.5 level. (C) Superposition of Cα chain traces of WT holo-CRBP1 (chartreuse green), WT delipidated apo-CRBP1 (orange), holo-Q108L/K40L/CRPB1 (pale cyan) and WT CRBP1 in complex with palmitic acid (raspberry). Figures A, B and C were drawn using PyMOL (Pymol version 1.7.4, Schrödinger, LLC).

Unexpectedly, in a first structure of the putative WT apo CRBP1, crystallized in the same conditions as the holo-protein and in the absence of any added ligand, a long linear electron density, slightly bent, was visible inside the binding cavity (Fig. 6.8B)**,** which indicates that upon its expression the protein bound an as-yet-uncharacterized compound produced by *E. coli* metabolism and possessing affinity for CRBP1. A saturated fatty acid was selected as a possible hydrophobic ligand, since the flexibility of its chain would allow a better fitting in the density. A palmitate molecule fitted quite well in the electron density, with an occupancy slightly lower than 1. The RMSD for the superposition of the equivalent Cα atoms of WT holo-CRBP1 and WT CRBP1 in complex with the hypothetical ligand is 0.16 Å, with the positions of all the side chains pointing inside the cavity perfectly conserved, in keeping with the rigidity of the protein scaffold. In our model, one of the oxygen atoms of the carboxylate group of the hypothetical fatty acid interacts with nitrogen of Lys40 (O2 – Nζ, 2.71Å) and with a water molecule (O2 – O, 2.72Å), the second with the oxygen of Gln108 (O1 – Oε1, 2.86Å) (Fig. 6.11D). It is also very likely that an ionic interaction takes place between the positively charged $–NH^{3+}$ group of Lys40 and the negatively charged carboxylate of the hypothetical fatty acid.

The latter observations suggested to us that the mutation K40L would lead to a lack of interaction between the ε-amino group of Lys40 and the carboxylate group of the aforementioned hypothetical ligand, with the possibility of generating an authentic apo-CRBP. Indeed, upon crystallization the X-ray analysis of the K40L/CRBP1 mutant form revealed that it was a real apo-

protein, in which the cavity was filled with 11 solvent molecules that do not mimic the position of the hypothetical ligand. This result is consistent with our hypothesis that CRBP1 has the ability to bind carboxylic acids, such as fatty acids, through the formation of an ionic bond with Lys40. The latter ligand, however, does not possess high binding affinity for CRBP1, since it can be easily and fully displaced by the retinol.

In order to determine the structure of an authentic WT apo-CRBP1, the WT CRBP1 obtained upon expression in *E. coli* was delipidated by treatment with a hydrophobic resin (see Material and Methods) and crystallized. The structures of WT apo-CRBP1 and apo-K40L/CRBP1 are nearly superimposable, the RMSD for the superposition of equivalent Cα atoms being 0.29 Å (Fig. 6.8C). The structures of WT apo-CRBP1 is at lower resolution (1.7Å) as compared to the HT apo-CRBP1. Despite that, most of the solvent molecules located inside the cavity are conserved in both structures, as well as in the case of the structure of apo-K40L/CRBP1 mutant form. The RMSD for equivalent Cα atoms upon superposition of WT apo-CRBP1 and HT apo-CRBP1 is higher, 0.85Å. In particular, we do not observe for our WT apo-CRBP1 the movement of the E-F loop (residues 73-81), which by itself shows a RMSD of 3.05 Å in the comparison between WT and HT apo-CRBP1, and of the nearby side chain of Tyr60, found for the HT apo-CRBP1. On the contrary, the latter region of WT apo-CRBP1 maintains essentially the conformation found for WT and HT holoproteins. In fact, WT holo- and apo-CRBP1 are very similar, the RMSD for the superposition of the equivalent Cα atoms being 0.21Å (Fig. 6.8C). The positions of all the side chains pointing inside the cavity are perfectly conserved, a clear indication of the rigidity of the protein scaffold. The C-terminus also points in the same way as the holoprotein.

Table 1 Statistics on data collection and refinement. 1800 frames of 0.1° each were collected. Numbers in parentheses refer to the last resolution shell.

X-ray data	Holo-CRBP1	Holo-Q108L/CRBP1	Holo-K40L/CRBP1	Holo-Q108L/K40L/CRBP1
Wavelength (Å)	0.91881	0.91881	1.000	0.91881
Space group	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$
Cell parameters [a,b,c, Å]	34.120, 49.142, 75.801	34.035, 48.563, 75.748	34.483,47.187, 81.033	34.544, 47.414, 80.595
Resolution (Å)	41.23 – 1.26 (1.33-1.26)	40.88–1.43 (1.51–1.43)	47.19-1.61 (1.67-1.61)	47.41-1.40 (1.48-1.40)
Independent reflections	34,582 (4,718)	23,562 (2,724)	17,701 (1718)	26,769 (3,676)
Multiplicity	6.0 (5.4)	3.3 (3.0)	3.9 (4.0)	6.1 (5.7)
Completeness (%)	99.0 (94.5)	94.4 (82.5)	99.6 (99.8)	99.6 (98.1)
$\langle I/\sigma(I)\rangle$	17.5 (2.9)	9.2 (2.2)	11.4 (1.6)	18.3 (2.7)
R_{merge}	0.038 (0.475)	0.063 (0.504)	0.058 (0.643)	0.043 (0.518)
R_{pim}	0.018 (0.241)	0.042 (0.293)	0.038 (0.421)	0.021 (0.255)
Refinement				
Protein atoms (no hydrogen) / solvent atoms /ligand /others	1136 / 179 / 21	1141 / 180 / 21	1107 / 195 / 21 /1	1102 / 112 /21
R_{cryst} / R_{free}	0.165 / 0.184	0.181 / 0.214	0.183 / .0234	0.194 / 0.223
Geometry				
Ramachandran favored /outliers (%)	97.0 / 0.0	97.0/0.0	97.0/ 0.0	97.0/0.0
Rotamer outliers	0.0%	0.8%		0.8%
Overall score	1.5	1.8		1.6
R.m.s.d. on bonds length [Å], angles (°)	0.011, 1.65	0.012, 1.71	0.012, 1.59	0.012, 1.57

* Completeness is low at the highest resolution, but data were left, since statistics are very good. The completeness is 99% at 1.37Å, which can be considered the actual resolution.

Table 1 Continue

X-ray data	Delipidated apo-CRBP1	Apo-K40L/CRBP1	CRBP1-palmitic acid
Wavelength (Å)	0.973186	1.00003	0.91881
Space group	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$
Cell parameters [a,b,c, Å]	34.061, 47.888, 76.064	34.531, 47.603, 77.944	34.340, 48.133, 77.555
Resolution (Å)	47.89-1.70 (1.76-1.70)	47.60-1.52 (1.61-1.52)	48.13-1.15 (1.21-1.15)
Independent reflections	14,255 (1,327)	20,143 (2720)	37,927 (1,492)
Multiplicity	9.4 (6.7)	6.3 (5.8)	5.4 (2.0)
Completeness (%)	99.8 (98.5)	99.0 (93.8)	80.6 (22.4)*
<I/σ(I)>	7.4 (0.8)	27.2 (3.2)	24.9 (3.6)
R_{merge}	0.132 (0.89)	0.031 (0.468)	0.033 (0.227)
R_{pim}	0.068 (0.663)	0.016 (0.228)	0.016 (0.170)
Refinement			
Protein atoms (no hydrogen) / solvent atoms /ligand/others	1119 / 124	1108 / 138 / 0 / 1	1129 / 234/ 18
R_{cryst} / R_{free}	0.206 / 0.251	0.171 / 0.198	0.153 / 0.173
Geometry			
Ramachandran favored / outliers (%)	97.0 / 0.0	96.2/ 0.0	96.3/ 0.0
Rotamer outliers	0.8	2.5	0.0
Overall score	1.3		1.7
RMSD on bonds length [Å], angles (°)	0.012, 1.14	0.012, 1.42	0.008, 1.24

Table 2 Hydrogen bond interactions between the hydroxyl oxygen atom of retinol and protein atoms. The actual distance is reported in parentheses. In the Q108L/K40L mutant the electron density is not clear for the retinol -C15–OH group, whilst the rest of the ligand is clearly visible.

CRBP1-retinol complexes	Interaction 1	Interaction 2	Electron density
WT	Oε1 Q108 (2.92 Å)	NζK40 (3.22 Å)	Very clear for all retinol
K40L	Oε1 Q108 (2.41 Å)	---	Very clear for all retinol
Q108L*	Sδ M62 (2.77 Å)	Nζ K40 (2.57 Å)	Reasonable density for the –OH group
Q108L/K40L**	---	---	No density for the -C_{15}–OH group

* The retinol is present in double conformation
** In the deposited structure, although its electron density is not defined, the retinol –OH group is at 2.29 Å from Sδof M62 and at 2.34 Å from a water molecule, well defined and held in place by three H-bonds with the CO groups of Y60 and I51 and Oγ of T53.

6.1.4 Effect of the His-Tag on the WT CRBP1: MD simulation

By means of a 100 ns molecular dynamics (MD) simulation, we have analyzed the effect of the presence of the His-Tag on the WT apo-CRBP structure, possibly responsible for protein conformational changes. To this end, a His_6-Tag was added to the WT apo-CRBP structure. To better reproduce the acidic experimental conditions, the pKa values of the ionizable residues were calculated taking into account the local structural environment (see Materials and Methods paragraph 5.1.2). As a result, at the crystallization condition of pH 5.5, all the His residues are likely to be mostly protonated. Therefore, a MD simulation was performed on the His-tagged structure with all the His residues in a charged state. Nevertheless, to exclude that the charge of His-residues was the main cause of any enhanced mobility of the protein, a second MD simulation was carried out in the same conditions but with all the His residues in the unprotonated form. Finally, a simulation on the WT apo-CRBP1 structure without the His-Tag was also carried out as a control. From the plot of the root mean square fluctuation (RMSF) of Cα atoms (calculated on the last 70 ns of simulation, during which the protein had already achieved a stable structure), we can clearly see that the His-Tagged structure in both cases shows a flexibility greater than that of the WT control, especially affecting the portal region (Fig. 6.9). The higher flexibility of the His-Tagged structure is confirmed by the analysis of B-factors: the HT apo CRBP1 structure (Silvaroli et al., 2016) shows clearly higher values for the whole portal region as compared to the WT apo CRBP1 (Fig. 6.10).

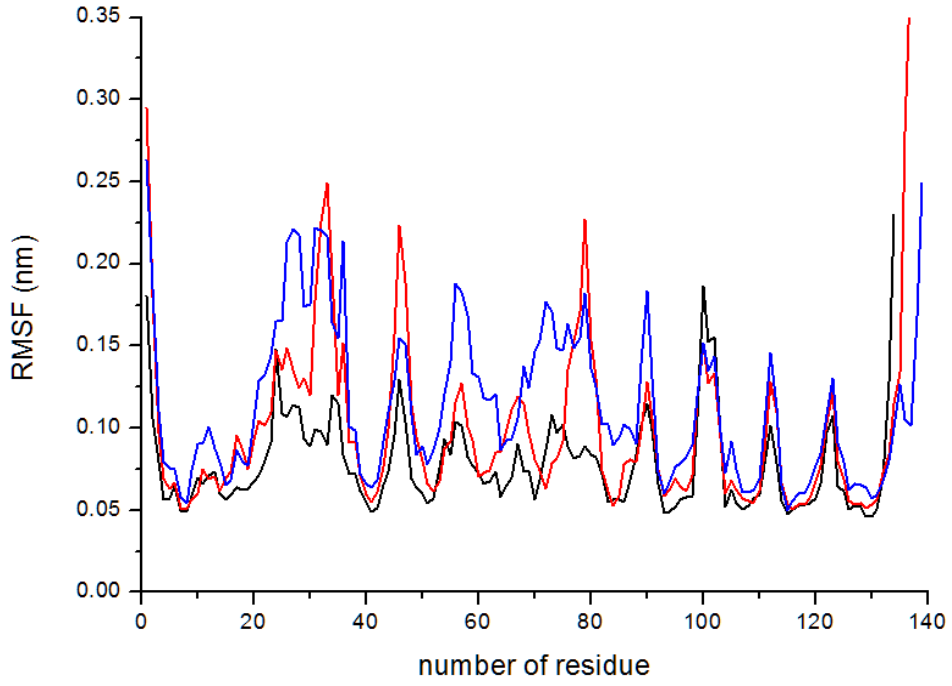


Figure 6.9 RMSF for Cα atoms of WT apo-CRBP1 without His-Tag (black line), with protonated His-Tag (blue line) and with unprotonated His-Tag (red line). RMSF values were calculated on the last 70 ns of MD simulation.

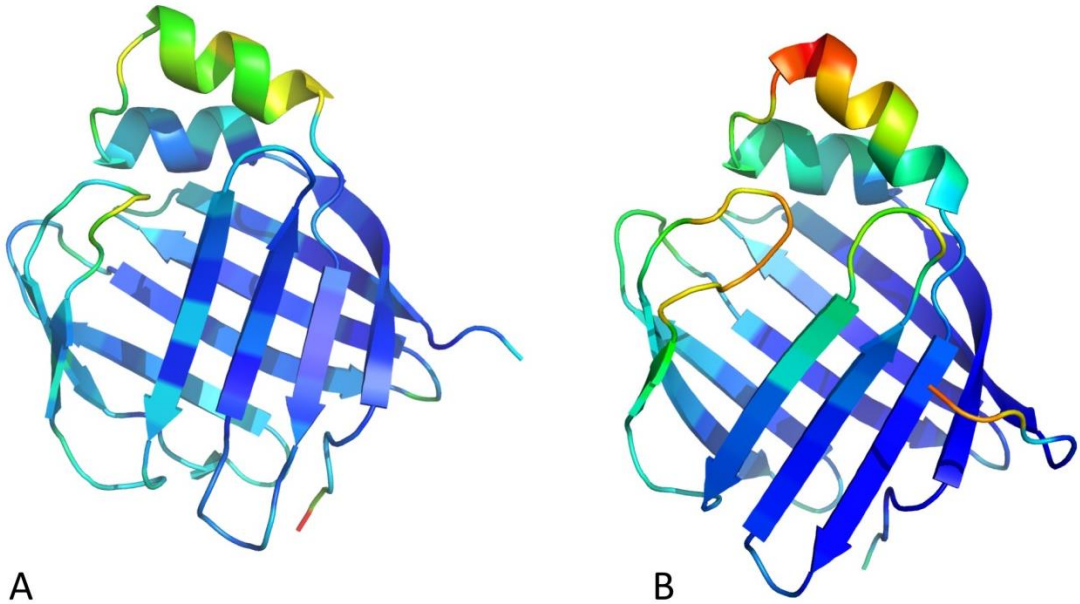


Figure 6.10 Cartoon structural models of delipidated WT apo-CRBP1 (A) and HT apo-CRBP1 (PDB ID 5H9A) (B), colored according to the temperature factors of the atoms, from blue (lowest) to red (highest).

6.1.5 Structures of mutant forms (K40L, Q108L and K40L/Q108L) of human holo CRBP1.

The structures of the holo mutant forms Q108L, K40L and K40L/Q108L of human CRBP1 are very similar to that of WT holo-CRBP1 (Fig. 6.8C), indicating that the presence of a Leu at position 40 and/or 108 does not minimally perturb the protein barrel that hosts the ligand. In fact, the RMSD

for the superposition of the equivalent Cα atoms of WT holo-CRBP1 with those of the holo mutant forms Q108L, K40L and K40L/Q108L are 0.16 Å, 0.41Å and 0.29Å, respectively.

As mentioned above, in WT holo-CRBP1 the retinol hydroxyl end group is bound in the inner part of the cavity and it forms two H-bonds, with Gln108 and Lys40. In the holo mutant forms the retinol is positioned in the inner protein cavity, occupying the same position as in the case of WT holo CRBP1. In the K40L mutant form, the retinol hydroxyl end group points towards Gln108, as in the WT protein, forming a H bond with its oxygen atom (OH – Oε1 distance 2.41Å, Fig. 6.11B). In the Q108L mutant form, the hydroxyl group is less defined, but still visible, in the electron density, since it is present in two alternate conformations: in one it points towards Met62 (O – Sδ62 distance, 2.77Å) and in the other it points towards Lys40 (O – Nζ40 distance, 2.57Å) (Fig. 6.11C). In the double mutant form K40L/Q108L, the density for the retinol moiety -CH_2-OH is not visible, indicating that this end portion of the ligand is mobile, due to the lack of any anchor for the -OH group.

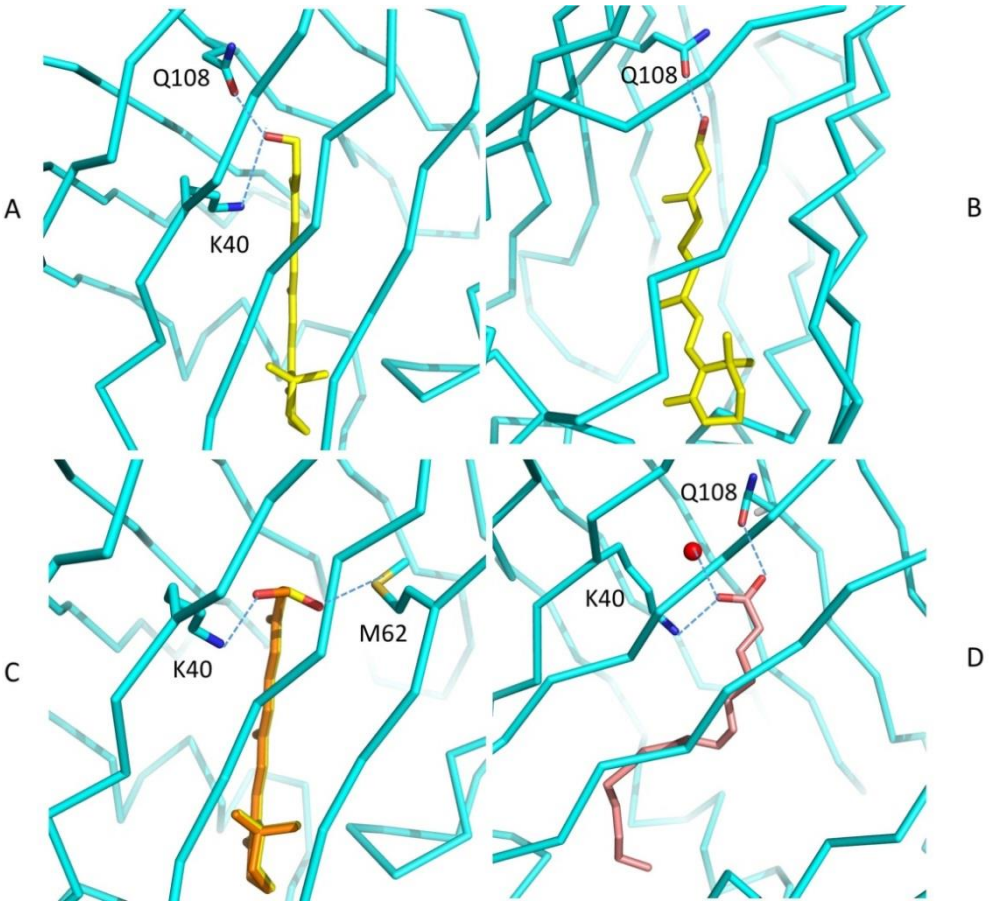


Figure 6.11 Interactions of the hydrophilic hydroxyl end group of retinol with CRBP1. A) WT CRBP1; B) K40L/CRBP1; C) Q108L/CRBP1; D) CRBP1 in complex with palmitic acid (pink). In panel C), the retinol is present in two different orientations (yellow and orange), with the –OH group pointing towards either Lys40 or Met62.

6.1.6 Thermal motions

The analysis of B-factors for the various structures we have determined is significant, since it clearly indicates that the α-helix-II, along with EF-loop and Arg58 (CD-loop), is the most flexible part of the protein, even in WT holo-CRBP1 (Fig. 6.12). This flexibility is increased in the presence of the different mutations for the holo-proteins, and in some cases it also propagates to the first helix and to some loops. The comparison of holo- and apo- WT CRBP1 (Fig. 6.12A and E, respectively) and of holo- and apo-K40L/CRBP1 (Fig. 6.12B and F, respectively) indicates that the absence of the ligand inside the cavity somehow further influences the flexibility of the protein, and in particular increases the mobility of the α-helix-II.

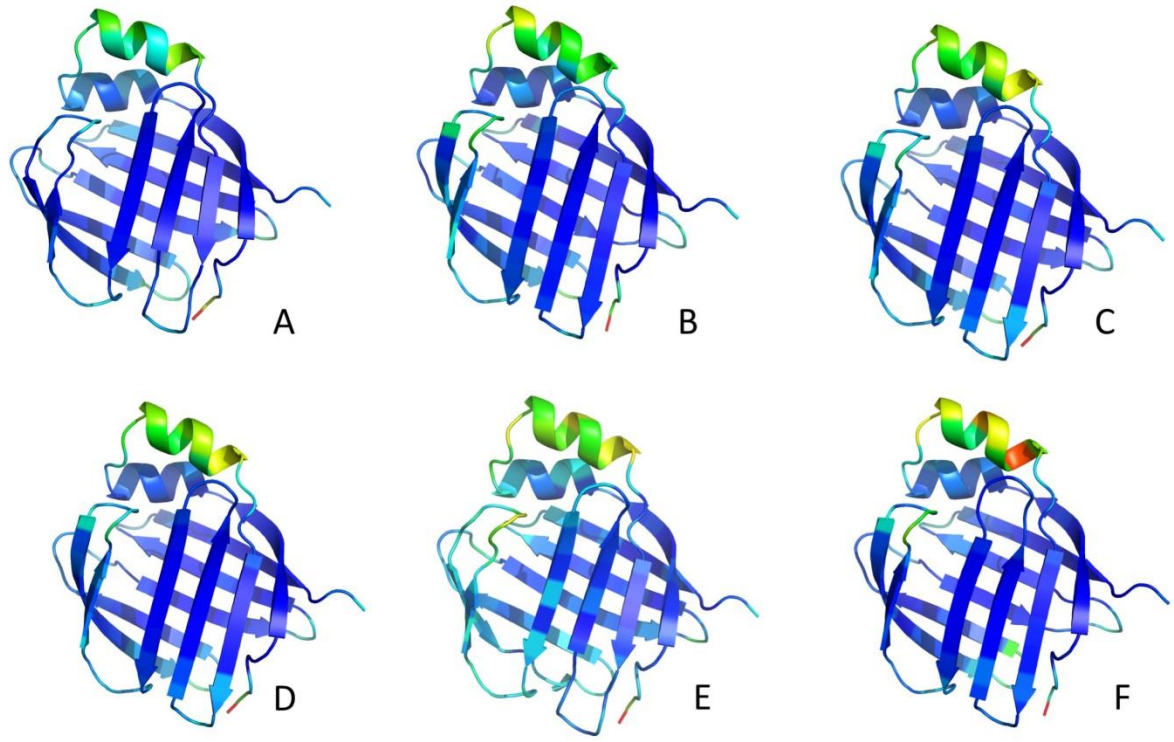


Figure 6.12 Cartoon panels showing some of the structures of CRBP1 determined here colored according to the temperature factors of the atoms, from blue (lowest) to red (highest). Models correspond to: (**A**) WT holo CRBP1; (**B**) holo K40L CRBP1; (**C**) holo Q108L CRBP1; (**D**) holo K40L-Q108L CRBP1; (**E**) delipidated apo CRBP1; (**F**) apo K40L CRBP1. Secondary structure assignment was automatically performed according to PyMOL definition (The PyMOL Molecular Graphics System, Version 1.5.0.4 Schrödinger, LLC).

6.1.7 Considerations about the key binding residues of CRBP1

The interactions of retinol bound inside the cavity of the various forms, WT and mutant forms, of CRBP1 are summarized in Table 2. The retinol in all cases is bound in the same orientation, and its -OH group interacts, when possible, with Lys40 and Gln108. When Lys40 is absent, the interaction

is limited to Gln108, whilst when the latter residue is absent the –OH is present in double conformation and it interacts with both Lys40 and Met62. In the K40L/Q108L/CRBP1 double mutant, the retinol end moiety –CH2-OH is mobile and its position is not visible in the electron density. Taken together, these data suggest that the main force for the binding of retinol inside the cavity of CRBP1 is its extensive interaction with hydrophobic residues that line the binding cavity, whilst few polar interactions established by the retinol hydroxyl group contribute to both binding affinity and specificity. The maintenance of the shape of the cavity is probably enough to hold the ligand in its position.

When the crystal structures of human WT apo and holo CRBP1 are compared, they are essentially identical. When such structures are compared with those containing a His6-Tag (Silvaroli et al., 2016), the holo-forms are highly similar, with the exception of the last residues at the C-terminus, which are influenced by the presence of the His-Tag. Instead, a striking difference between the E-F loops of the two apo-forms is observed. In this respect, it should be noted that the holo-crystals of HT CRBP1 and the apo- and holo-crystals of WT CRBP1 are nearly isomorphous, whilst the apo-crystals of HT and WT CRBP1 are not isomorphous, despite the fact that they belong to the same space group P212121. On the contrary, crystals of WT holo, apo, and mutant forms of CRBP1 are all isomorphous. Notably, NMR studies have previously shown that the structures in solution of rat WT apo- and holo-CRBP1 are also highly similar (Mittag et al., 2006), in agreement with our results on the X-ray analyses of the crystal structures of human WT apo- and holo-CRBP1.

Regarding the difference between HT and WT apoproteins, it should be pointed out that the His-Tag tail in the crystal is close to the region 57-60, one of the two areas involved in conformational differences between holo- and apo- proteins (the distance between the imidazole ring of His140 and the closest atom of Asn59 is about 5Å, and that between His139 and the closest atom of Arg52 is less than 4 Å). In regard to this point, we could hypothesize that the destabilization of this area in HT apo CRBP1 is transmitted to the nearby 75-79 loop. This destabilization is likely to be possible due to the lack of a ligand bound inside the cavity, which by itself is able to stabilize the 3D structure in the holoprotein. On the contrary, in the absence of the ligand a less rigid protein scaffold could be destabilized by the interaction of the protein surface with the His-Tag tail. His residues are likely to be mostly protonated at the acidic pH of the crystallization medium (Silvaroli et al., 2016). However, the detection of a greater structural mobility also in the simulation with unprotonated His-residues, indicates that the presence of the His positive charges is not

responsible for the higher flexibility. The results of molecular dynamics simulations also reveal interactions between the His-Tag with the protein surface that lead to a destabilization of the portal region. Examples of other proteins are present in the literature indicating that the His-Tag may have an impact on critical properties of proteins ((Randolph, 2012), and references therein). Interestingly, in a previous study the high resolution crystal structures of WT apo and holo human cellular retinoic acid binding type II (CRABPII) were found to be very similar, while the crystallization of the His6-tagged protein resulted in poorly diffracting crystals (Vaezeslami, Mathes, Vasileiou, Borhan, & Geiger, 2006).

The retinol molecule is deeply buried inside its binding cavity in CRBP1. This feature ensures the complete shield and protection of the ligand from the external environment. Since the superimposition of apo- and holo-structures of WT CRBP1 reveals for both a closed conformation and only minimal changes between them, as-yet-uncharacterized conformational changes affecting the putative portal region are expected to occur upon retinol uptake and release. The analysis of thermal motions, based on crystallographic B-factors, indicates for WT CRBP1 a high flexibility of the region comprising α-helix-II and CD- and EF-loops, especially in the case of the apoprotein. Such flexibility may account for the occurrence of transient conformational changes affecting the portal region, necessary to allow the retinol to access the CRBP1 binding cavity as well as its release from the cavity.

6.2 MD simulations of the Apo-forms

With the aim to clarify the differences between the four human CRBPs and to identify putative regions involved in the up-take mechanism, we have carried out four MD simulations in standard conditions, using as starting coordinate the crystallographic apo-structures of CRBP1, 2, 3 and 4.

6.2.1 Molecular dynamics simulation of apo-CRBP1

We have performed a 100 ns molecular dynamics simulation on the wt apo-CRBP1 structure to investigate its flexibility. As we can see in figure 6.13, where we have reported the plot of the secondary structure during the simulation, we can’t observe significant changes: the beta structure is highly stable and even the alpha helices preserve their structure all the time.

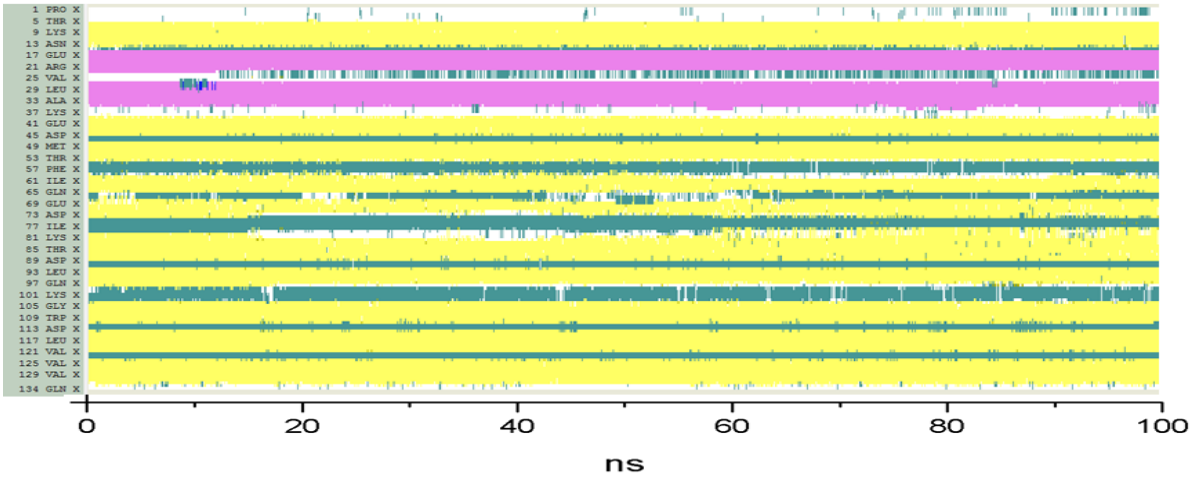


Figure 6.13 Secondary structure plot of hCRBP1 during 100ns MD simulation, in standard condition. We can't observe significant changes: both alpha helices and beta regions are conserved during the simulation.

The Root Mean Square Fluctuation (RMSF) calculation allows the identification and the analysis of residues flexibility (see below, Fig 6.28). With this method, we can confirm what already observed by NMR studies of Franzoni and co-workers (Franzoni et al., 2002), that the residues forming the already known "portal region, constituted by the CD loop, the EF loop and the alpha helix II, have the highest flexibility, with respect to the whole backbone (apart from some short connection loops at the opposite end of the barrel). This behavior is visible also in figure 6.14 in which we have superimposed 17 frames of the CRBP1 trajectory, saved every 5 ns of simulation (this low sampling was used only in this figure, for the sake of clarity).

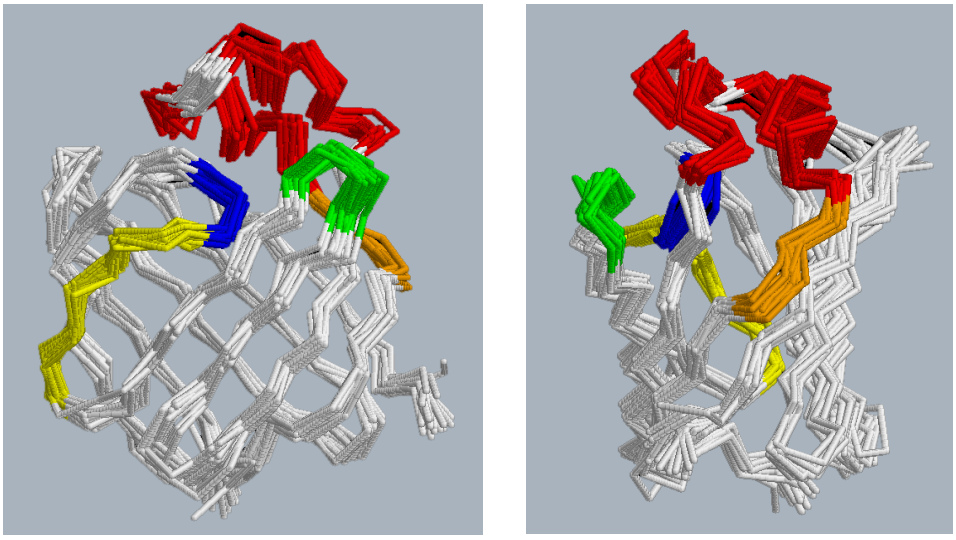


Figure 6.14 Superimposition of 17 CRBP1 structures, saved every 5 ns from the MD simulation trajectory (this low sampling was used only in this figure for the sake of clarity). The two images illustrate two different orientations of the same protein. Relevant regions were colored with different shades: alpha helices (residues 15-24 and 27-35) in red, CD loop (residues 55-58) in green, EF loop (residues 75-78) in blue, βE (residues 68-74) in yellow and the ending part of βA in orange.

Regarding the two key residues of the binding site, our data are again in agreement with the previously mentioned NMR studies (Franzoni et al., 2002). Gln108 has a very low mobility, due to the localization of NH2 group close to the center of the Phe4 ring; this amino-aromatic interaction probably allows the correct positioning of Gln108 in anticipation of the arrival of the ligand. Lys40, unlike Gln108, is a quite flexible residue. Monitoring the distance between Cα and Nz of Lys40 during the MD, we have observed that fluctuation ranges from 4.5 to 6.5 Å (Fig. 6.15): the long chain of this residue permits a spring behavior that may be necessary for the capture of the hydroxyl group of retinol or to attract its poly-isoprene tail, during the ligand uptake.

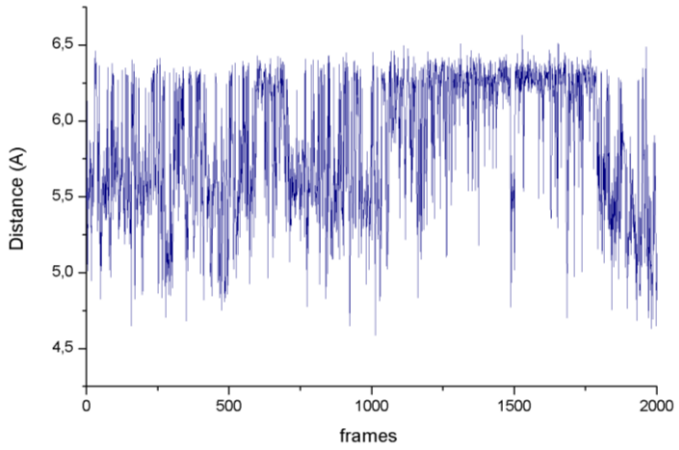


Figure 6.15 Distance plot between carbon alpha of Lys40 and its Nz atoms. We have observed a fluctuation range around 4.5-6.5Å.

Observing the crystal structures of CRBP1 and 2, we can see that Tyr60 can assume two alternative positions which can obstruct or not the binding cavity. We have monitored the behavior of this residue during the dynamics on CRBP1, observing the presence of a persistent H-bond between Tyr60 and Glu72; only in sporadic cases Tyr60 is directed towards the inner of the cavity in a position that obstructs the binding pocket. However, it is worthwhile to observe that the position in which the Tyr60 is bound to the Glu72, even if keep the cavity space accessible, does not affect the opening of the entrance to the binding site, that seems to be led from another residue, namely Phe57. Phe57 is part of a hydrophobic cluster, formed by six residues belonging to alpha helix II (Ala28, Leu29, Ile32, Ala33, Leu35 and Leu36) plus one residue of EF loop (Ile77), that are located in the “portal region” of hCRBP1 (Mittag et al., 2006) (Ragona et al., 2014). This cluster remains well compact during the whole simulation with the exception of Phe57, which in some timepoints leaves the hydrophobic cluster. We have identified a persistent cation-π interaction between Phe57 and Arg58. When this interaction is lost, Phe57 can bend outward from its position,

creating a little opening on the protein surface (Fig. 6.16). When the cation-π interaction is present, the two residues of Phe57 and Arg58 form a molecular cap which obstructs the opening (Fig. 6.17). Even if the persistence of the open conformation is statistically low, we will see that it will be important for the ligand uptake.

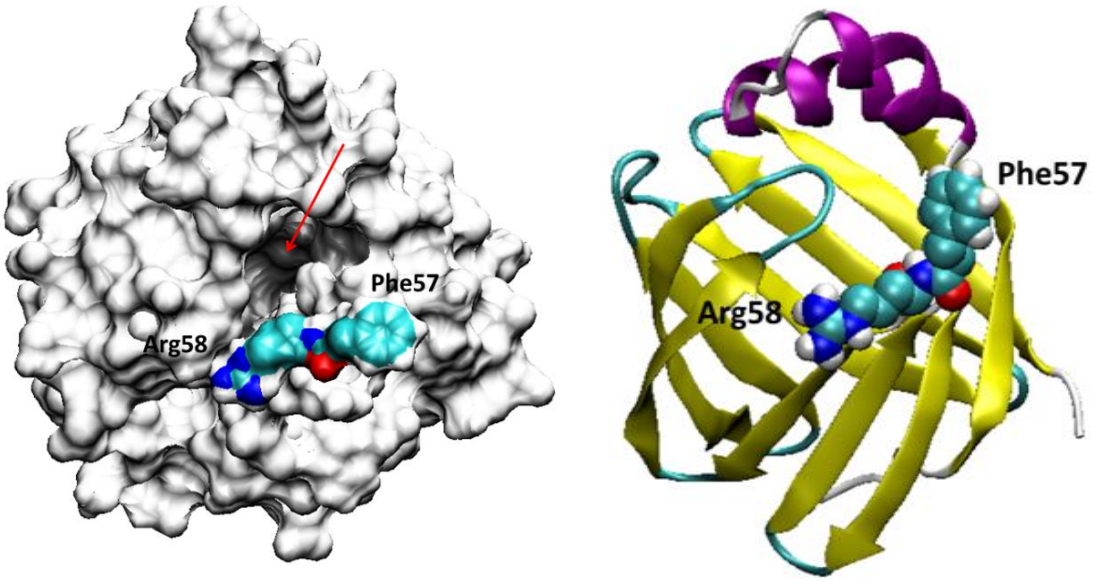


Figure 6.16 The molecular surface of hCRBP1 WT is shown on the left: a little opening, indicated with a red arrow, is visible on the surface and due to the re-positioning of Phe57 and Arg58, which are colored by atom type. On the right we have reported the same situation in a cartoon representation; Phe57 and Arg58 are shown in spacefill and colored by atom type.

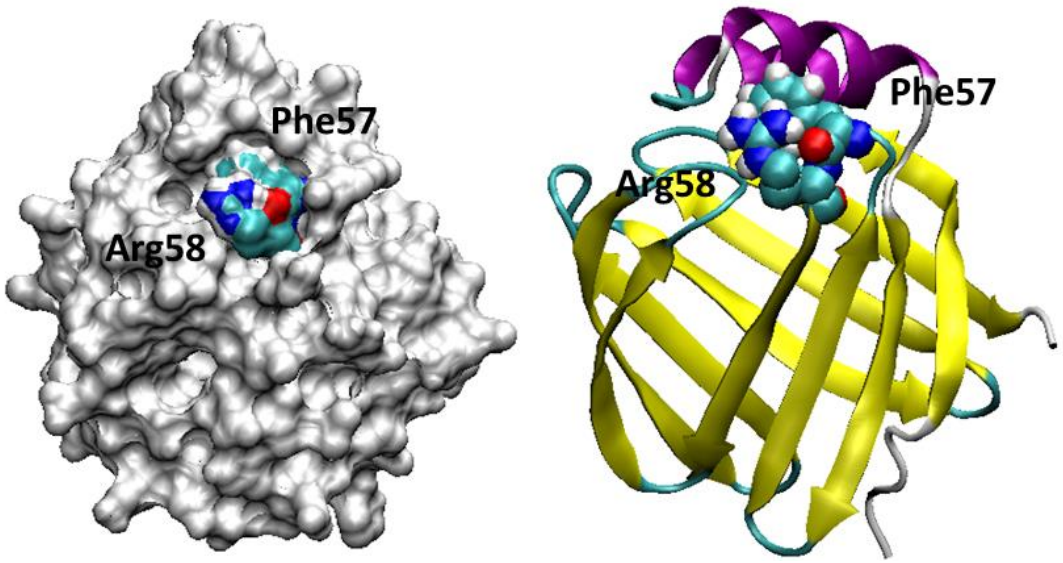


Figure 6.17 The molecular surface of hCRBP1 WT is shown on the left: the opening is closed by the "molecular cap" formed by Arg58 and Phe57 (colored by atom type). On the right we have reported the same situation in the cartoon representation; Phe57 and Arg58 are shown in spacefill and colored by atom type.

Among the most important structural features of intracellular LBPs there is the so-called "gap region" (see the Introduction section), an interruption in the beta barrel due to a large distance between the D and E strands that prevents the formation of H-bonds (Franzoni, Cavazzini, Rossi, & Lücke, 2010). Monitoring the distance between the carbonyl oxygen atoms of residues 61 and 73, to observe the gap opening during the MD, we can see that its great fluctuation (from a minimum of 3.7 Å to a maximum of 12 Å, Figs. 6.18-6.19), starting just after 20 ns, is not correlated with the opening of the binding cavity observed above.

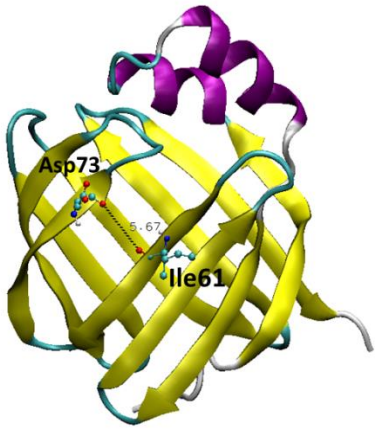


Figure 6.18 Cartoon representation of hCRBP1: residues 61 and 73, belonging to the "gap region", are shown in ball and sticks and colored by atom type.

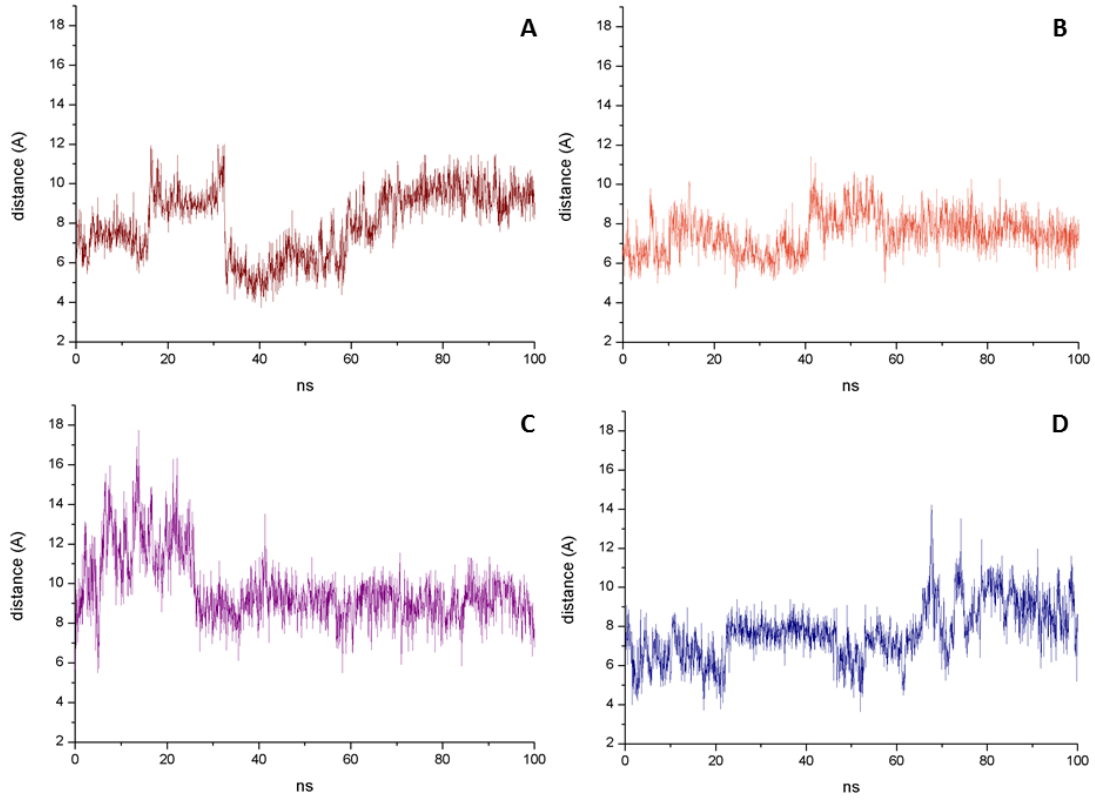


Figure 6.19 Plot of the distances between carbonyl hydroxyl atoms of residue 61 and 73: (**A**) for CRBP1, (**B**) for CRBP2, (**C**) for CRBP3 and (**D**) for CRBP4. The major enlargement of the gap region is observed in CRBP3 and 4.

The upper part of the gap region is stabilized by the above mentioned H-bond between Tyr60 and Glu72. However, this region could be a point of flexibility of the protein and could be of help to the passage of the ligand through the portal region.

6.2.2 Molecular dynamics simulations of apo-CRBP2

The same strategy used for CRBP1 was applied to CRBP2, performing a 100 ns MD simulation on the apo-protein. As we can see below in figure 6.28, where the RMSF of the carbon alpha atoms of the protein is plotted, CRBP2 shows quite high rigidity as in the case of CRBP1. The secondary structure plot, shown in figure 6.20, highlights again a remarkable stability of the beta barrel with the exception of the βA terminal part (red rectangle), the first part of helix I and their linker region. Helix I seems to be more flexible in CRBP2 that in the isoform 1. The flexibility of the βA terminal part is evident also in the RMSF plot (see below Fig. 6.28) (Franzoni et al., 2010).

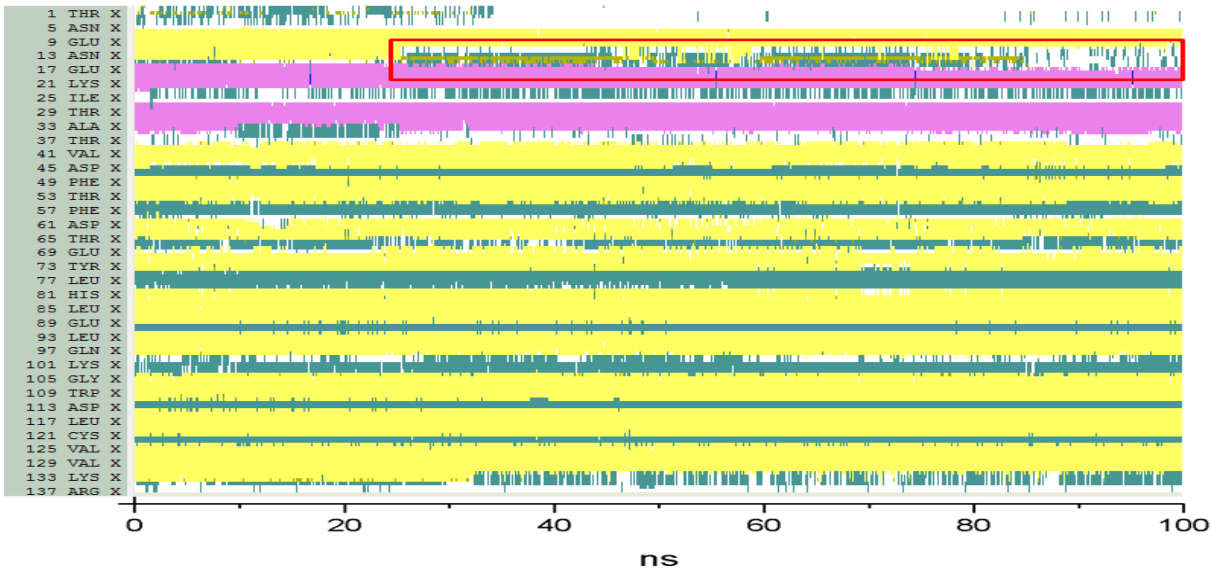


Figure 6.20 Secondary structure plot of hCRBP2 during 100 ns MD simulation, in standard condition. The beta barrel is clearly stable during the simulation with the exception of the ending of beta A (squared in red).

The calculation of the RMSF, see below in figure 6.28, indicates that the alpha helices region is the most flexible part of the protein. In figure 6.21 we show, as for CRBP1, the superimposition of 17 structures saved every 5 ns from the trajectory file of the MD simulation. We can observe also in this case the quite high rigidity of the beta barrel, while the alpha helices region and the ending part of βA (colored in orange) seem to be affected by greater flexibility.

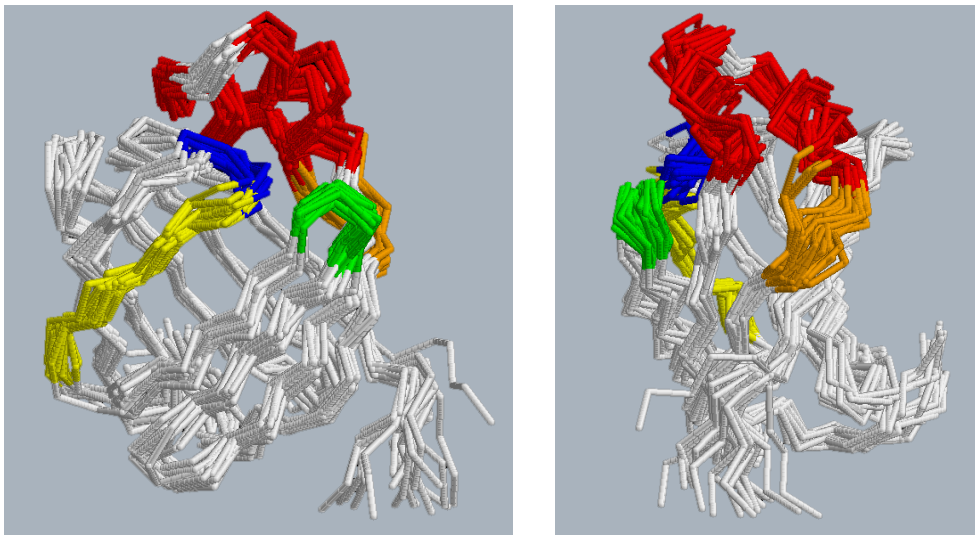


Figure 6.21 Superimposition of 17 CRBP2 structures saved, every 5 ns, from the trajectory obtained after the MD simulation. The two images illustrate two different orientation of the same protein. Important regions were colored with different shades: alpha helices (residues 15-24 and 27-35) in red, CD loop (residues 55-58) in green, EF loop (residues 75-78) in blue, βE (residues 68-74) in yellow and the ending part of βA in orange.

Contrarily to CRBP1, in the simulation performed on apo-CRBP2 we can't observe the formation of any opening on the protein surface.

The distance between the carbonyl oxygen atoms of residues 61 and 73 (Fig. 6.19) belonging to the gap region, calculated during the MD, ranges from a minimum of 4.7 Å to a maximum of 11.4 Å, indicating a lower fluctuation and a smaller enlarging of this region if compared with CRBP1.

6.2.3 Molecular dynamics simulations of apo-CRBP3

The analysis of the 100 ns MD simulation performed on apo-CRBP3 showed a different behavior with respect to the previous ones, which is characterized by a higher flexibility. In figure 6.28 below and in figure 6.22, where we show the superimposition of the 17 backbone structures as explain previously, is clearly notable a widespread flexibility which affects, in particular, the EF loop and the beta strand E. On the contrary, alpha helices region and the ending part of beta strand A seem to be quite stable during the simulation similarly to CRBP1 and despite the major general mobility of the whole structure.

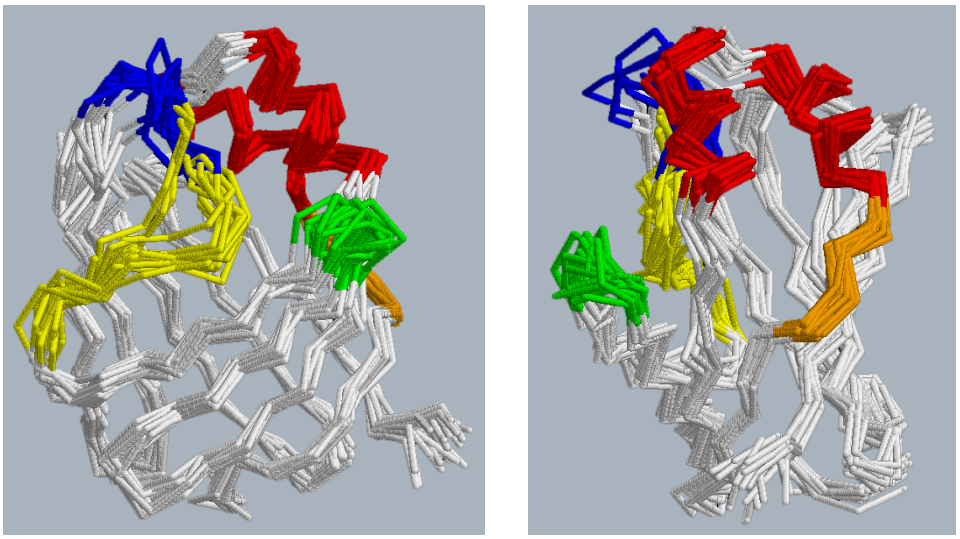


Figure 6.22 Superimposition of the backbone structures of CRBP3, taken from the trajectory as explained above. The two images illustrate two different orientation of the same protein. Important regions were colored with different shades: alpha helices (residues 15-24 and 27-35) in red, CD loop (residues 55-58) in green, EF loop (residues 75-78) in blue, βE (residues 68-74) in yellow and the ending part of βA in orange.

Monitoring the distance in the gap region (between the carbonyl oxygen atoms of residues 61 and 73) we can observe a greater fluctuation (5.5-17.74 Å) than what we observed in apo-CRBP1 and 2, but limited to the first part (6 to 25 ns) of the simulation (Fig. 6.19). During this time is clearly visible a big opening on the protein surface (Fig. 6.23) in such a way that the inner cavity becomes highly exposed to the solvent, including the two key residues (His108 and Lys40). The opening is only temporary and, from about 26 ns to the end of the simulation, D and E beta strands close again the gap, remaining quite stable. However beta strand E and F are highly flexible; the secondary structures plot, shown in figure 6.24, indicates that the ending part of strand E and the initial part of strand F assume a coiled structure, enlarging the EF loop (Fig. 6.19).

These data suggest the possibility that CRBP3 could host a ligand sterically bulkier than retinol.

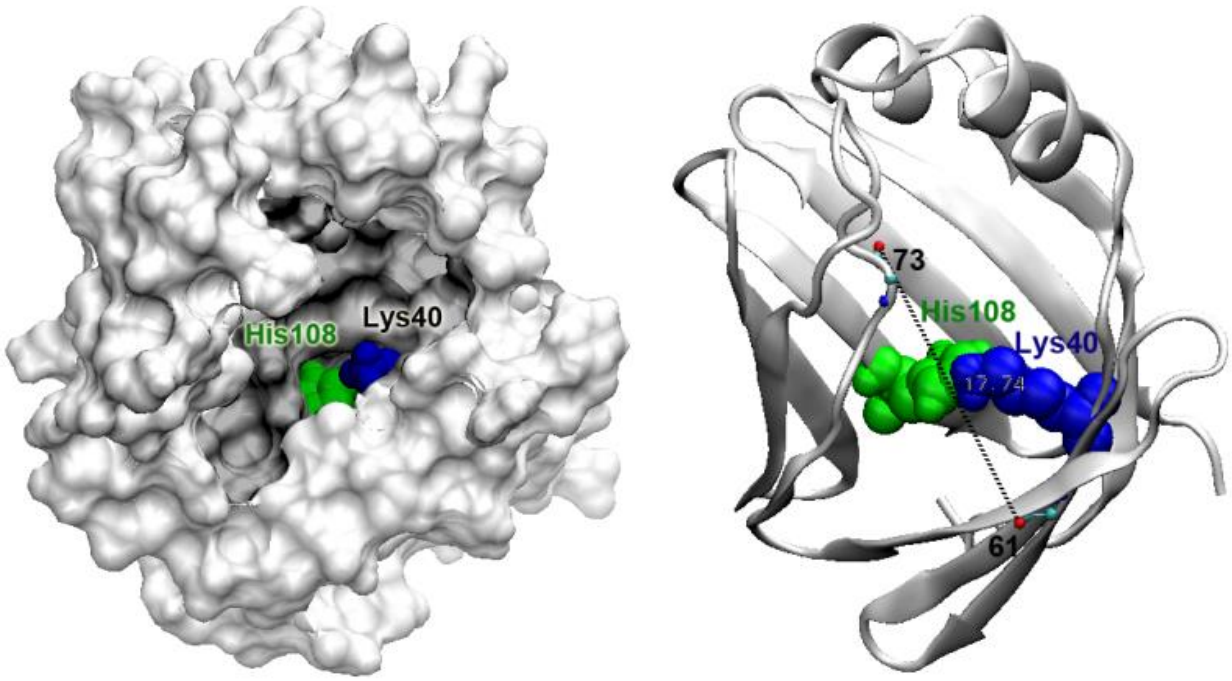


Figure 6.23 The molecular surface of hCRBP3 is shown on the left: the formation of a big opening exposes the key residues (His108 and Lys40) to the solvent. On the right we have reported the same situation, with the protein in a cartoon representation; Lys40 (blue) and His108 (green) are shown in spacefill.

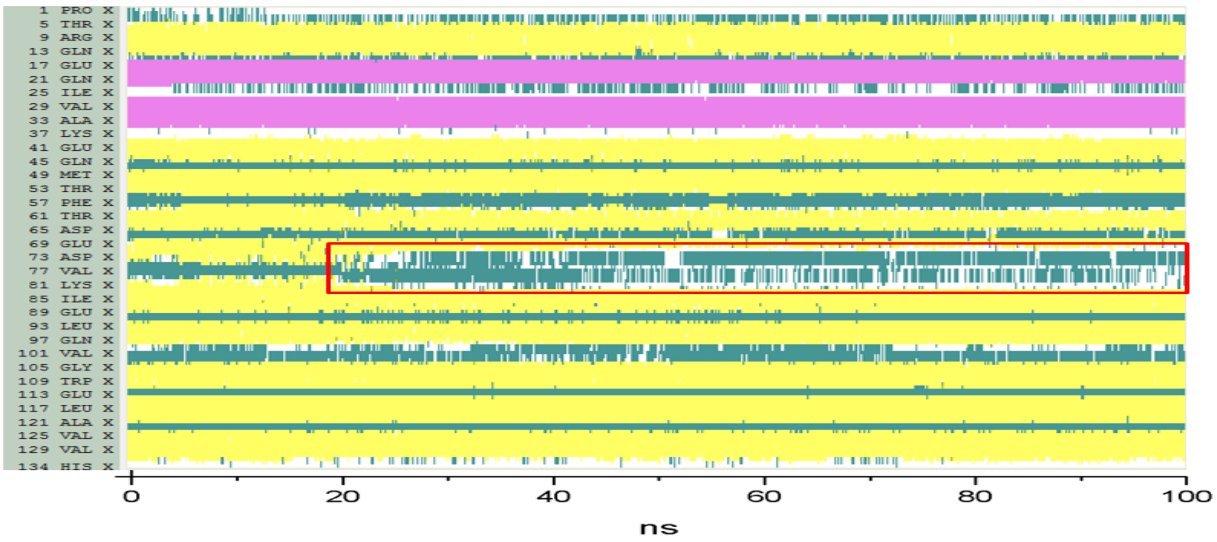


Figure 6.24 Secondary structures plot of hCRBP3 during the 100ns MD simulation, in standard condition. In red we have highlight the ending part of strand E and the initial part of strand F: this region assumes a coiled structure since 20ns until the end of the simulation.

6.2.4 Molecular dynamics simulations of apo-CRBP4

The RMSF analysis (see below Fig. 6.28) performed on the MD trajectory of apo-CRBP4 has revealed that the alpha helices region is the very most flexible part of the protein. In figure 6.25 the secondary structure plot is shown; we can notice the high stability of the beta barrel even for

the ending part of beta strand A which in CRBP2 was instable, while the helices region is affected by major changes.

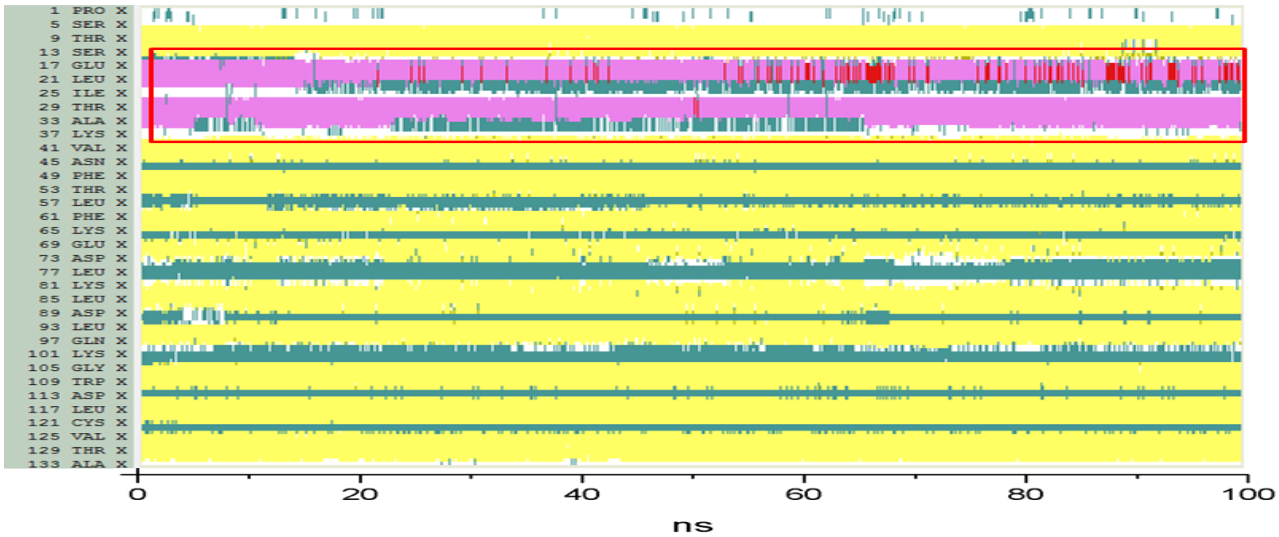


Figure 6.25 Secondary structure plot of hCRBP4 during 100ns MD simulation, in standard condition. In red we have highlight the alpha helices region which show the lowest stability.

The superimposition of backbone structures collected during the MD, shown in figure 6.26, highlights a general flexibility of the whole structure, also greater than CRBP3. Apart from helices, we can observe a significant flexibility even in CD loop, while beta strand E seems to be relatively stable.

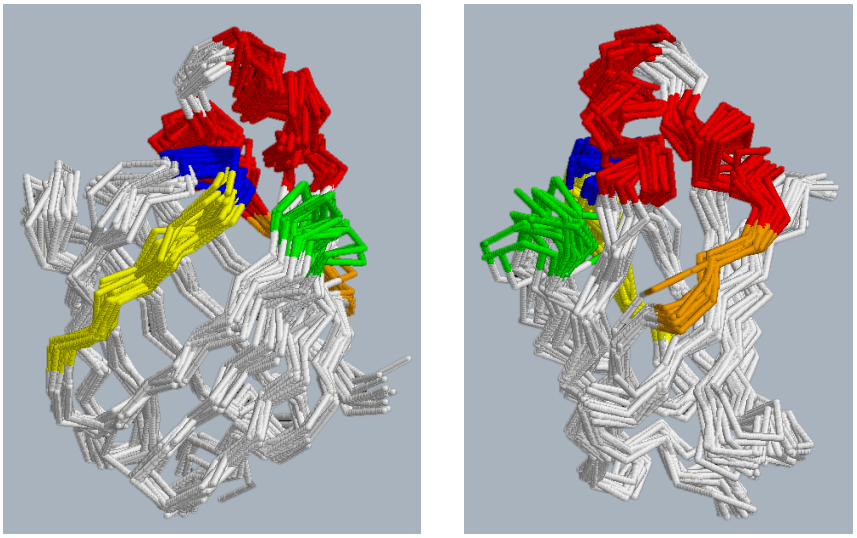


Figure 6.26 Superimposition of backbone structures collected every 5 ns over the 100 ns of simulation performed on CRBP4. The two images illustrate two different orientation of the same protein. Important regions were colored with different shades: alpha helices (residues 15-24 and 27-35) in red, CD loop (residues 55-58) in green, EF loop (residues 75-78) in blue, βE (residues 68-74) in yellow and the ending part of βA in orange.

The distance of the gap region between strand D and E lays in a range of 3.6-14.2 Å (Fig. 6.19).

This fluctuation is comparable to that recorded for CRBP1. Even if we can't observe the formation of a big enlargement of the gap region like CRBP3, the high mobility of CRBP4 induces the formation of several openings on the protein surface. In fact, even in the crystal structure are already present some openings, positioned in the gap region, in the alpha helices portion and at the portal region; this characteristic is probably due to the high flexibility of this protein. The biggest opening is visible after about 75 ns of simulation and is due to the movement of CD loop, which bends out, and to the convergence of alpha helix II to alpha helix I. These changes, affecting the portal region, cause the exposure of the inner cavity to the solvent, included the two key residues (Fig. 6.27). However this opening doesn't correspond to the maximum distance recorded in the gap region.

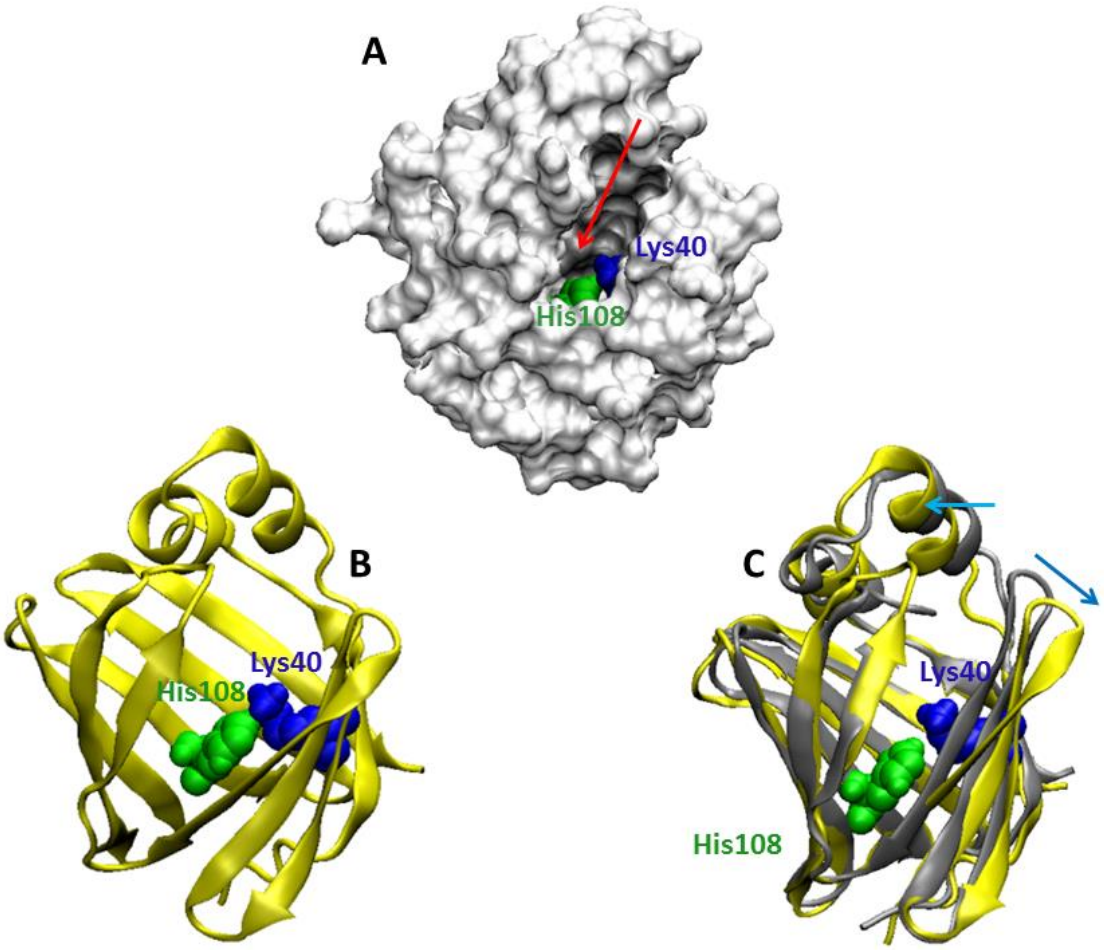


Figure 6.27 In panel A, the fully open structure of CRBP4 represented as molecular surface. His108 (green) and Lys40 (blue) are represented in spacefill. In panel B, the same configuration is shown in cartoon representation. In panel C is shown an overlap of initial coordinates (grey) of CRBP4 whit the fully open conformation (yellow): we can observe changes in the position of CD loop and alpha helix II.

6.2.5 Considerations about the MD simulations of apo-proteins

The MD simulations analyses have revealed important characteristics which are different between the four isoforms.

CRBP 3 and 4 are markedly more flexible than CRBP 1 and 2, and distinct regions show different mobility. In figure 6.28, the RMSF plot of the cα atoms clearly highlights these differences.

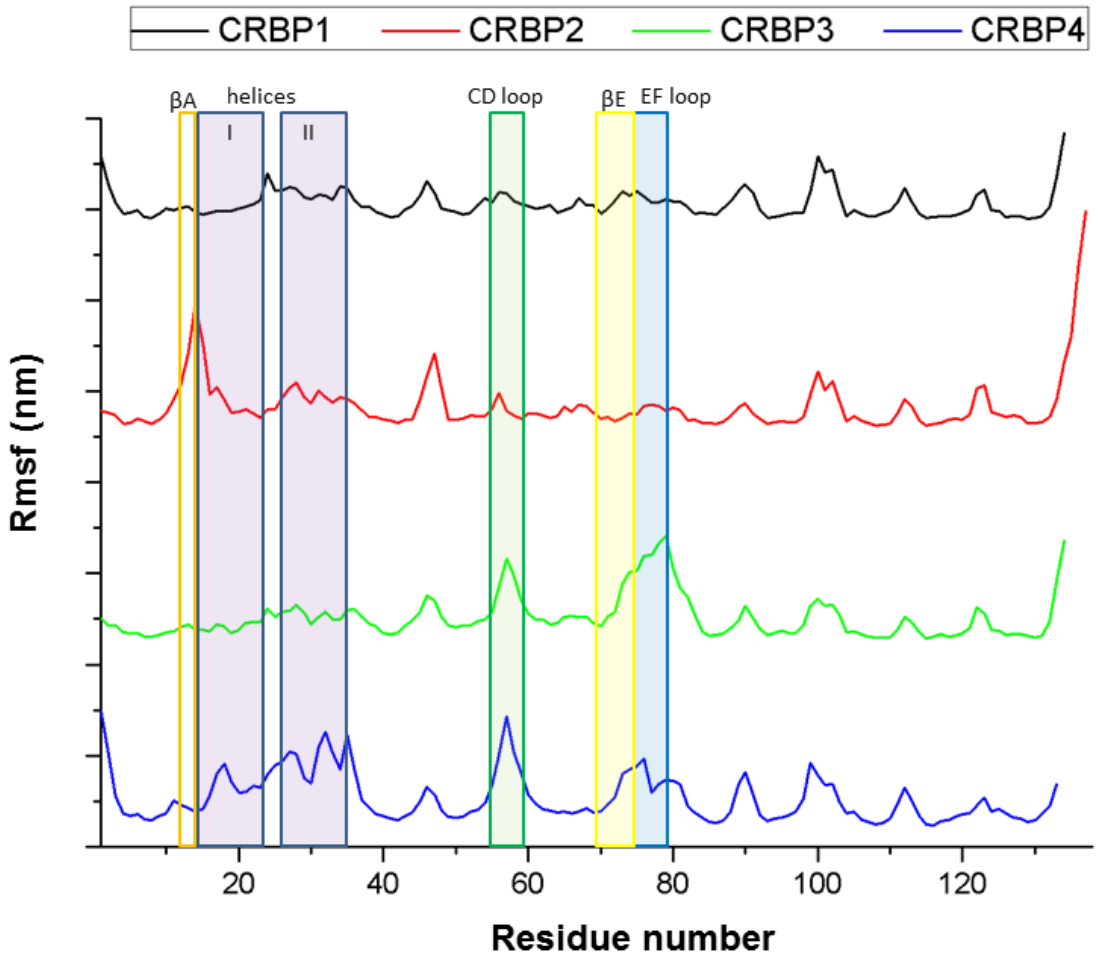


Figure 6.28 The plots show the RMSF of the Cα atoms, relatively to the four isoforms. Remarkable protein regions are highlight with rectangles: strand A in orange, the alpha helices region in purple, CD loop in green, strand B in yellow and the EF loop in blue. Each vertical ticks corresponds to 0.2 nm.

CRBP1 is the most rigid protein and no particular regions seem to be affected by elevate mobility, with the exception of GH loop, even if (apart from short connection loops) the portal region is the most flexible part of the protein as suggested by NMR studies performed on apo- rat CRBP1 (Franzoni et al., 2010). However, we have identified a little opening, discussed above, due to the concerted movements of the side chains of Phe57 and Arg58; the formation of this opening doesn't seem to involve larger movement of structural elements. Despite a major diffuse

flexibility, CRBP3, like CRBP1, shows a low mobility of the helices. Instead, the beta strand E and the EF loop, together with the CD loop, show the highest RMSF values. In fact, the analysis of the MD simulation has revealed that the enlargement of the gap region led to the formation of a big stable opening, suggesting the possible ability of CRBP3 to bind a ligand bigger than retinol. In CRBP2, the beta strand A and the helix-turn-helix region (in particular alpha helix I) are the most flexible part; these data are in agreement with what observed by Franzoni and co-workers in their NMR study performed on apo- rat CRBP2 (Franzoni et al., 2010). Also CRBP4 shares high mobility in the helix-turn-helix region, even if not around the beta strand A; in addition, in CRBP4 a great flexibility affects also all the relevant regions highlighted in the other isoforms, i.e. the CD loop, the beta strand E and the EF loop. This behavior leads to the formation of several different openings during the dynamics, with the biggest one concerning the portal region.

A phylogenetic study of iLBP family, performed by Schaap and co-workers (Schaap, van der Vusse, & Glatz, 2002), revealed that hCRBP4 is related more closely to CRBP2, while CRBP1 seems to be evolutionary more related to CRBP3. However, the four isoforms derive from the same ancestral gene duplication event. The study of the flexibility which we have carried out confirms this trend, even if it seems that CRBP3 and 4 are affected by a major flexibility than isoforms 1 and 2.

6.3 MD simulations of retinol up-take

The iLBPs up-take mechanism is a broadly debated topic. Several NMR studies were performed on different members of this family, but only in few cases this studies provided clear evidence about the entrance of ligands. In particular on human CRBPs are available only NMR studies on rat isoforms 1 and 2; no works on CRBP3 and 4 concern their up-take mechanism. MD simulations techniques are the best approach to investigate this process at a molecular level. Therefore, we carried out MD in the presence of retinol to provide new insights about the entrance of the ligand into the binding cavity of CRBPs. On these bases, experiments were then set up to confirm the simulation data.

A number of simulations were performed before defining the optimal conditions to see the entrance of retinol. Due to its lipophilic characteristics, retinol prefers hydrophobic regions on the protein surface. CRBPs have many hydrophobic patches on their surface and this feature causes the aspecific contact of retinol with these regions. The ligand, after the contact with these surfaces, remains for a long time trapped during a simulation. Moreover, a combination between a well-defined “open” protein conformation and a suitable orientation of the ligand in close proximity of the entrance are needed. Substantially, the ligand must be in the up-take region of the protein when the conformation of this region is favorable. While *in vitro* we have a lot of protein and retinol molecules, in a MD simulation only one protein and one ligand are present, decreasing the probability of occurrence of a mutual favorable position. We also tried to simulate more than one ligand molecules at the same time, arising from an MD work on a liver FABP (Tsfadia et al., 2007), but with no results.

One of the methods used in MD simulations to accelerate the atom motions consists to increase the temperature (Hu & Liu, 2014); in this way we also supply energy to the system to overcome possible potential barriers that trapped the ligand at the surface. Therefore, we carried out MD simulations on human CRBPs at 350K, in the presence of one retinol molecule (see Materials and Methods). In less than 40ns the up-take of the ligand was observed.

6.3.1 MD simulation of CRBP1 in the presence of retinol

The MD simulation of CRBP1, after the addition of one retinol molecule, revealed important insights about the up-take mechanism. The possibility to compare our data with NMR data performed on rat CRBP1 allowed to propose a model for the ligand up-take entrance into the binding site. In figure 6.29 is shown a sequence of snapshots which describes the fundamental steps of ligand entrance. Retinol, positioned at the start as shown in image A (see Material and methods), is moved by the peculiar residues distribution presents on the CRBP1 surface, positioning near the portal region (Fig. 6.29B).

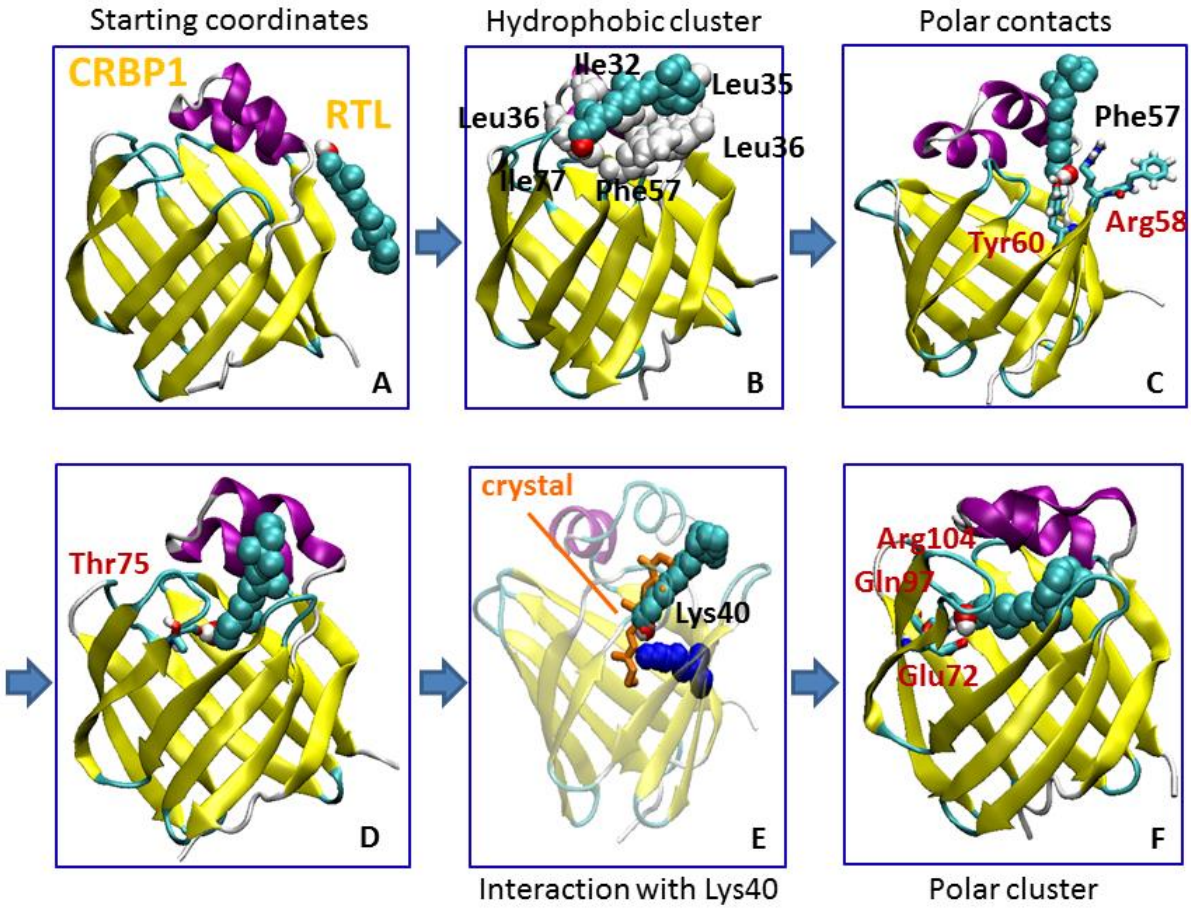


Figure 6.29 Snapshots A-D describe the entrance of the retinol into CRBP1 binding cavity. The protein is represented in cartoon; retinol, colored by atom types, and the hydrophobic residues (white) are shown in Van der Waals representation, while polar residues are in sticks and colored by atom types. In panel E the Lys40, one of the two key residues, is shown in Van der Waal representation and colored in blue; in this image we have colored in orange and represented in sticks the retinol in the crystal position.

This region of CRBP1 is rich in hydrophobic residues, which form a cluster with which the long hydrophobic tail and ring of the retinol can interact. The opening made by the displacement of Phe57 and Arg58, that we had identified in the apo-simulation, is already present at this time; driven by polar contacts between Arg58 itself, Tyr60 and the retinol hydroxyl group, the ligand

enters into the cleft and then, dragged by other polar residues (Thr65 for example, Fig. 6.29D) like climbing down a ladder, penetrates into the binding site. Subsequently, retinol contacts one of the two key residues, Lys40, but it fails to descend at the bottom of the cavity. In the crystal structure, in fact, the ligand interacts also with Gln108, while in the MD simulation (Fig. 6.29E), it is unable to contact both the key residues, preferring a horizontal position in which the hydroxyl group interacts with a polar cluster composed by Glu72, Gln97 and Arg104. Even if the ligand continuously switches from this horizontal orientation to a more vertical one interacting with Lys40, the first is statistically more populated.

Despite the final position of retinol isn't that of the crystal structure, a lot of information deriving by NMR studies on CRBP1, confirm our up-take model. Franzoni and co-workers (Franzoni et al., 2002) proposed a role for Phe57 in the control of the access to the binding cavity. Moreover, in the same work it was noticed that a concerted movement of both Phe57 and Arg58 would enlarge the opening of the portal region; such side-chain movements, together with little changes in the CD and EF loops, could probably be sufficient to allow retinol to enter the binding cavity. All our data confirm these observations and are in disagreement with other hypothesis, which suggested that the enlargement of the gap region of CRBP1 is needed for the up-take mechanism (Aalten, Findlay, Amadei, & Berendsen, 1995). However, an experimental validation of this mechanism is in progress in our laboratory by means of the design of selected mutants.

6.3.2 MD simulation of CRBP2 in the presence of retinol

Contrarily to what expected, the MD simulation of CRBP2 in the presence of retinol showed a different site of entrance. In figure 6.30 are summarized the principal events of the retinol up-take. Despite at the start of the simulation we have positioned the ligand near the portal region (Fig. 6.30A), retinol prefers to move itself in a different region, i.e. between the two alpha helices. The basic residues R35 and K31 drag the -OH group towards the cleft between the two helices, where retinol finds a charged cluster around -OH and a hydrophobic floor for its tail and its ring. Due to the movement of the helices, behavior that we have already seen in the apo-MD simulation, a number of hydrophobic residues belonging to the alpha helices become exposed to the solvent, establishing an optimal surface for the retinol interaction. In fact, the ligand sinks in this hydrophobic "matrix" and as soon as the first polar residues, namely Gln38 and Thr56, can

come into contact with the hydroxyl group of retinol, the ligand rises in a vertical position. Only when Phe16 shifts apart, retinol can enter into the binding cavity. Other polar residue, like Tyr19, Asn16, Gln38, Thr53 and Lys40 drag the ligand in depth (Fig. 6.30D, 6.30E). Lys40, one of the two key residues, seems to be important to capture the retinol alcoholic group, dragging the ligand at the bottom, where it can interact also with Gln108. Finally, the ligand is able to retrieve a position similar to that observed in the crystal structure (Fig. 6.30F). A horizontal position is also occupied by retinol, which come into contact with the same polar cluster (Glu72, Gln97, Arg104) mentioned for CRBP1. However, in this case, the crystal position seems to be statistically more populated. Although the passage of retinol through the helix-turn-helix cleft wasn't an expected result, many other simulation attempts were performed in the same conditions, changing the starting position of the ligand.

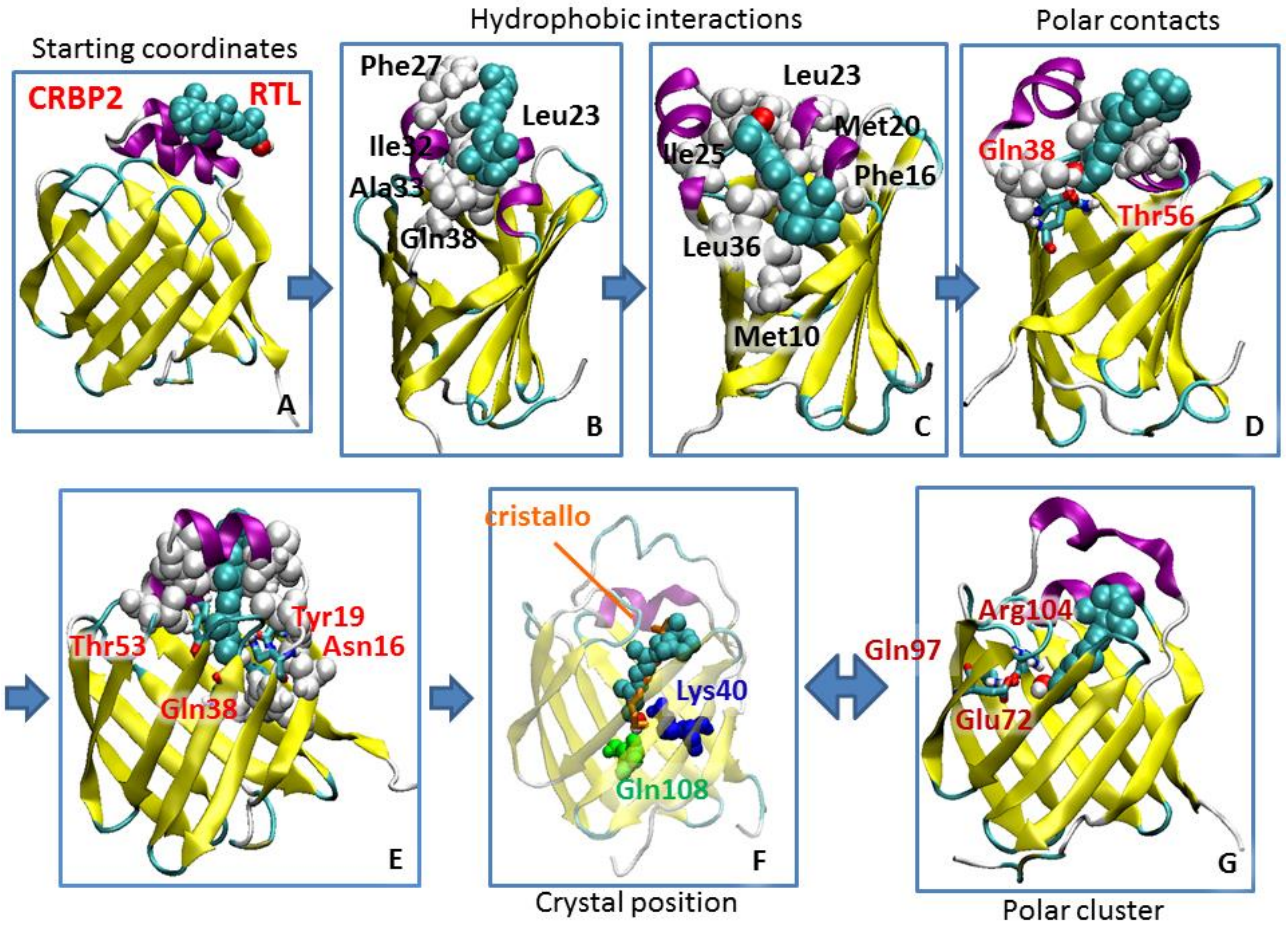


Figure 6.30 Snapshots A-G describe the entrance of the retinol into CRBP2 binding cavity. The protein is represented in cartoon; retinol, colored by atom types, and hydrophobic residues are shown in Van der Waals representation, while polar residues are in sticks and colored by atom types. In panel F, Lys40 and Gln108, the two key residues, interact with the –OH group of retinol; its position seems to be very similar to the one observed in the crystal structure (orange).

Differently from CRBP1, for which experimental information about the up-take mechanism were available, no model for ligand entry has been proposed, at date, for CRBP2. Franzoni and co-workers (Franzoni, Cavazzini, Rossi, & Lücke, 2010), from their NMR analyses, noticed little differences between solution structures of rat CRBP1 and 2; these include particularly a higher flexibility of the alpha helix region and the beta strand A, in agreement with what we have observed in the MD simulation of apo-CRBP2, suggesting the reliability of the up-take model that we have proposed.

6.3.3 MD simulation of CRBP3 in the presence of retinol

Although NMR studies or crystal structure of the holo form of CRBP3 aren't available, the addition of retinol to CRBP3 produces, in absorption analysis, a peak with a maximum at 350 nm, as in the case of CRBP1 and 2, probably attributed to the retinol bound to the protein (Folli C. et al., 2001). The result of the MD simulation on CRBP3 in the presence of retinol shows another different mechanism than those observed for isoform 1 and 2. The ligand, driven by the distribution of surface residues, reaches a position near the portal region (Fig. 6.31B), as in the case of CRBP1, but a different event occurs: the opening of the gap region (Fig. 6.31C). This behavior was observed even in the apo-simulation discussed previously. Despite the large opening formed, the ligand remains at the portal region and after some attempted entries, in which the β-ionone ring points towards the cavity, it reorients itself introducing its hydroxyl group into the binding pocket. In the binding site, retinol interacts with the same polar cluster formed by Glu72, Gln97 and Arg104, interactions conserved in all the four isoforms, while the gap region returns to a close conformation. The ligand remains in this position until the end of the simulation, completely buried inside the cavity and shielded from the solvent. Unfortunately, no information about the binding mode of retinol to CRBP3 are available in literature, therefore we can't compare our data with any other structural information.

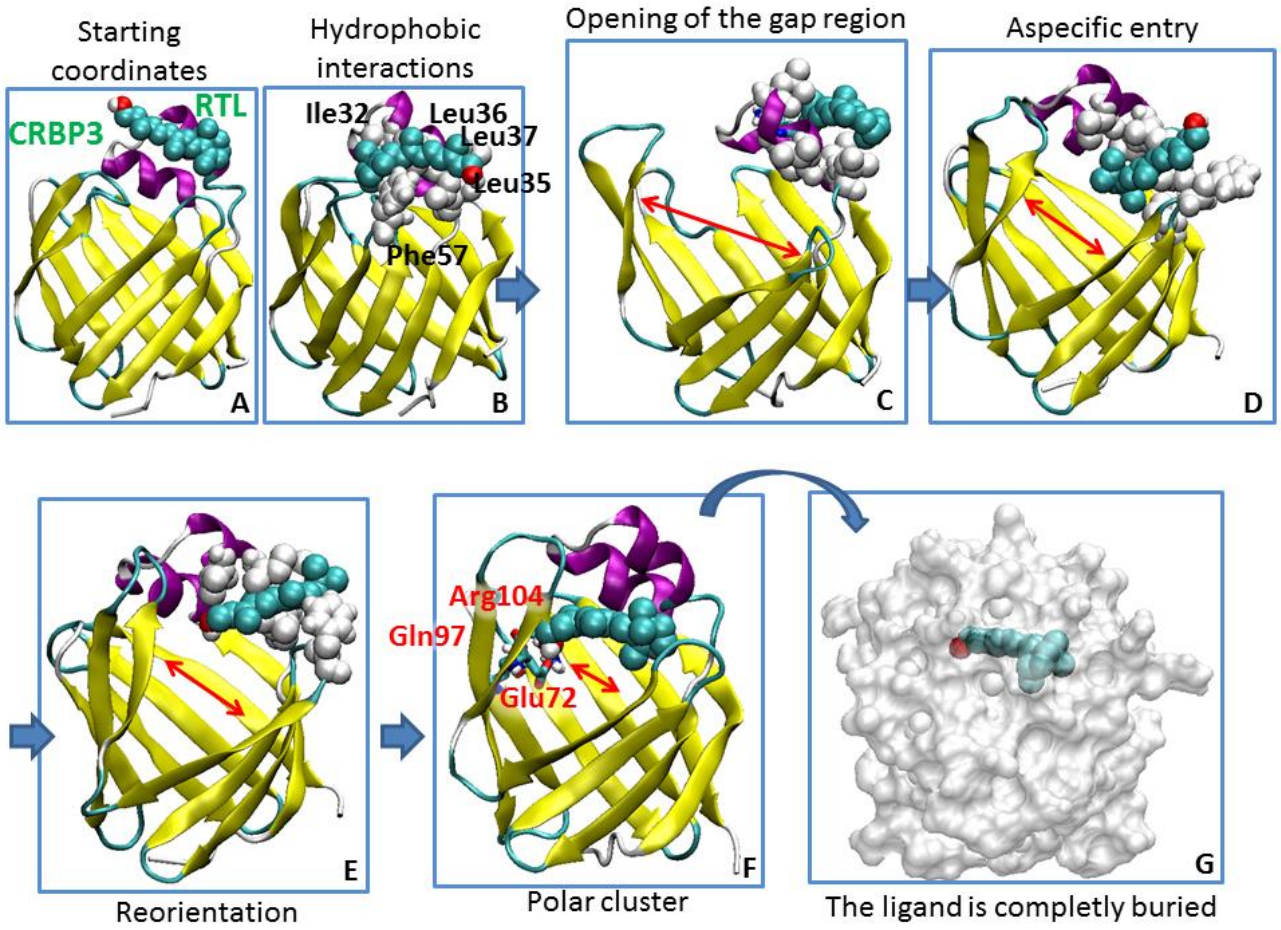


Figure 6.31 Snapshots A-G describe the entrance of the retinol into CRBP3 binding cavity. The protein is represented in cartoon; retinol, colored by atom types, and hydrophobic residues are shown in Van der Waals representation, while polar residues are in sticks and colored by atom types. In panel D we can observed an aborted attempt of ligand entry; then the reorientation of retinol, in panel E, permits its entry into the cavity where it interacts with the polar cluster shown in panel F. At the end of the simulation, retinol is completely buried as in the panel G, in which the molecular surface of CRBP3 is shown in semi-transparency and the ligand appears colored by atom types, in Van der Waals representation.

6.3.4 MD simulation of CRBP4 in the presence of retinol

The experimental data available about CRBP4 already suggest a difference in the retinol-binding mode. In fact, the addition of this ligand to the protein promotes the formation of a peak with a maximum at about 330 nm which cannot be distinguished from the peak of free retinol in buffer (325-330 nm). The MD simulation in the presence of retinol is summarized in figure 6.32 in which some snapshots describe the entrance of retinol. As for CRBP1 and 3, retinol is attracted to the portal region (Fig. 6.32B), but different kind of residues from those present in isoforms 1 and 2 constitute this region; in fact as soon as the CD loop bends out, the β-ionone ring of the ligand insert itself into the cavity, while the hydroxyl group interacts with two polar residues belonging to alpha helix II, namely Asp26 and Thr29 (Fig. 6.32C). This situation is unstable and the ligand tries

repeatedly to leave the cavity (Fig. 6.32D). However, on the ending of the simulation, contrarily to the other CRBPs, retinol inserts into the cavity its hydrophobic ring, while its hydroxyl group is exposed to the solvent as shown in figure 6.32E.

We have carried out three replica of this simulation and all of them confirm the same entrance pathway with the same retinol orientation.

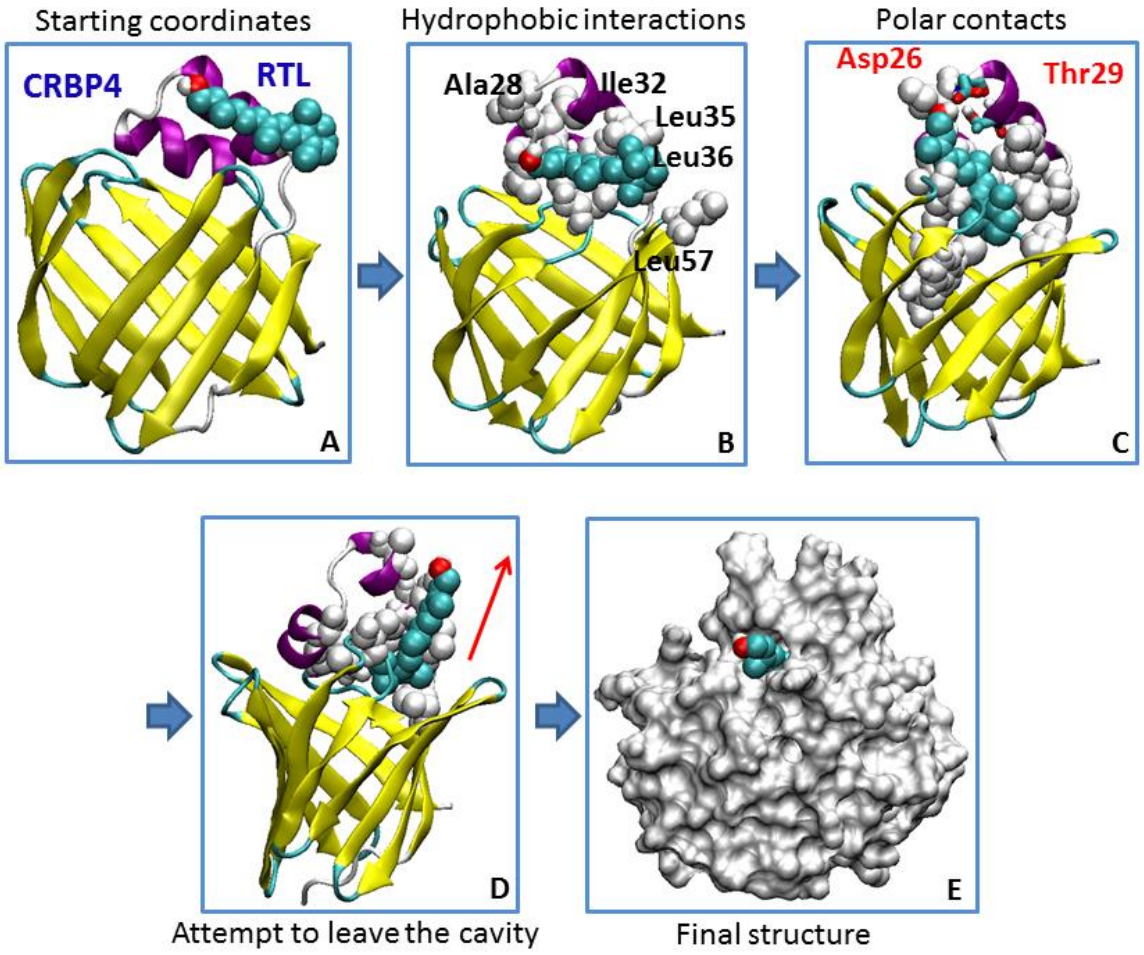


Figure 6.32 Snapshots A-G describe the entrance of the retinol into CRBP4 binding cavity. The protein is represented in cartoon; retinol, colored by atom types, and hydrophobic residues are shown in Van der Waals representation, while polar residues are in sticks and colored by atom types. In panel D the red arrow indicates an attempt of retinol to leave the cavity. However, at the end of the simulation (panel E), retinol is inserted into the cavity in a backwards orientation with respect to what observed for the other isoforms. In this picture the protein surface of CRBP4 is shown and the ligand appears colored by atom types, in Van der Waals representation.

The impossibility to compare the MD data with a crystal structure of holo-CRBP4 makes difficult their interpretation. However, the structure of another retinol transporter, namely RBP and mentioned in the Introduction section, seems to support our finding. The plasma retinol-binding protein, in fact, binds its endogenous ligand retinol in a backwards orientation, if compared with a typical CRBP-retinol complex. In RBP the hydroxyl group of retinol is exposed to the solvent, while

the hydrophobic part of the ligand (i. e. the β-ionone ring and the poly-isoprene tail) is inserted into the binding pocket (Fig. 6.33). Curiously, the absorption spectrum of retinol-RBP complex shows a peak with a maximum at 330 nm, similarly to retinol-CRBP4 complex (Fig. 6.33).

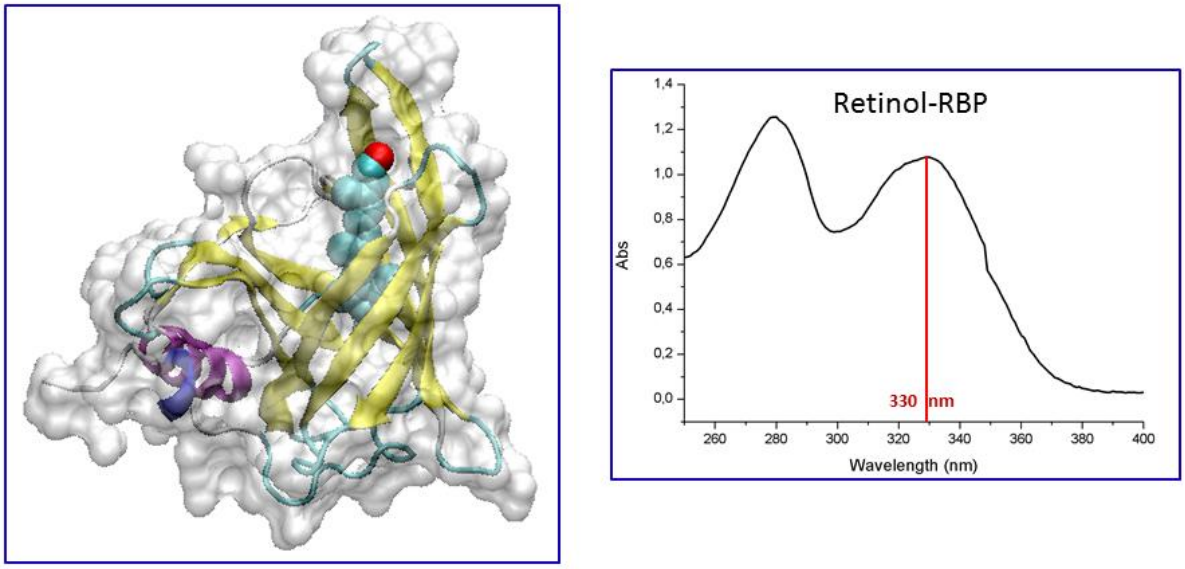


Figure 6.33 The picture on the left shows the molecular surface of RBP in semi-transparency and the retinol inside the cavity, represented in Van der Waals and colored by atom types. On the right is shown the absorption spectrum of the retinol-RBP complex.

6.3.5 The surface residues

Despite the high structural conservation between the four CRBP isoforms, their amino acid sequences show a low percentage of identity (lower than 60%) and many sequence differences occur at residues located at the protein surface. In figure 6.34 is shown the distribution of residues at the surface of CRBP1, 2, 3 and 4, with particular interest for the portal region. Considering only this region, CRBP1 and 3 share a similar residues pattern, different from that of CRBP2 and 4. Comparing CRBP1 and 2, a difference in the portal region is evident: less hydrophobic residues are present in the isoform 2. This feature could be the reason for the different entrance of the ligand. Even in the case of CRBP4 we can notice less hydrophobic residues in the portal region than CRBP1; however, the dynamics of CRBP4 in the presence of retinol, have shown the entry of the ligand with an upside down orientation and with high instability (see below, paragraph 6.5.3, the experimental results of the stability test). These observations have led to the hypothesis for CRBP4 of an endogenous ligand different from retinol, perhaps with a highest affinity.

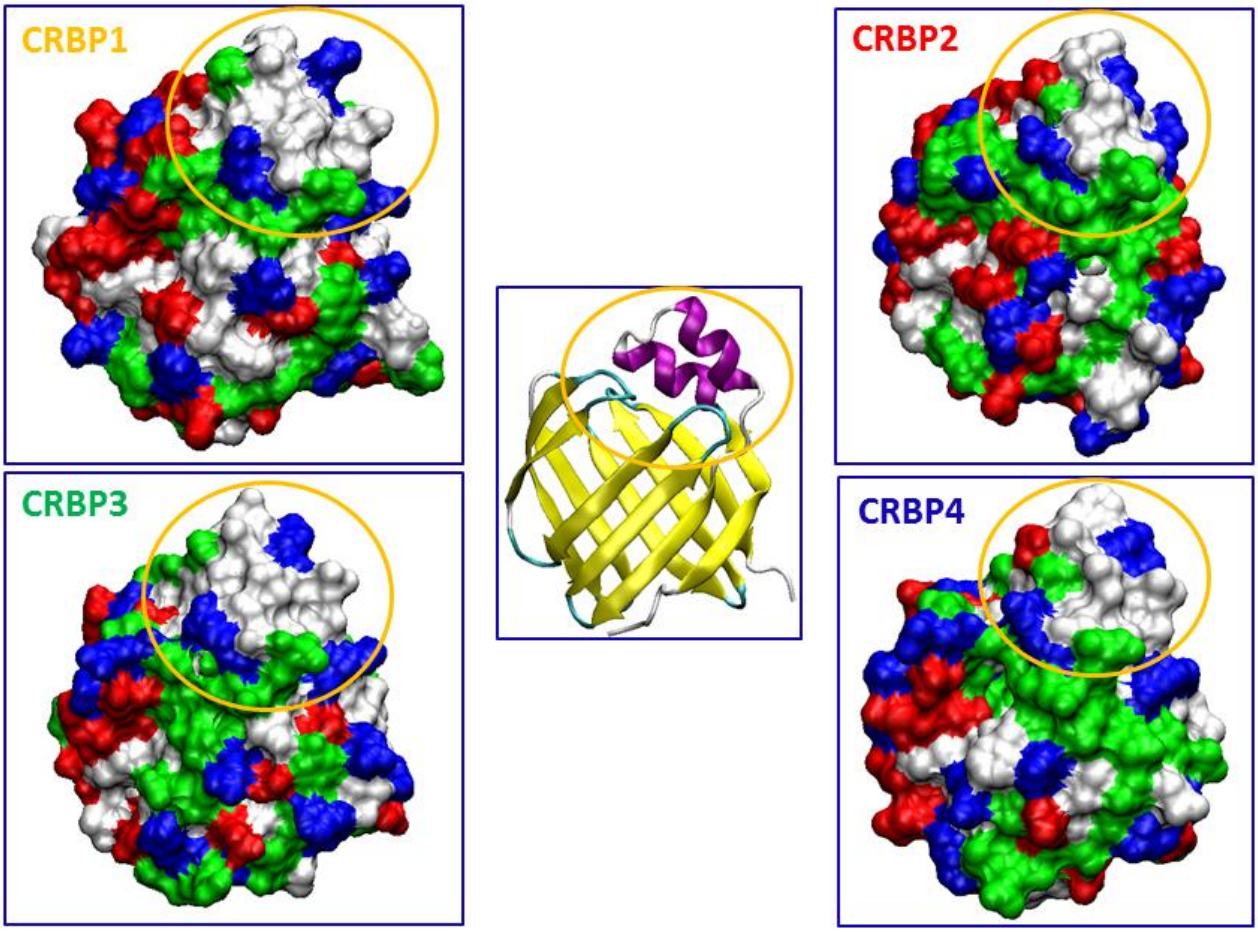


Figure 6.34 Different distributions of surface residues in the four diverse CRBPs, represented with their molecular surface. In white we have colored hydrophobic residues, in green polar amino acids, in blue and red are respectively shown residues with positive and negative charges. All proteins are shown in the same orientation depicted in the central image. We have highlighted the portal region with a yellow circle.

6.4 Experimental validation of the up-take mechanism of CRBP1

The MD simulations on CRBP1 in the presence of retinol, have supplied important suggestions about the up-take mechanism of retinol. With the aim to validate these data we obtained some mutants of important residues, identify by the MD simulation analysis. In agreement with other works (Franzoni et al., 2002) (Mittag et al., 2006), Phe57 and Arg58 seem to be the most important, because their concerted movement creates, in the simulation, a little opening in which retinol can enter. In addition, Tyr60 seems to contribute, together with Arg58, to the first steps of the insertion of the ligand into the binding pocket, by means of the interaction with its hydroxyl group. At the end of the simulation retinol prefers a horizontal position, contacting a polar cluster and interacting in particular with Glu72.

The role of Phe57 was already investigated by Napoli and Penzes (Penzes & Napoli, 1999), which substituted this residue with an alanine to reduce the hydrophobic effect. This mutation resulted in a 3-fold lower affinity of CRBP1 for retinol. Phe57 is a hydrophobic bulky residue which, from the MD data, seems to obstruct the cavity opening; substituting this amino acid with a little hydrophobic and non-aromatic residue like alanine, could interfere with the movement concerted with Arg58 and the opening could be statistically less probable, determining a lower affinity.

The substitutions of Arg58, Tyr60, and Glu72 with amino acids with different chemical-physical characteristics should allow to clarify the role of these residues. Therefore, using the same strategies of mutagenesis, expression and protein purification described for the key residues mutants, we obtained three new mutant forms: Y60F, Y60F/R58L and E72L.

6.4.1 Expression and purification of up-take mutant forms

The expression of the three mutants was verified by SDS-PAGE analysis. In figure 6.35 we report the results after the expression and the cell lysis: all mutants are expressed successfully and a large amount of protein is soluble.

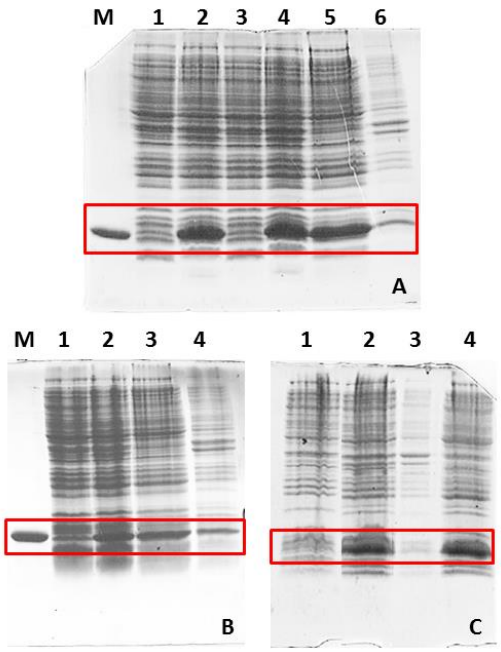


Figure 6.35 SDS-PAGE analysis of A) Y60F/CRBP1: marker corresponding to hCRBP1 (M), non-induced culture (lanes 1 and 3), induced culture (lanes 2 and 4), soluble protein and pellet, after cell lysis (lanes 5 and 6 respectively); B) E72L/CRBP1: non-induced culture (lane 1), induced culture (lane 2), soluble protein and pellet, after cell lysis (lanes 3 and 4 respectively); C) Y60F/R58L/CRBP1: non-induced culture (lane 1), induced culture (lane 2), pellet and soluble protein, after cell lysis (lanes 3 and 4 respectively). Proteins of interest are highlight with red rectangles.

The soluble part of Y60F mutant was purified by two chromatographic passages: an ion exchange and a size exclusion chromatography were carried out. In figure 6.36A the results of an SDS-PAGE on fractions deriving from ion exchange chromatography are shown. Significant fractions (39-42) were collected and concentrated to proceed with the second chromatographic passage. After the size exclusion chromatography (Fig. 6.36B), fractions containing the almost pure Y60F mutant (24-29) were newly collected and concentrated.

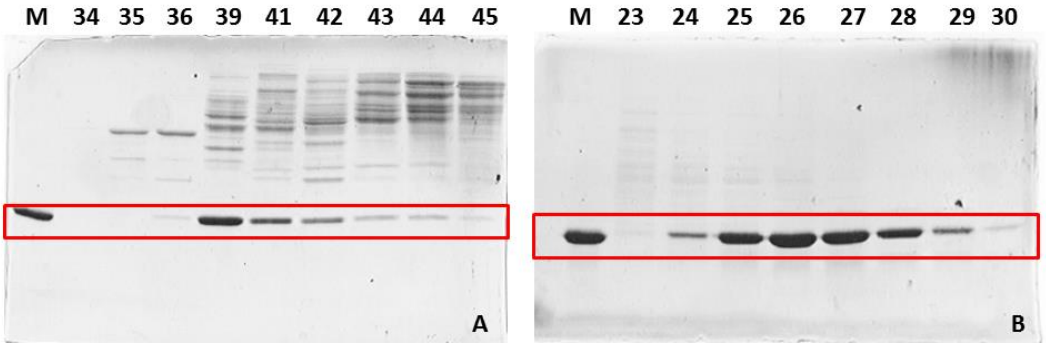


Figure 6.36 SDS-PAGE analysis of fractions with significant absorption values at 280nm, deriving from ion exchange chromatography (A) and size exclusion chromatography (B) performed on Y60F/CRBP1 mutant. Protein of interest is highlight with red rectangles.

In a previous purification of Y60F/R58L double mutant, the ion exchange chromatography was used and the protein was lost; to solve this problem, the double mutant was purified repeating for two times the size exchange chromatography (Fig. 6.37A and B).

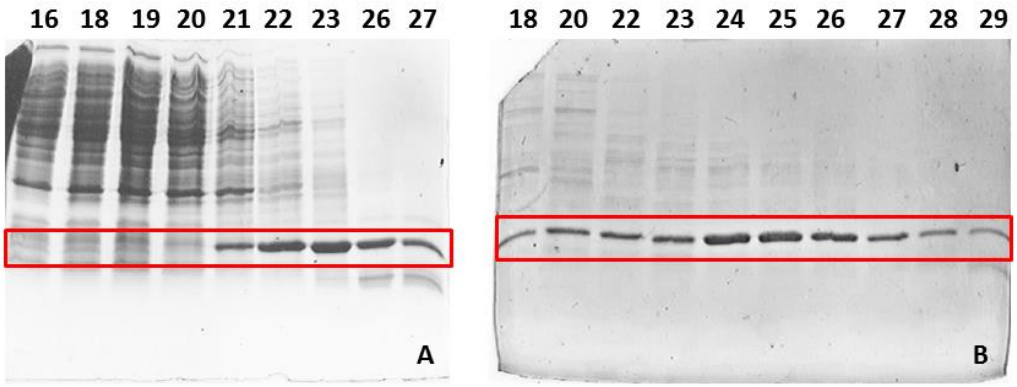


Figure 6.37 SDS-PAGE analysis of fractions with significant absorption values at 280nm, deriving from two size exclusion chromatographic passages (A) and (B) performed on Y60F/R58L/CRBP1 mutant. Protein of interest is highlight with red rectangles.

Fractions deriving from the second passage and containing the almost pure protein (23-29) were collected and concentrated.

Regarding the E72L mutant, we weren't able to find the protein after the ion exchange chromatography; even if we repeated the expression and the purification phases three times, but the protein was always lost during the purification phase. Glutamate 72, conserved in all CRBP isoforms, seems to have a structural function; when this residue is substitute with a hydrophobic amino acid, like leucine, the polar cluster of which it is part (Glu72, Gln97 and Arg104), with its network of H-bonds, could be damaged. In addition, the only polar interaction in the gap region is the hydrogen bond between Glu72 and Tyr60; the substitution of glutamate with a hydrophobic amino acid determines the loss of this interaction and probably makes the gap region too flexible. This suggestion is in agreement with a mutational study (Zimmerman, Rademacher, Ruterjans, Lucke, & Veerkamp, 1999) relative to the E72S substitution in the FABP3 protein, that indicates a significant loss of stability of this mutant, ascribing to glutamate 72 - conserved in the iLBP family - a structural stabilizing role.

At the end of these procedures we obtained pure form of Y60F and Y60F/R58L mutants and we have proceeded with their characterization.

6.4.2 Characterization of Y60F/CRBP1 mutant

By the procedures described in the Materials and methods section, we obtained a holo-protein crystal of the Y60F/CRBP1 mutant. With the collaboration of Professor Zanotti, from the University of Padova, who performed the X-Ray diffraction and the data analysis, we are able to confirm the presence of retinol inside the binding cavity of the mutant. The 3D structure of the mutant is very similar to the one of the WT form. The hydroxyl group of retinol interacts with the two key residues, and the position of the ligand is superimposable with its position in the WT.

The binding assay, performed adding retinol to the mutant solution and recording the absorption spectrum, shows the formation of a peak with a maximum at about 350 nm (Fig. 6.38), similar to that obtained for the WT.

To evaluate the affinity of the mutant for retinol we carried out a fluorescence titration, repeating the measures four times. Y60F have a Kd of approximatively 7.5 ± 2 nM (Fig. 6.38). In comparison with the WT, for which the known Kd is about 0.1 nM (Malpeli, Stoppini, Zapponi, Folli, & Berni, 1995), our mutant shows a 75-fold higher Kd than the WT.

The substitution of tyrosine with phenylalanine compromises the polarity of the residues, but not its steric hindrance; thus, the lower affinity is probably due to the loss of the polar interaction with retinol, seen in the MD simulation. We also noticed an increased instability of this mutant, which seems to be particularly sensitive to freeze-thaw cycles. As mentioned above, Tyr60 interacts with Glu72 forming the only hydrogen bond present in the gap region; it is reasonable to think that the noticed instability may be due to the loss of this interaction when a phenylalanine replaced the tyrosine in this position.

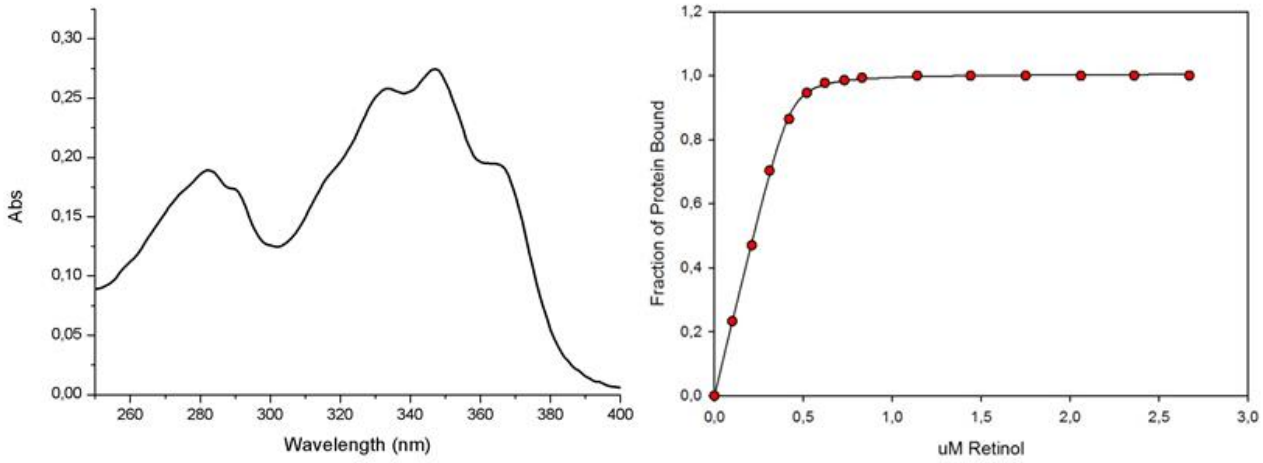


Figure 6.38 On the left, the absorption spectrum recorded after the addition of a slight molar excess of all-trans-retinol to the Y60F/CRBP1 mutant. On the right is shown the fitting to a theoretical curve of the experimentally determined fractional saturation, obtained by the spectrofluorometric titration of Y60F/CRBP1 mutant with retinol.

6.4.3 Characterization of the Y60F/R58L double mutant of CRBP1

The Y60F/R58L double mutant is affected by high instability, in particular after freeze-thaw cycles. However we were able to record an absorption spectrum just after the addition of a slight molar excess of *all-trans*-retinol. In this case a peak different from WT protein, with a maximum at about 330 nm, was observed (Fig. 6.39). This absorption spectrum is similar to that observed for retinol-CRBP4 complex, which suggests a different and labile bound of the ligand.

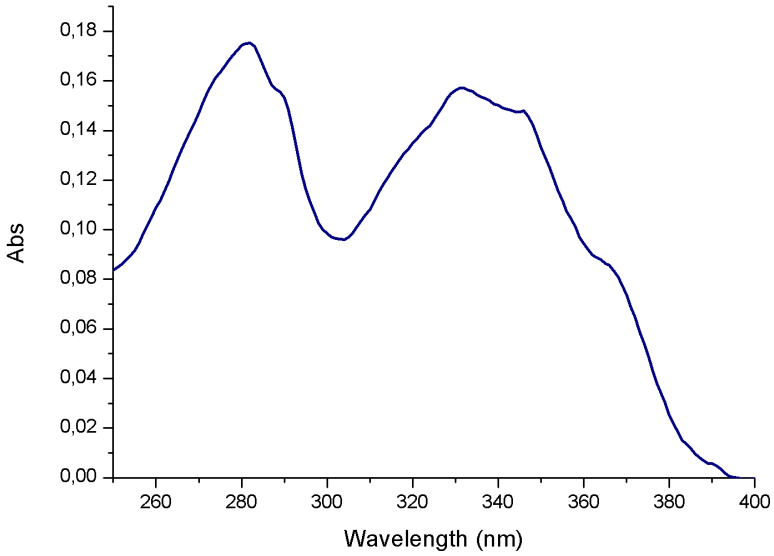


Figure 6.39 The absorption spectrum recorded after the addition of a slight molar excess of all-trans-retinol to the Y60F/R58L double mutant.

Attempts to perform a fluorescence titration failed, because of the instability of the protein. The fitting of the experimentally determined fractional saturation to a theoretical curve, reports an *n* value - which indicates the number of binding site in the protein - of about 0.4 instead of 1, value typically obtained for CRBPs. This data suggests that only a little fraction of the protein used is able to bind retinol.

The Y60F/R58L mutant was projected with the idea of study the combined effects of both these residues. However, the impossibility to work with a stable protein has led to plan the R58L single mutant. The expression, purification and characterization of the latter, are currently underway.

6.5 CRBP3 and 4

The finding during the X-Ray analysis of an E.coli ligand bound to the WT CRBP1 (see paragraph 6.1.3), suggested the possibility that the different spectrum of the retinol-CRBP4 complex could be due to an E.coli ligand bound to CRBP4, during the expression phase, with greater affinity than retinol. In this case, the addition of *all-trans*-retinol should not be sufficient to replace the other ligand. To solve this doubt we expressed and purified CRBP4, trying to eliminate the undesired ligand possibly bound to this isoform. To this aim, we used the protocol of delipidation described in the Material and methods section, then performing again the binding assay in absorption.

CRBP3 was also expressed to perform binding assays with a retro-retinoid, as discussed below.

6.5.1 Expression and purification of CRBP3

To express CRBP3, the protocol indicated by Folli et al. (C Folli et al., 2001) was initially used, but a very low quantitative of soluble protein was obtained. We optimized the expression incubating the culture at 4°C for 16 hours, after the addition of a lower concentration of IPTG than that used in the past. With these new conditions we obtained an increase of the yield, which, nevertheless, remains lower that the yield of other expressed CRBP (CRBP1 and 4). In figure 6.40 is shown the result of SDS-PAGE analysis performed to verify the expression.

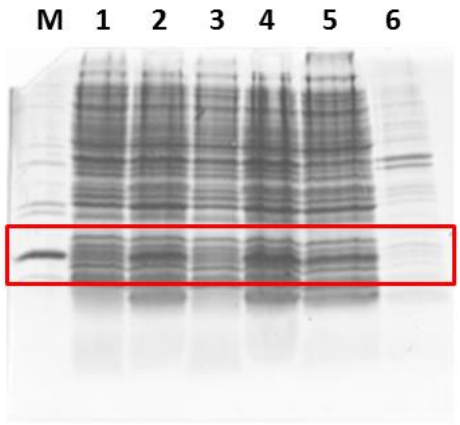


Figure 6.40 SDS-PAGE analysis of CRBP3 expression: marker corresponding to hCRBP3 (M), non-induced culture (lanes 1 and 3), induced culture (lanes 2 and 4), soluble protein and pellet, after cell lysis (lanes 5 and 6 respectively). Protein of interest is highlight with red rectangle.

The soluble fraction was purified in the same way used for CRBP1: an ion exchange and a size exclusion chromatography were carried on. After the first chromatographic passage, fractions which showed significant absorption values at 280 nm were analyzed by SDS-PAGE (Fig. 6.41A); fractions containing the protein of interest (45-48) were collected and concentrated to proceed with the size exclusion chromatography. Even for these fractions an SDS-PAGE analysis was performed (Fig. 6.41B) and only fractions which had a low level of contaminants were collected and concentrate. The pure protein was, then, quantified, delipidated and used for the binding assays.

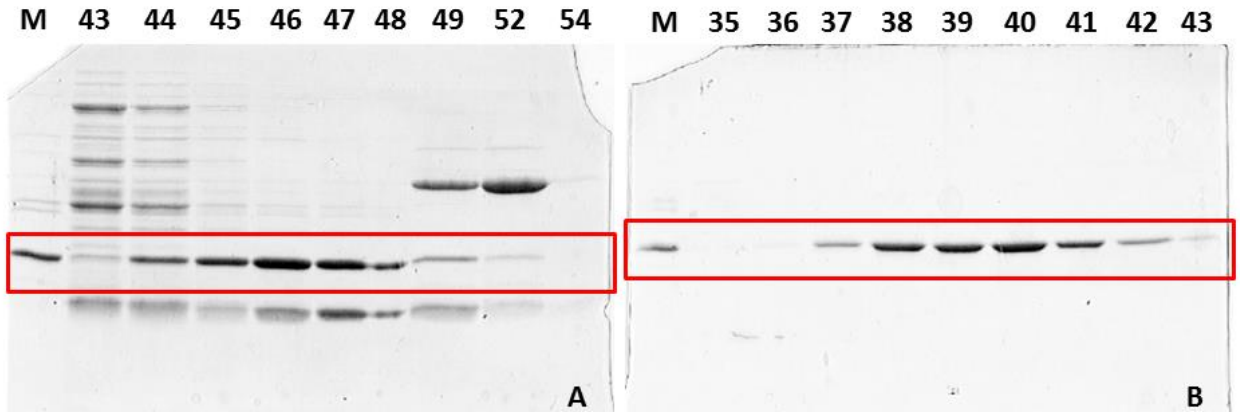


Figure 6.41 SDS-PAGE analysis of fractions with significant absorption values at 280 nm, deriving from ion exchange chromatography (A) and size exclusion chromatography (B), performed on CRBP3. Protein of interest is highlight with red rectangles.

6.5.2 Expression and purification of CRBP4

Even in the case of CRBP4 expression, the same protocol used for CRBP3 was followed, limiting the part of the protein enclosed in inclusion bodies using a lower temperature. In figure 6.42, the result of SDS-PAGE analysis indicates a good expression and that the major part of the protein is soluble.

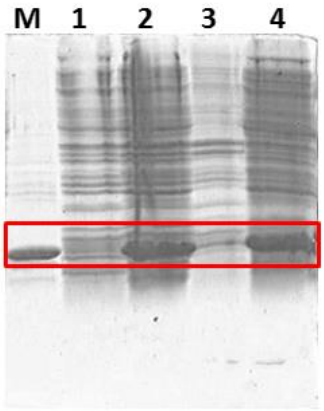


Figure 6.42 SDS-PAGE analysis of CRBP4 expression: marker corresponding to hCRBP4 (M), non-induced culture (lane 1), induced culture (lane 2), pellet and soluble protein after cell lysis (lanes 3 and 4 respectively). Protein of interest is highlight with red rectangle.

Despite CRBP4 is positively charged at pH of 7.5, we have however performed an anion exchange chromatography to separate from the protein (that was eluted in the flow-through) the large amount of negatively charged DNA contained in the lysate. From the SDS-PAGE analysis we identified fractions, of the flow-through, containing the protein of interest. A size exclusion

chromatography was carried out to completely purify CRBP4. Fractions which had shown the presence of the isoform 4 (Fig. 6.43B) were collected and concentrated. The pure protein was then delipidated and used for the binding assays.

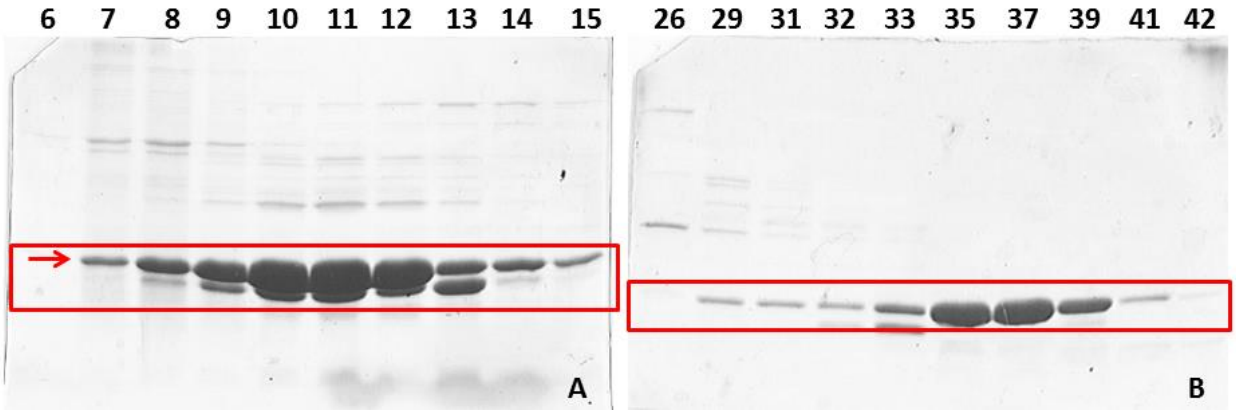


Figure 6.43 SDS-PAGE analysis of fractions with significant absorption values at 280nm, deriving from ion exchange chromatography (A) and size exclusion chromatography (B), performed on CRBP4. Protein of interest is highlight with red rectangles.

Interestingly, during the first expression and purification of CRBP4, an unknown ligand bound the protein during the passage into the ion exchange column. This fact suggested to attempt the characterization of this new ligand (see below at paragraph 6.6). No size exclusion chromatography was carried out on this preparation and all the protein obtained was used for this characterization.

6.5.3 Binding assay on delipidated CRBP4

After delipidation, CRBP4 was used for the binding assay with *all-trans*-retinol. As described in the Materials and methods section, a slight molar excess of ligand was added to the protein and the absorption spectrum was recorded. We also monitored the stability of the complex after two hours from the addition of the compound.

The absorption spectrum recorded after the addition of retinol to the delipidated CRBP4 is identical to that reported by Folli and co-workers (Claudia Folli, Calderone, Ramazzina, Zanotti, & Berni, 2002): despite the delipidation passage, the addition of retinol to CRBP4 doesn't produce the typical spectrum observed for the other CRBPs. In addition, monitoring the stability after only two hours, the almost complete disappearance of the peak at 330 nm is observed (Fig. 6.44). These data suggest that CRBP4 binds *all-trans*-retinol with high lability.

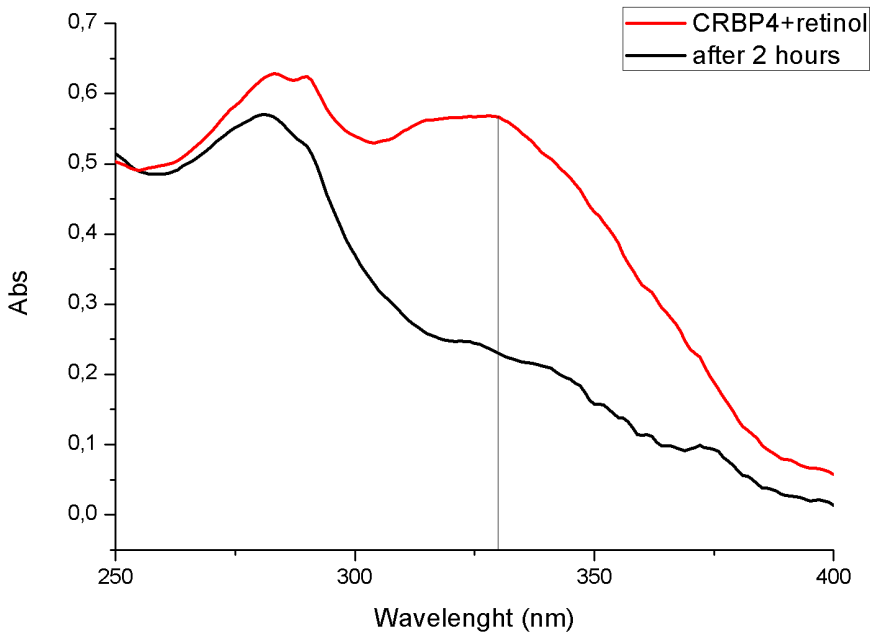


Figure 6.44 The absorption spectrum recorded after the addition of a slight molar excess of all-trans-retinol (red line) to CRBP4 after delipidation, showing the formation of a peak with a maximum at about 330 nm. The black line corresponds to the absorption spectrum of the same complex recorded after 2 hours, during which the sample was maintained at room temperature, in the dark.

6.6 Characterization of the unknown ligand, bound to CRBP4

As previously mentioned, during the purification of CRBP4 an unknown ligand was found in complex with this isoform. An absorption spectrum of the complex was recorded and is shown in figure 6.45A: three defined peaks, at 337, 354 and 374 nm, are present.

To be sure that the ligand was bound into the cavity and not aspecifically bound on the protein surface, we utilized a mini-size-exclusion column (Micro BioSpin Chromatography column, BioRad) which serves to elute big macromolecule (>6KDa) and hold back little unbound compounds. After recording an absorption spectrum of the protein eluted, we were able to confirm that the ligand was bound inside the CRBP4 binding cavity.

To identify the nature of the unknown ligand, it was needed to extract the compound using the lipidex resin; the ligand was, then, eluted adding methanol to the resin. In figure 6.45B is shown the absorption spectrum of the free unknown ligand which exhibits three defined peak at 334, 348, 367 nm.

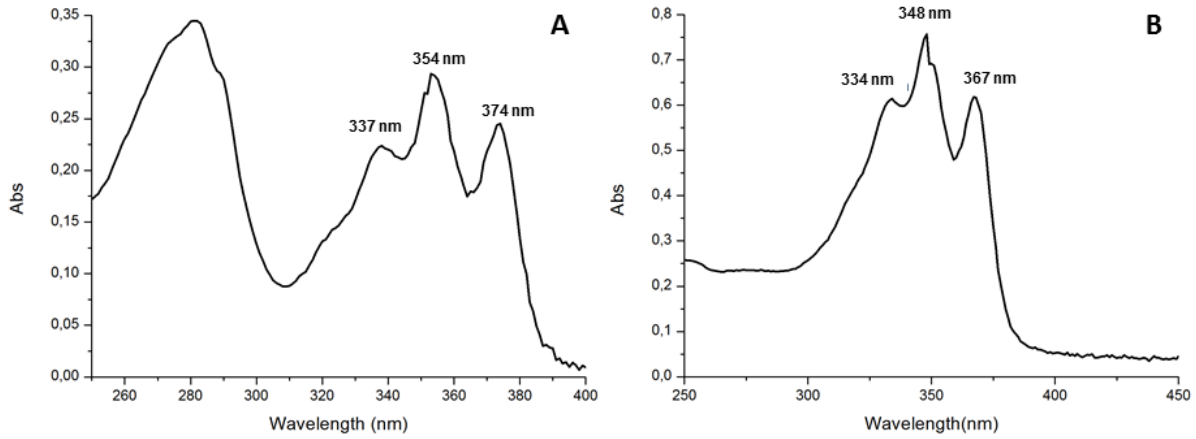


Figure 6.45 Absorption spectra of: A) CRBP4 bound to an unknown ligand; B) the unknown ligand in methanol.

The free compound absorption spectrum obtained, is very similar to that of 14-Hydroxy-4,14-retro-retinol (14HRR) (Buck et al. 1991), a natural retro-retinoid involved in the proliferation of T cells belonging to the immune system (Derguini, Nakanishi, Haemmerling, & Buck, 1994). The directed infusion of few microliter of compound, dissolved in methanol, in a mass spectrometer instrument, revealed the presence of a ligand having $C_{20}H_{30}O_2$ as empirical formula, compatibles with the retro-retinoid aforementioned. The MALDI analyses confirmed the release of a compound of approximately 304.2 Da only in the sample in which denaturant conditions were applied (figure 6.46, red spectrum). All these data have led to the identification of 14HRR.

It is not clear how this retro-retinoid was present into the ion exclusion column. Given that this column was also used for the purification of the plasma retinol-binding protein, it is possible that a part of the retinol unbound to RBP was hold by the resin, remaining trapped for a long time. Into the matrix, retinol may be subjected to modifications, due to the medium acidification or to the light exposition. It is also possible that some E.coli enzymes, contained in the lysate samples consecutively passed through the column, have converted retinol to 14HRR, which is remained bound to the resin. When CRBP4 passed through the column, the protein selectively bound this compound.

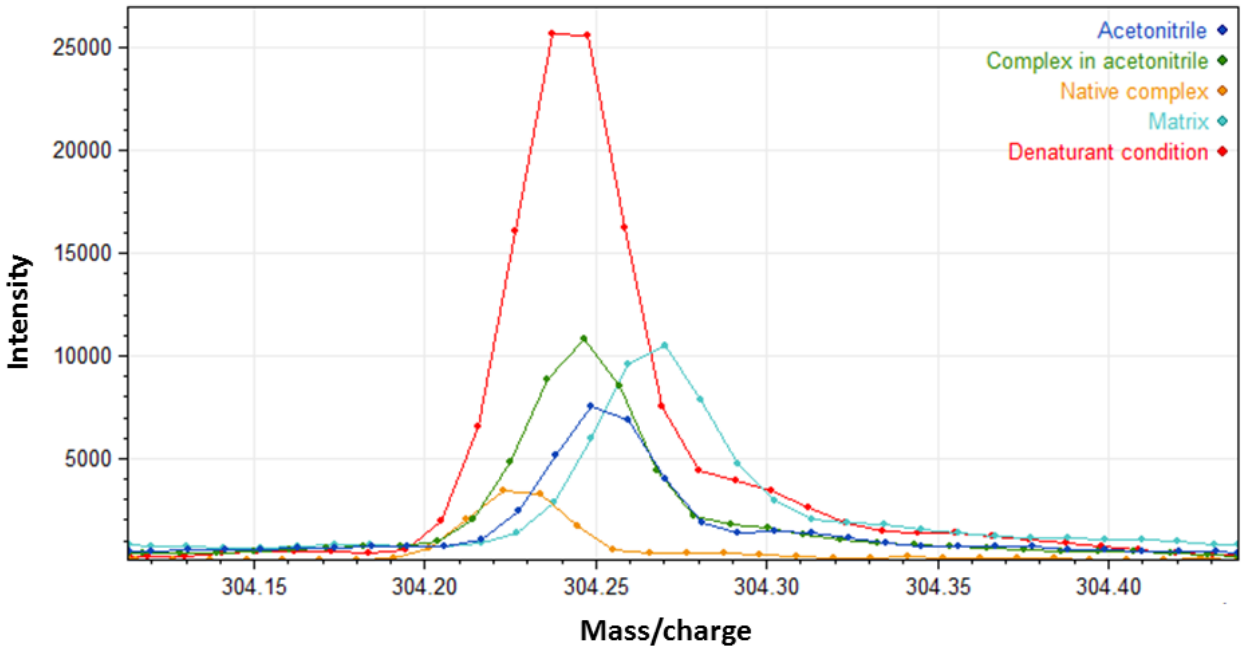


Figure 6.46 MALDI-TOF mass spectra: blue) sample of matrix and acetonitrile, green) the ligand-CRBP4 complex in presence of acetonitrile, orange) the native complex, cyan) the matrix, red) the denaturated sample. Information about the samples preparation is available in the material and methods section.

6.6.1 Chemical synthesis of 14HRR

To be sure that 14HRR was, effectively, the identity of the unknown ligand, we toke advantage from the collaboration with Professor Zanardi and Professor Curti, of the Pharmaceutical Department of the University of Parma, who performed the chemical synthesis of the compound. Both the enantiomeric forms (R and S) of 14HRR, whose formulas are shown in figure 6.47, were obtained. The synthesis was carried out as describe by Corey and co-workers (Corey, Noe, & Guzman-Perez, 1995).

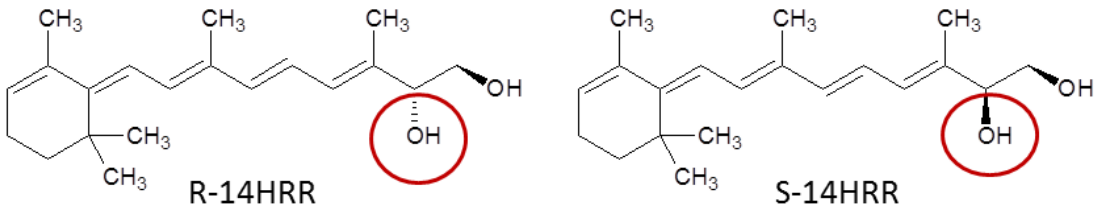


Figure 6.47 Structural formulas of the two enantiomeric forms of 14HRR.

The addition of a slight molar excess of 14HRR (S or R) to CRBP4 lead to the formation of an absorption spectrum (Fig. 6.48) similar to that obtained after the purification of this isoform (Fig. 6.45): the same three peaks were obtained, concluding that 14HRR was effectively the identity of the unknown ligand.

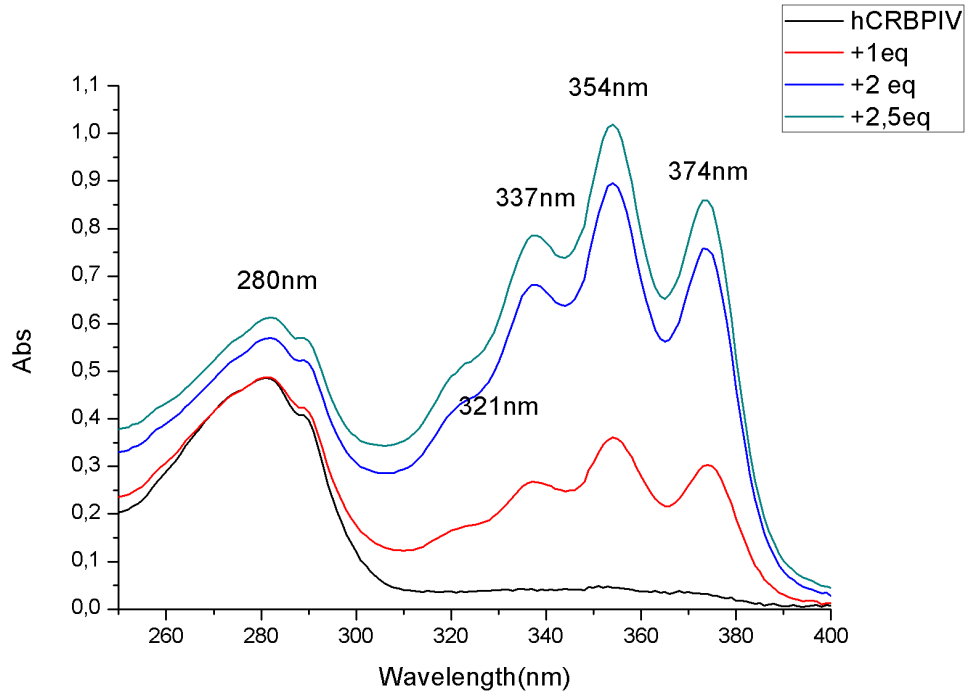
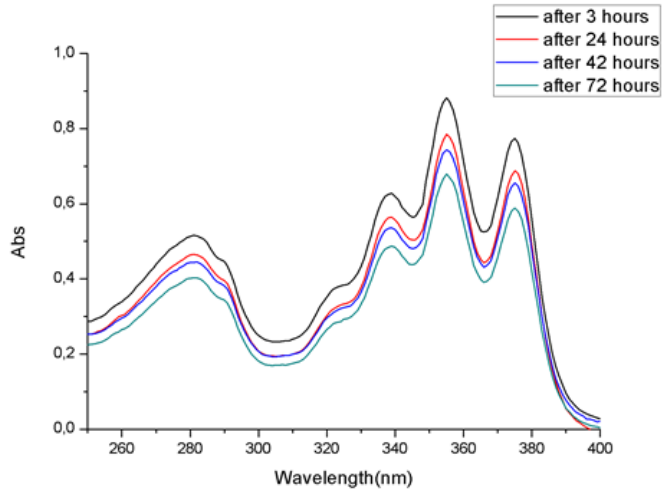


Figure 6.48 Absorption spectra of CRBP4 in Tris Hcl 30 mM + NaCl 0,5 M (black line); CRBP4-14HRR complex after the addition of 1 molar equivalent (red line), 2 molar equivalents (blue line) and 2.5 molar equivalent (green line) of 14HRR.

6.6.2 Binding of 14 HRR to CRBP1, 3 and 4

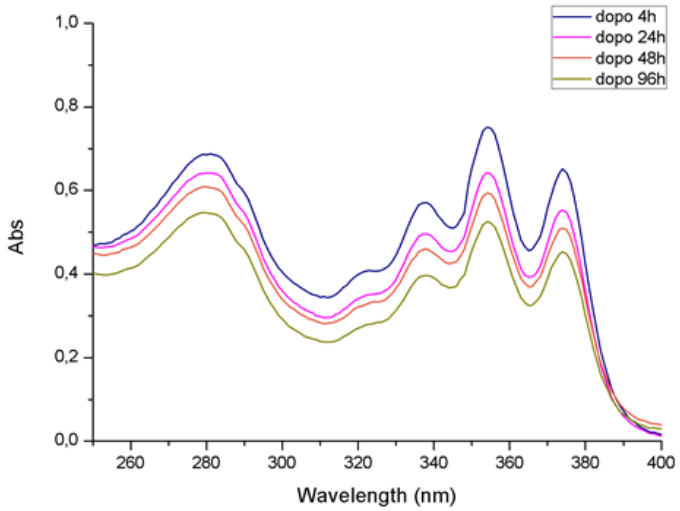
We carried out some preliminary analysis to evaluate the binding of 14HRR to the three isoforms currently available in our laboratory: CRBP1, 3 and 4. Monitoring in absorption the intensity of the peaks attributed to the 14HRR-CRBP complex, we were able to follow the stability of the complex in the time. The decrease of the ratio between the absorbance measured at 345 nm and that at 280 nm, calculated at each stated time, indicates the progressive loss of ligand bound to CRBP. We calculated the percentage of the decrease of retinol bound to CRBP, dividing the ratio value recorded at each time with the same ratio value obtained after 4 hours from the retinoid addition.

The absorption spectra of the 14HRR(R)-CRBP complexes and relative ratios are shown in figures reported below.



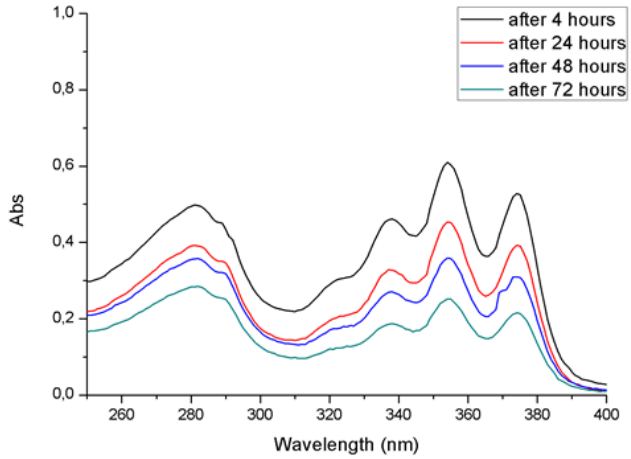
CRBP1	$OD_{280\ nm}$	$OD_{355\ nm}$	A_{355}/A_{280}	Decrease (%)
4h	0.51	0.88	1.72	/
24h	0.46	0.78	1.70	1.16%
48h	0.44	0.74	1.68	2.34%
72h	0.40	0.67	1.67	2.9%

Figure 6.49 The picture on the top show the absorption spectrum of 14HRR(R)-CRBP1 recorded after 4, 24, 42, 72 hours from the addition of a slight molar excess of the compound. The table summarizes the ratio values and the percentage of the decrease of the ligand bound to the protein.



CRBP3	$OD_{280\ nm}$	$OD_{354\ nm}$	A_{354}/A_{280}	Decrease (%)
4h	0.68	0.75	1.09	/
24h	0.64	0.65	1.02	6.4%
48h	0.61	0.59	0.97	11%
96h	0.57	0.55	0.96	11.9%

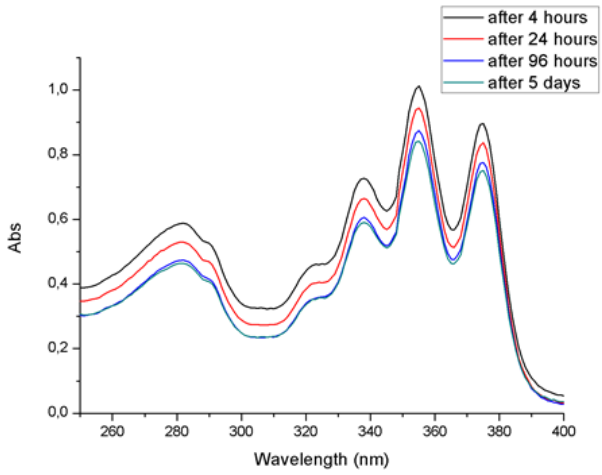
Figure 6.50 The picture on the top show the absorption spectrum of 14HRR(R)-CRBP3 recorded after 4, 24, 48, 96 hours from the addition of a slight molar excess of the compound. The table summarizes the ratio values and the percentage of the decrease of the ligand bound to the protein.



CRBP4	$OD_{280\ nm}$	$OD_{355\ nm}$	A_{355}/A_{280}	Decrease (%)
4h	0.49	0.61	1.24	/
24h	0.39	0.45	1.15	7.3%
48h	0.36	0.36	1.00	19.3%
72h	0.28	0.25	0.89	28.2%

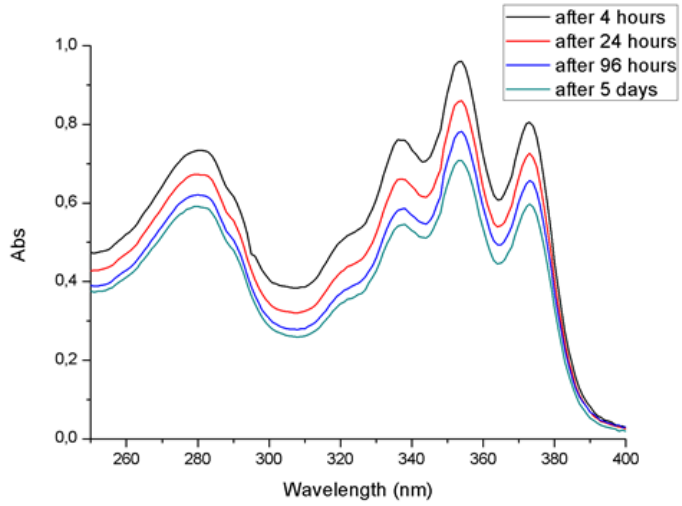
Figure 6.51 The picture on the top show the absorption spectrum of 14HRR(R)-CRBP4 recorded after 4, 24, 48, 72 hours from the addition of a slight molar excess of the compound. The table summarizes the ratio values and the percentage of the decrease of the ligand bound to the protein.

In the following are shown the absorption spectra of the 14HRR(S)-CRBP complexes, and the relative ratio values.



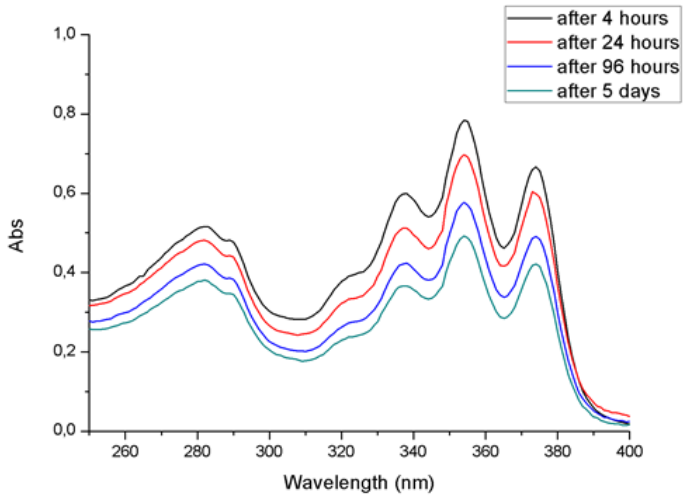
CRBP1	$OD_{280\ nm}$	$OD_{355\ nm}$	A_{355}/A_{280}	Decrease (%)
4h	0.58	1.01	1.74	/
24h	0.53	0.94	1.77	0%
96h	0.47	0.87	1.85	0%
5gg	0.46	0.84	1.83	0%

Figure 6.52 The picture on the top show the absorption spectrum of 14HRR(S)-CRBP1 recorded after 4, 24, 96 hours and 5 days from the addition of a slight molar excess of the compound. The table summarizes the ratio values and the percentage of the decrease of ligand bound to the protein.



CRBP3	$OD_{280\ nm}$	$OD_{353\ nm}$	A_{353}/A_{280}	Decrease (%)
4h	0.73	0.96	1.31	/
24h	0.67	0.86	1.28	2.29%
96h	0.62	0.78	1.26	3.8%
5gg	0.59	0.71	1.20	8.4%

Figure 6.53 The picture on the top show the absorption spectrum of 14HRR(S)-CRBP3 recorded after 4, 24, 96 hours and 5 days from the addition of a slight molar excess of the compound. The table summarizes the ratio values and the percentage the decrease of the ligand bound to the protein.



CRBP4	$OD_{280\ nm}$	$OD_{354\ nm}$	A_{354}/A_{280}	Decrease (%)
4h	0.51	0.78	1.53	/
24h	0.48	0.70	1.46	4.6%
96h	0.42	0.58	1.38	9.8%
5gg	0.38	0.49	1.29	15.7%

Figure 6.54 The picture on the top show the absorption spectrum of 14HRR(S)-CRBP4 recorded after 4, 24, 96 hours and 5 days from the addition of a slight molar excess of the compound. The table summarizes the ratio values and the percentage of the decrease of the ligand bound to the protein.

Measurements after 24 hours (summarized in table 3), common to all experiment, suggest that S form of 14HRR is the enantiomer that forms the ore stable complex with all CRBPs tested. Moreover 14HRR(S)-CRBP1 seems to be the complex with the major stability, followed by 14HRR(S)-CRBP3 and finally 14HRR(S)-CRBP4 is the complex with the lowest stability. However, after 5 days the absorption of the 14HRR(S) bound to CRBP4 is still visible, while the addition of *all-trans*-retinol to CRBP4 led to the almost complete loss of signal in the 320-400 nm range after only two hours.

Table 3 Summary of the percentage drop of ligand bound to CRBPs after 24 hours.

	CRBP1	CRBP3	CRBP4
(R)14HRR	1.16%	6.4%	7.3%
(S)14HRR	0%	2.29%	4.6%

CONCLUSIONS AND PERSPECTIVES

The study of key binding residues conserved in CRBP1 and 2, have provided important information about the interaction of retinol into the binding cavity. Glu108 and Lys40, as is clearly demonstrate by the crystal structures of retinol-CRBP1 and 2 complexes, interact with the retinol hydroxyl group by hydrogen bonds. The substitution of one of the two key residues with a hydrophobic one, like leucine, has led to the increasing of the dissociation constant, i.e. a loss of affinity. Both residues seem to be important for the strong anchoring of retinol in the binding site. We are not able to suggest which of the two key residues is the most important for the binding because the Kd of the two mutants are very similar. Nevertheless, our results seem to indicate a predominant anchoring role for Glutamine 108, while Lysine 40 could be involved in the catch of the hydroxyl group of the retinol and in the interaction with the poly-isoprene tail, given that its substitution lead to a change in the absorption spectrum. The cation-π interaction between Lys40 and the retinol tail was already proposed by NMR studies but never demonstrated. In our work we have evidenced that Lys40 is critical for determining the characteristic chromophoric proprieties of retinol-CRBP1 complex and that could be involved in the retinol up-take mechanism, as observed in the MD simulation in the presence of ligand. The cation π-interaction could be also important to drag the ligand to the bottom of the binding cavity, after the OH-group capture, to permit its anchoring to Gln108. The double mutant, in which both key residues are replaced, preserves the capability to bind retinol even if with lowest affinity, suggesting that the hydrophobic interactions between the lipophilic part of retinol and the hydrophobic residues of the cavity are sufficient to hold the ligand in the binding pocket. However, the retinol ending part shows an increased mobility due to the absence of polar interactions with the key residues. In addition, the presence of the ligand in the crystal structure of the double mutant indicates that the up-take mechanism is maintained and the key binding residues seem not to be involved in the entrance of retinol. This first conclusion suggest that other unidentified factors, distinct from the substitution of glutamine 108 with histidine in CRBP3 and 4, are responsible for the different affinities of the four CRBPs. Despite the two key binding residues are conserved in CRBP1 and 2, these two isoforms show different affinity for both retinol and retinal. A lot of mutants of CRBP1 and 2 concerning residues of the cavity, which show several differences in the two primary isoforms, were characterized by various groups to investigate the reasons of these different affinities, without reaching a conclusion. Most sequence differences between CRBP1 and 2 occur at residues on the protein

surface, suggesting a possible role of the distribution of these surface residues in the ligand pre-selection.

Molecular dynamics simulations were useful to study the protein flexibility and have provided new insight in the ligand up-take mechanism of CRBPs.

The MD simulations performed on CRBP1 have revealed an overall rigidity of the entire structure; however, in accord with NMR data (Franzoni et al., 2002), the portal region, formed by the helix II and the CD and EF loops seems to be affected by a greater flexibility, suggesting that this region could be involved in the entry of the ligand. In addition, a little opening has been identified during the simulation, due to the concerted movement of Phe57 and Arg58. The entrance of the ligand involves this open conformation of the portal region, that retinol reaches by means of a strong interaction with a hydrophobic cluster, and in which it enters driven by consecutive polar interactions with its hydroxyl group, in particular by Arg58 and Tyr60. Therefore, we can conclude that little changes are sufficient for the entry of the ligand, as previously suggested by NMR studies.

We have then proceeded with the experimental validation of the computational data, obtaining mutants of the relevant residues identified during the simulation. Only the Y60F mutant was useful for the analysis, and the calculation of its dissociation constant shows a 75-fold higher Kd than that of the WT, suggesting the importance of this residue in the retinol up-take. The expression, purification and characterization of the R58L single mutant are currently underway in our laboratory, while the double mutant Y60F/R58L is very instable and it is not possible to calculate its dissociation constant. In the up-take simulation, retinol reaches a final position in which the hydroxyl group is captured by a polar cluster conserved in all CRBPs and formed by Glu72, Gln97 and Arg104. To clarify why the ligand is not able to contact Glu108 and to remain anchored in a crystal-like position, we have expressed the E72L mutant form, but this mutant was lost during each purification tried, suggesting a structural role for this residue. In addition, we are currently dealing with the optimization of retinol topology parameters, to verify the reliability of the ones available in the used force field. Our experimental studies on the key binding residues, in fact, have suggested that the correct parametrization of the poly-isoprene tail could be very important, due to its observed interaction with the positively charged Lys40.

MD simulations performed on CRBP2 suggest a different behavior. Even if also this protein is on the whole quit rigid, in both apo- and up-take simulations we have observed the major movement of the helix-turn-helix region, together with the ending part of beta-strand A. Quite surprisingly, in the simulation of the up-take we have observed a different ligand entry point from that observed for CRBP1: retinol penetrates into the cavity through a cleft between the two alpha helices, positioning itself, at the end of the simulation, in the same way as in the crystal structure. We have therefore proposed a CRBP2 up-take model different from that of CRBP1. Our data are in accord with NMR data obtained about CRBP2 (Franzoni, Cavazzini, Rossi, & Lücke, 2010), that indicate the major flexibility of the alpha helix I and the beta strand A. Experimentally, we are working to isolate the human sequence of CRBP2 from a cDNA library of human intestine, to express the isoform 2. The design of a R58L mutant also for CRBP2, could shed light on the differences of the two up-take mechanisms.

Differently from what observed for CRBP1 and 2, the simulations on CRBP3 have shown a high flexibility of this isoform, in particular affecting the gap region. However, despite the formation of a big opening due to the gap enlargement, retinol enters into the cavity through the portal region. This behavior, together with the lowest affinity for retinol, has suggested the possibility for CRBP3 to host a ligand more bulky than retinol. Even the computational study on CRBP4 has shown a high flexibility of this isoform and the formation of various openings on the protein surface. However, the simulation in presence of ligand has shown a different retinol entry mode: the ligand enters in a backward orientation and the binding is unstable. The absorption spectrum of retinol-CRBP4 complex exhibits a maximum at about 330 nm, similar to that obtained for the complex of retinol with RBP protein, in which retinol binds the carrier again in a backwards orientation. This correlation, together with the very low affinity of retinol for CRBP4, suggests that retinol could not be the endogenous ligand of isoform 4. Instead, we have identified a new natural compound, namely 14HRR, which is able to bind not only CRBP4, but also CRBP1 and 3. The ligand remains bound to the protein after more than 5 days, making this compound a putative ligand for CRBP4. Despite we are not able to correlate the role that 14HRR has in the immune system with the localization of CRBP4 in human organism, we are trying to obtain the crystal structure of 14HRR-CRBPs complexes, to understand the binding mode of this compound. The fact that CRBP4 can bind 14HRR and perhaps not retinol is interesting given that the retro-retinoid is, simplifying, retinol with the addition of a second alcoholic group. Thus, a bulkier molecule binds CRBP4 while a

smaller one can't. This conclusion sustains the aforementioned hypothesis of a pre-selection of the ligand obtained by the different surface residues distribution of CRBPs. In this case, the increased polarity of the compound (14HRR) compared to retinol, could be useful to promote the interaction of the ligand with CRBP4. We can't exclude that a different compound could enter in a different way than that observed for retinol in the CRBP4 MD simulation.

Ligand up-take simulations were performed at 350K, while apo-simulations were carried out in standard condition. However, the major changes seen in all up-take simulations were already observed in the corresponding apo-protein simulations, excluding the possibility that these changes can be caused by temperature.

In conclusion we have observed that an important difference between the four isoforms is their flexibility: CRBP1 and 2 show quite high rigidity, while CRBP3 and 4 are much more flexible. Different regions belonging to distinct isoforms are interested by different mobility. The helices' motion is remarkable in CRBP2 and 4 while it is much less evident in the other two isoforms. Comparing the surface residues distribution of the portal region, we can observe a similarity between CRBP1 and 3, and CRBP2 and 4. Evolutionary studies, mentioned in literature (Schaap, van der Vusse, & Glatz, 2002), reveal a phylogenetic correlation between isoforms 2 and 4, while CRBP1 seems to be most related with CRBP3, in agreement with our finding.

Remain to clarify if isoforms 3 and 4 are truly retinol-binding proteins or if, despite the sequence homology, these carriers can bind compounds different from retinol. Retinoid extraction from different animal tissues could be a useful method to identify putative ligands for CRBP3 and 4.

Bibliography:

Aalten, D. M. F. Van, Findlay, J. B. C., Amadei, A., & Berendsen, H. J. C. (1995). Essential dynamics of the cellular retinol-binding protein evidence for ligand-induced conformational changes. Protein Engineering, Design and Selection, 8(11), 1129–1135.

Adams, P. D., Afonine, P. V., Bunkóczi, G., Chen, V. B., Davis, I. W., Echols, N., … Zwart, P. H. (2010). PHENIX: A comprehensive Python-based system for macromolecular structure solution. Acta Crystallographica Section D: Biological Crystallography, 66(2), 213–221.

Banaszak, L et al. 1994. "Lipid-Binding Proteins: A Family of Fatty Acid and Retinoid Transport Proteins." Advances in protein chemistry 45: 89–151.

Bashor, M M, D O Toft, and F Chytil. 1973. "In Vitro Binding of Retinol to Rat-Tissue Components." Proceedings of the National Academy of Sciences of the United States of America 70(12): 3483–87.

Blomhoff, R et al. 1982. "In Vivo Uptake of Chylomicron [3H]retinyl Ester by Rat Liver: Evidence for Retinol Transfer from Parenchymal to Nonparenchymal Cells." Proceedings of the National Academy of Sciences of the United States of America 79(23): 7326–30.

Blomhoff, Rune, and Heidi Kiil Blomhoff. 2006. "Overview of Retinoid Metabolism and Function." Journal of Neurobiology 66(7): 606–30.

Buck, J., Derguini, F., Levi, E., Nakanishi, K., & Hämmerling, U. (1991). Intracellular signaling by 14-hydroxy-4,14-retro-retinol. Science (New York, N.Y.), 254(5038), 1654–6.

Bunt Milam, A. H., and J. C. Saari. 1983. "Immunocytochemical Localization of Two Retinoid-Binding Proteins in Vertebrate Retina." Journal of Cell Biology 97(3): 703–12.

Cai, Jun et al. 2012. "Solution Structure and Backbone Dynamics of Human Liver Fatty Acid Binding Protein: Fatty Acid Binding Revisited." Biophysical journal 102(11): 2585–94.

Calderone, Vito et al. 2002. "Identification and Structural Analysis of a Zebrafish Apo and Holo Cellular Retinol-Binding Protein." Journal of Molecular Biology 321(3): 527–35.

Corey, E. J., Noe, M. C., & Guzman-Perez, A. (1995). Catalytic enantioselective synthesis of (14R)-14-hydroxy-4,14-retro-retinol from retinyl acetate. Tetrahedron Letters, 36(24), 4171–4174.

Cowan, Sandra W., Marcia E. Newcomer, and T.Alwyn Jones. 1993. "Crystallographic Studies on a Family of Cellular Lipophilic Transport Proteins." Journal of Molecular Biology 230(4): 1225–46.

Crow, J a, and D E Ong. 1985. "Cell-Specific Immunohistochemical Localization of a Cellular Retinol-Binding Protein (Type Two) in the Small Intestine of Rat." Proceedings of the National Academy of Sciences of the United States of America 82(14): 4707–11.

Derguini, F et al. 1995. "13,14-Dihydroxy-Retinol, a New Bioactive Retinol Metabolite." The Journal of biological chemistry 270(32): 18875–80.

Derguini, Fadila, Koji Nakanishi, Ulrich Haemmerling, and Jochen Buck. 1994. “Intracellular Signaling Activity of Synthetic (14R)-, (14S)-, and (14RS)-14-Hydroxy-4,14-Retro-Retinol.” Biochemistry 33(3): 623–28.

Doldo, Elena et al. 2015. “Vitamin A, Cancer Treatment and Prevention: The New Role of Cellular Retinol Binding Proteins.” BioMed research international 2015: 624627.

Dolinsky, T. J., Nielsen, J. E., McCammon, J. A., & Baker, N. A. (2004). PDB2PQR: An automated pipeline for the setup of Poisson-Boltzmann electrostatics calculations. Nucleic Acids Research, 32(WEB SERVER ISS.).

Dong, Diane, Sarah E. Ruuska, David J. Levinthal, and Noa Noy. 1999. “Distinct Roles for Cellular Retinoic Acid-Binding Proteins I and II in Regulating Signaling by Retinoic Acid.” Journal of Biological Chemistry 274(34): 23695–98.

Dougherty, D. A. (2013). The cation-π interaction. Accounts of Chemical Research, 46(4), 885–893.

Emsley, P., & Cowtan, K. (2004). Coot: Model-building tools for molecular graphics. Acta Crystallographica Section D: Biological Crystallography, 60(12 I), 2126–2132.

Episkopou, V et al. 1993. “Disruption of the Transthyretin Gene Results in Mice with Depressed Levels of Plasma Retinol and Thyroid Hormone.” Proc Natl Acad Sci U S A 90(6): 2375–79.

Eriksson, U. et al. 1984. “Cellular Retinol-Binding Protein. Quantitation and Distribution.” Journal of Biological Chemistry 259(21): 13464–70.

Evans, P. (2006). Scaling and assessment of data quality. In Acta Crystallographica Section D: Biological Crystallography (Vol. 62, pp. 72–82).

Everts, Helen B., John P. Sundberg, and David E. Ong. 2005. “Immunolocalization of Retinoic Acid Biosynthesis Systems in Selected Sites in Rat.” Experimental Cell Research 308(2): 309–19.

Folli, C., Calderone, V., Ottonello, S., Bolchi, A., Zanotti, G., Stoppini, M., & Berni, R. (2001). Identification, retinoid binding, and x-ray analysis of a human retinol-binding protein. Proceedings of the National Academy of Sciences of the United States of America, 98(7), 3710–5.

Folli, C., Calderone, V., Ramazzina, I., Zanotti, G., & Berni, R. (2002). Ligand binding and structural analysis of a human putative cellular retinol-binding protein. Journal of Biological Chemistry, 277(44), 41970–41977.

Franzoni, L., Cavazzini, D., Rossi, G. L., & Lücke, C. (2010). New insights on the protein-ligand interaction differences between the two primary cellular retinol carriers. Journal of Lipid Research, 51(6), 1332–1343.

Franzoni, L., Lücke, C., Pérez, C., Cavazzini, D., Rademacher, M., Ludwig, C., … Rüterjans, H. (2002). Structure and backbone dynamics of apo- and holo-cellular retinol-binding protein in solution. Journal of Biological Chemistry, 277(24), 21983–21997.

Futterman, S, J C Saari, and S Blair. 1977. “Occurrence of a Binding Protein for 11-Cis-Retinal in Retina.” J Biol.Chem. 252(10): 3267–71.

Gasteiger, E., Hoogland, C., Gattiker, A., Duvaud, S., Wilkins, M. R., Appel, R. D., & Bairoch, A. (2005). Protein Identification and Analysis Tools on the ExPASy Server. In The Proteomics Protocols Handbook (pp. 571–607). Totowa, NJ: Humana Press.

Glatz, J. F. C., & Veerkamp, J. H. (1983). A radiochemical procedure for the assay of fatty acid binding by proteins. Analytical Biochemistry, 132(1), 89–95.

Glover, J, and E.R Redfearn. 1954. “The Mechanism of the Transformation of Beta-Carotene into Vitamin A in Vivo.” The Biochemical journal 58(331st Meeting): xv–xvi.

Guex, N., & Peitsch, M. C. (1997). SWISS-MODEL and the Swiss-PdbViewer: An environment for comparative protein modeling. Electrophoresis, 18(15), 2714–2723.

Hanhoff, Thorsten, Christian Lücke, and Friedrich Spener. 2002. “Insights into Binding of Fatty Acids by Fatty Acid Binding Proteins.” Molecular and Cellular Biochemistry 239(1–2): 45–54.

He, Xiaoqin, Joel Lobsiger, and Achim Stocker. 2009. “Bothnia Dystrophy Is Caused by Domino-like Rearrangements in Cellular Retinaldehyde-Binding Protein Mutant R234W.” Proceedings of the National Academy of Sciences of the United States of America 106(44): 18545–50.

Hodsdon, M E, and D P Cistola. 1997. “Ligand Binding Alters the Backbone Mobility of Intestinal Fatty Acid-Binding Protein as Monitored by 15N NMR Relaxation and 1H Exchange.” Biochemistry 36(8): 2278–90.

Hu, Y., & Liu, H. (2014). Case study on temperature-accelerated molecular dynamics simulation of ligand dissociation: Inducer dissociation from the Lac repressor protein. Journal of Physical Chemistry A, 118(39), 9272–9279.

Humphrey, W., Dalke, A., & Schulten, K. (1996). VMD: Visual molecular dynamics. Journal of Molecular Graphics, 14(1), 33–38.

Imam, A et al. 2001. “Retinoids as Ligands and Coactivators of Protein Kinase C Alpha.” FASEB journal : official publication of the Federation of American Societies for Experimental Biology 15(1): 28–30.

Jamison, R S, M E Newcomer, and D E Ong. 1994. “Cellular Retinoid-Binding Proteins: Limited Proteolysis Reveals a Conformational Change upon Ligand Binding.” Biochemistry 33(10): 2873–79.

Jamison, R. S. et al. 1995. “Test of the Contribution of an Amino-Aromatic Hydrogen Bond to Protein Function.” Biochemistry 34(35): 11128–32.

Jiang, Weiya, and Joseph L. Napoli. 2012. “Reorganization of Cellular Retinol-Binding Protein Type 1 and Lecithin:retinol Acyltransferase during Retinyl Ester Biosynthesis.” Biochimica et Biophysica Acta (BBA) - General Subjects 1820(7): 859–69.

Kabsch, W. (2010). XDS. Acta Crystallogr D Biol Crystallogr, 66(Pt 2), 125–132.

Kakkad, B. P., and D. E. Ong. 1988. “Reduction of Retinaldehyde Bound to Cellular Retinol-Binding Protein (Type II) by Microsomes from Rat Small Intestine.” Journal of Biological Chemistry 263(26): 12916–19.

Kiefer, C et al. 2001. “Identification and Characterization of a Mammalian Enzyme Catalyzing the Asymmetric Oxidative Cleavage of Provitamin A.” The Journal of biological chemistry 276(17): 14110–16.

Kono, Nozomu, and Hiroyuki Arai. 2015. “Intracellular Transport of Fat-Soluble Vitamins A and E.” Traffic 16(1): 19–34.

Lapshina, Elena A., Olga V. Belyaeva, Olga V. Chumakova, and Natalia Y. Kedishvili. 2003. “Differential Recognition of the Free versus Bound Retinol by Human Microsomal Retinol/sterol Dehydrogenases: Characterization of the Holo-CRBP Dehydrogenase Activity of RoDH-4.” Biochemistry 42(3): 776–84.

Lebedev, A. A., Vagin, A. A., & Murshudov, G. N. (2007). Model preparation in MOLREP and examples of model improvement using X-ray data. In Acta Crystallographica Section D: Biological Crystallography (Vol. 64, pp. 33–39).

Levin, L. B.-A., Nachliel, E., Gutman, M., & Tsfadia, Y. (2009). Molecular dynamics study of the interaction between fatty acid binding proteins with palmitate mini-micelles. Molecular and Cellular Biochemistry, 326, 29–33.

Levin, M S et al. 1988. “Comparison of the Ligand Binding Properties of Two Homologous Rat Apocellular Retinol-Binding Proteins Expressed in Escherichia Coli.” The Journal of biological chemistry 263(33): 17715–23.

Li, Ellen et al. 1991. “Fluorine Nuclear Magnetic Resonance Analysis of the Ligand Binding Properties of Two Homologous Rat Cellular Retinol-Binding Proteins Expressed in Escherichia Coli.” Journal of Biological Chemistry 266(6): 3622–29.

Li, Ellen, and Patrick Tso. 2003. “Vitamin A Uptake from Foods.” Current opinion in lipidology 14(3): 241–47.

Lovell, S. C., Davis, I. W., Arendall, W. B., de Bakker, P. I. W., Word, J. M., Prisant, M. G., … Richardson, D. C. (2003). Structure validation by Calpha geometry: phi,psi and Cbeta deviation. Proteins, 50(3), 437–450.

Lu, J et al. 1999. "The Structure and Dynamics of Rat Apo-Cellular Retinol-Binding Protein {II} in Solution: Comparison with the X-Ray Structure." J Mol Biol 286: 1179–95.

Lu, J., Lin, C.-L., Tang, C., Ponder, J. W., Kao, J. L. ., Cistola, D. P., & Li, E. (2000). Binding of retinol induces changes in rat cellular retinol-binding protein II conformation and backbone dynamics. Journal of Molecular Biology, 300(3), 619–632.

MacDonald, P N, and D E Ong. 1987. "Binding Specificities of Cellular Retinol-Binding Protein and Cellular Retinol-Binding Protein, Type II." J Biol Chem 262(22): 10550–56.

MacDonald, P. N., & Ong, D. E. (1988). Evidence for a lecithin-retinol acyltransferase activity in the rat small intestine. The Journal of Biological Chemistry, 263(25), 12478–82. Retrieved from

Malpeli, G., Stoppini, M., Zapponi, M. C., Folli, C., & Berni, R. (1995). Interactions with retinol and retinoids of bovine cellular retinol-binding protein. European Journal of Biochemistry / FEBS, 229(2), 486–493.

Mittag, T., Franzoni, L., Cavazzini, D., Schaffhausen, B., Rossi, G. L., & Günther, U. L. (2006). Retinol modulates site-specific mobility of apo-cellular retinol-binding protein to promote ligand binding. Journal of the American Chemical Society, 128(30), 9844–9848.

Moore, T. 1930. "Vitamin A and Carotene: The Absence of the Liver Oil Vitamin A from Carotene. VI. The Conversion of Carotene to Vitamin A in Vivo." The Biochemical journal 24(3): 692–702.

Murshudov, G. N., Skubák, P., Lebedev, A. A., Pannu, N. S., Steiner, R. A., Nicholls, R. A., … Vagin, A. A. (2011). REFMAC5 for the refinement of macromolecular crystal structures. Acta Crystallographica Section D: Biological Crystallography, 67(4), 355–367.

Napoli, J L, and K R Race. 1988. "Biogenesis of Retinoic Acid from Beta-Carotene. Differences between the Metabolism of Beta-Carotene and Retinal." The Journal of biological chemistry 263(33): 17372–77.

Napoli, Joseph L. 1999. "Interactions of Retinoid Binding Proteins and Enzymes in Retinoid Metabolism." Biochimica et Biophysica Acta - Molecular and Cell Biology of Lipids 1440(2–3): 139–62.

Newcomer, M E, and D E Ong. 2000. "Plasma Retinol Binding Protein: Structure and Function of the Prototypic Lipocalin." Biochim Biophys Acta 1482(1–2): 57–64.

Nossoni, Zahra et al. 2014. "Structures of Holo Wild-Type Human Cellular Retinol-Binding Protein II (hCRBPII) Bound to Retinol and Retinal." Acta Crystallographica Section D: Biological Crystallography 70(12): 3226–32.

Noy, N. 2000. “Retinoid-Binding Proteins: Mediators of Retinoid Action.” The Biochemical journal 348 Pt 3: 481–95.

Olson, J A, and O Hayaishi. 1965. “The Enzymatic Cleavage of Beta-Carotene into Vitamin A by Soluble Enzymes of Rat Liver and Intestine.” Proceedings of the National Academy of Sciences of the United States of America 54(5): 1364–70.

Oostenbrink, C., Villa, A., Mark, A. E., & Van Gunsteren, W. F. (2004). A biomolecular force field based on the free enthalpy of hydration and solvation: The GROMOS force-field parameter sets 53A5 and 53A6. Journal of Computational Chemistry, 25(13), 1656–1676.

Palczewski, K et al. 1994. “Rod Outer Segment Retinol Dehydrogenase: Substrate Specificity and Role in Phototransduction.” Biochemistry 33(46): 13741–50.

Penzes, Peter, and Joseph L. Napoli. 1999. “Holo-Cellular Retinol-Binding Protein: Distinction of Ligand-Binding Affinity from Efficiency as Substrate in Retinal Biosynthesis.” Biochemistry 38(7): 2088–93.

Posch, K C, M H Boerman, R D Burns, and J L Napoli. 1991. “Holocellular Retinol Binding Protein as a Substrate for Microsomal Retinal Synthesis.” Biochemistry 30(25): 6224–30.

Ragona, L., Pagano, K., Tomaselli, S., Favretto, F., Ceccon, A., Zanzoni, S., … Molinari, H. (2014). The role of dynamics in modulating ligand exchange in intracellular lipid binding proteins. Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics, 1844(7), 1268–1278.

Randolph, T. W. (2012). The two faces of His-tag: Immune response versus ease of protein purification. Biotechnology Journal, 7(1), 18–19.

Ronne, H et al. 1983. “Ligand-Dependent Regulation of Intracellular Protein Transport: Effect of Vitamin a on the Secretion of the Retinol-Binding Protein.” The Journal of cell biology 96(3): 907–10.

Sacchettini, J C, J I Gordon, and L J Banaszak. 1989. “Crystal Structure of Rat Intestinal Fatty-Acid-Binding Protein. Refinement and Analysis of the Escherichia Coli-Derived Protein with Bound Palmitate.” Journal of molecular biology 208(2): 327–39.

Sayle, R. A., & Milner-White, E. J. (1995). RASMOL: biomolecular graphics for all. Trends in Biochemical Sciences, 20(9), 374. Retrieved from

Schaap, F. G., van der Vusse, G. J., & Glatz, J. F. C. (2002). Evolution of the family of intracellular lipid binding proteins in vertebrates. Molecular and Cellular Biochemistry, 239(1–2), 69–77.

Schaefer, W H et al. 1989. “Purification, Primary Structure Characterization, and Cellular Distribution of Two Forms of Cellular Retinol-Binding Protein, Type II from Adult Rat Small Intestine.” The Journal of biological chemistry 264(7): 4212–21.

Schug, Thaddeus T. et al. 2007. “Opposing Effects of Retinoic Acid on Cell Growth Result from Alternate Activation of Two Different Nuclear Receptors.” Cell 129(4): 723–33.

Schüttelkopf, A. W., & Van Aalten, D. M. F. (2004). PRODRG: A tool for high-throughput crystallography of protein-ligand complexes. Acta Crystallographica Section D: Biological Crystallography, 60(8), 1355–1363.

Silvaroli, J. A., Arne, J. M., Chelstowska, S., Kiser, P. D., Banerjee, S., & Golczak, M. (2016). Ligand Binding Induces Conformational Changes in Human Cellular Retinol-binding Protein 1 (CRBP1) Revealed by Atomic Resolution Crystal Structures. Journal of Biological Chemistry, 291(16), 8528–8540.

Strohalm, M., Kavan, D., Novák, P., Volný, M., & Havlíček, V. (2010). mMass 3: A Cross-Platform Software Environment for Precise Analysis of Mass Spectrometric Data. Analytical Chemistry, 82(11), 4648–4651.

Tarter M., Capaldi S., Carrizo M.E, Ambrosi E., Perduca M., and Monaco L.H. 2007. “Crystal Structure of Human Cellular Retinol-Binding Protein 1.1 to 1.2 A Resolution.” Proteins 70(2): 1626–30.

Tsfadia, Y., Friedman, R., Kadmon, J., Selzer, A., Nachliel, E., & Gutman, M. (2007). Molecular dynamics simulations of palmitate entry into the hydrophobic pocket of the fatty acid binding protein. FEBS Letters, 581(6), 1243–1247.

Vaezeslami, S., Mathes, E., Vasileiou, C., Borhan, B., & Geiger, J. H. (2006). The Structure of Apo-wild-type Cellular Retinoic Acid Binding Protein II at 1.4 Å and its Relationship to Ligand Binding and Nuclear Translocation. Journal of Molecular Biology, 363(3), 687–701.

Van Der Spoel, D., Lindahl, E., Hess, B., Groenhof, G., Mark, A. E., & Berendsen, H. J. C. (2005). GROMACS: Fast, flexible, and free. Journal of Computational Chemistry.

Winter, N S, J M Bratt, and L J Banaszak. 1993. “Crystal Structures of Holo and Apo-Cellular Retinol-Binding Protein II.” Journal of molecular biology 230(4): 1247–59.

Wyss, R., F. Bucheli, and R. Hartenbach. 1998. “Determination of 13-Cis-3-Hydroxyretinoic Acid, All-Trans-3-Hydroxy-Retinoic Acid and Their 4-Oxo Metabolites in Human and Animal Plasma by HPLC with Automated Column Switching and UV Detection.” Journal de Pharmacie de Belgique 53(3): 268.

Xu, Z, D A Bernlohr, and L J Banaszak. 1993. “The Adipocyte Lipid-Binding Protein at 1.6-A Resolution. Crystal Structures of the Apoprotein and with Bound Saturated and Unsaturated Fatty Acids.” The Journal of biological chemistry 268(11): 7874–84.

Yost, R W, E H Harrison, and A C Ross. 1988. “Esterification by Rat Liver Microsomes of Retinol Bound to Cellular Retinol-Binding Protein.” The Journal of biological chemistry 263(35): 18693–701.

Young, a C et al. 1994. “Structural Studies on Human Muscle Fatty Acid Binding Protein at 1.4 A Resolution: Binding Interactions with Three C18 Fatty Acids.” Structure (London, England : 1993) 2(6): 523–34.

Zanotti, Giuseppe et al. 2008. “Structural and Mutational Analyses of Protein-Protein Interactions between Transthyretin and Retinol-Binding Protein.” FEBS Journal 275(23): 5841–54.

Zanotti, Giuseppe, and Rodolfo Berni. 2004. “Plasma Retinol-Binding Protein: Structure and Interactions with Retinol, Retinoids, and Transthyretin.” Vitamins and Hormones 69: 271–95.

Zimmerman, A. W., Rademacher, M., Ruterjans, H., Lucke, C., & Veerkamp, J. H. (1999). Functional and conformational characterization of new mutants of heart fatty acid-binding protein. Biochemical Journal, 344, 495–501.

Zimmerman, A., van Moerkerk, H., & Veerkamp, J. (2001). Ligand specificity and conformational stability of human fatty acid-binding proteins. The International Journal of Biochemistry & Cell Biology, 33(9), 865–76.

Ringraziamenti

Desidero ringraziare la Prof.ssa Eugenia Polverini per la grande professionalità e il fondamentale aiuto nella progettazione delle simulazioni; il Prof. Rodolfo Berni per avermi permesso di intraprendere questo entusiasmante percorso nel quale ho potuto unire aspetti computazionali e sperimentali, imparando molto. La Prof.ssa Claudia Folli per i suggerimenti dati e per avermi sollevato il morale nei momenti più angoscianti.

Non posso dimenticare i mei colleghi, Alberto e Koro, con i quali ho condiviso risate, attimi di panico e qualche tirocinante originale.

Ringrazio la mia famiglia, che da sempre mi sostiene e sopporta i miei momenti di tensione e sconforto. Un grazie alla mia Gioia per l'infinita pazienza.

Infine, ma non da ultimi, ringrazio tutti i miei amici, i miei soci di tennis, e un particolare ringraziamento ai miei compagni d'avventura del Circolo di Volontariato Kaleidos.