UNIVERSITA' DEGLI STUDI DI VERONA

DIPARTIMENTO DI
BIOTECNOLOGIE

DOTTORATO DI RICERCA IN
BIOTECNOLOGIE MOLECOLARI, INDUSTRIALI ED AMBIENTALI

CICLO XXIII

TESI DI DOTTORATO

NMR Structural and interaction studies of bile acid binding proteins complexed with physiological ligands and bile acid-derived contrast agents

S.S.D. CHIM/06

Coordinatore: Prof. Roberto Bassi

Tutor: Prof.ssa Henriette Molinari

Dottoranda: Dott.ssa Serena Annalisa Zanzoni

Contents

Abstract

The research project of my PhD concerns essentially the Nuclear Magnetic Resonance (NMR) structural studies of protein-ligand complexes. Studying protein interactions at the molecular level is crucial to the understanding of many biological processes, such as human diseases and drug design. NMR spectroscopy is particularly well suited to the investigation, at the atomic level, of the interactions between proteins and other molecules in solution. NMR can be used to evaluate the structural, thermodynamic and kinetic aspects of a binding reaction, even for weak protein-ligand interactions.

In this contest, during my PhD I addressed the structural investigation and the study of the molecular determinants of binding of different bile acid binding proteins (BABPs) and their native ligands or synthetic drugs.

BABPs are small cytosolic proteins that display their function in the hepatocytes and enterocytes where they act as bile acids transporters participating to the enterohepatic circulation.

The first part of this thesis is related to the structural determination of chicken liver-BABP (cL-BABP) in complex with a bile acid-based gadolinium(III)-chelate, a potential hepatospecific contrast agent for magnetic resonance imaging (MRI). The rationale of this study derived from a search for new hepatospecific MRI contrast agents for the discrimination and diagnosis of focal lesions or hepatic malignancies.

The second part of this thesis is focused on the characterization of the interactions of chicken ileal BABP (cI-BABP) with bile acids. This study has been performed using different approaches, such as NMR spectroscopy and calorimetric measurements.

The third part of the my thesis work is based on the NMR structural determination of the ternary complex between cI-BABP and two molecules of glycochenodeoxycholic acid (GCDA). Finally, a single

and double mutant of cI-BABP (A101S and H99Q/A101S) were produced to investigate the molecular determinants of binding cooperativity.

The obtained data allow shedding light on the chemical basis of intracellular bile acid transport. This notion may open new avenues for exploration of strategies for prevention and treatment of metabolic diseases.

Riassunto

Il progetto svolto durante il mio Dottorato di Ricerca riguarda essenzialmente la cartterizzazione strutturale di complessi proteina-legante tramite spettroscopia di Risonanza Magnetica Nucleare (NMR). Studiare le interazioni proteiche a livello molecolare è di fondamentale importanza per acquisire informazioni su molti processi biologici, quali lo sviluppo di malattie nell'uomo o il relativo disegno di farmaci. La Risonanza Magnetica Nucleare è una tecnica particolarmente adatta per la caratterizzazione, a livello atomico, delle interazioni tra proteine e altre molecole in soluzione. La Risonanza Magnetica Nucleare può essere utilizzata per valutare gli aspetti strutturali, termodinamici e cinetici di una reazione di interazione proteina-legante e può rilevare anche le interazioni più deboli.

In questo contesto, durante il mio Dottorato di Ricerca ho affrontato l'indagine strutturale e lo studio dei determinanti molecolari dell'interazione di differenti proteine, le bile acid binding proteins (BABPs) in complesso sia con i loro leganti fisiologici sia legate a farmaci sintetitici. Le BABPs sono piccole proteine citosoliche che svolgono la loro funzione negli epatociti ed enterociti, dove agiscono come trasportatori intracellulari degli acidi biliari, permettendo il loro ricircolo tramite la circolazione enteroepatica.

Nella prima prima parte di questa tesi viene descritta la determinazione strutturale tramite NMR della BABP di fegato di pollo (cL-BABP) in complesso con l'addotto acido biliare-Gd-DTPA, un potenziale agente di contrasto epatospecifo per la Risonanza Magnetica per Immagine (MRI). Il razionale di questo studio deriva da una ricerca di nuovi agenti di contrasto epatospecifici per la discriminazione e la diagnosi di lesioni focali o tumori nel fegato, tramite la Risonanza Magnetica per Immagine.

La seconda parte di questa tesi è focalizzata sulla caratterizzazione delle interazioni della BABP ileale di pollo (cI-BABP) con acidi biliari.

Questo studio è stato svolto utilizzando diversi approcci, così come la spettroscopia NMR e misure calorimetriche.

La terza parte del mio lavoro di tesi si basa sulla determinazione strutturale tramite NMR del complesso ternario tra la CI-BABP e due molecole di acido glicochenodeossicolico (GCDA). Infine, un singolo e doppio mutante della proteina CI-BABP (A101S e H99Q/A101S) sono stati prodotti per studiare i determinanti molecolari della cooperatività di legame.

I dati ottenuti permettono di capire le basi chimiche del trasporto intracellulare degli acidi biliari. Questa conoscenza può aprire nuove vie per l'esplorazione di strategie di prevenzione e trattamento delle malattie metaboliche.

Chapter 1: Introduction

1.1 Bile acids

Bile acids (BAs) are amphipathic steroidal compounds derived from the enzymatic catabolism of cholesterol in the liver and are the most important constituents of bile. The predominant bile acids in humans are chenodeoxycholic acid (CDCA) and cholic acid (CA). These primary BAs are synthesized within hepatocytes and conjugated at the terminal (C24) carboxyl group with the amino acids taurine and glycine. This amidation reaction has the effect of increasing bile acid solubility and renders them more readily excretable into bile. Bile is an important fluid secreted by the liver into the intestine and serves two main functions; digestion and absorption of lipid-soluble nutrients and elimination of endogenous and exogenous substances such as bilirubin, phospholipids, cholesterol, drugs and toxins (1). Bile is released into the small intestine upon food consumption. In the ileum, after facilitating and lipid-soluble vitamins absorption, most of the primary bile acids are absorbed from intestinal lumen and transported back to the liver, for reuse. This BAs recycling is called enterohepatic circulation (2). The existence of the recycling system ensures the presence of adequate concentrations of BAs (pool size) in the intestinal lumen for digestion: each molecule is utilized many times. About 5% of the pool size of BAs is lost into the large intestine and finally into the feces. The fecal loss of BAs, which is compensated by de novo BA biosynthesis in the liver, is involved in the biliary secretion, as it permits the elimination of substrates that cannot be eliminated via renal excretion (3).

1.1.1 Bile acid biosynthesis and modification

The fraction of BAs that has to be synthesized to preserve the adequate pool size in the intestine comes mainly from catabolism of cholesterol. Cholesterol degradation to primary BAs can be initiated by either a microsomal cholesterol 7alpha-hydroxylase (the classic or neutral pathway) or by mitochondrial sterol 27-hydroxylase (the alternative or acidic pathway). In the classic pathway, the principal steps leading to

formation of primary BAs include the saturation of the double bond, epimerization of the 3β-hydroxyl group, and hydroxylation at the 7α and 12α-positions on the sterol nucleus.

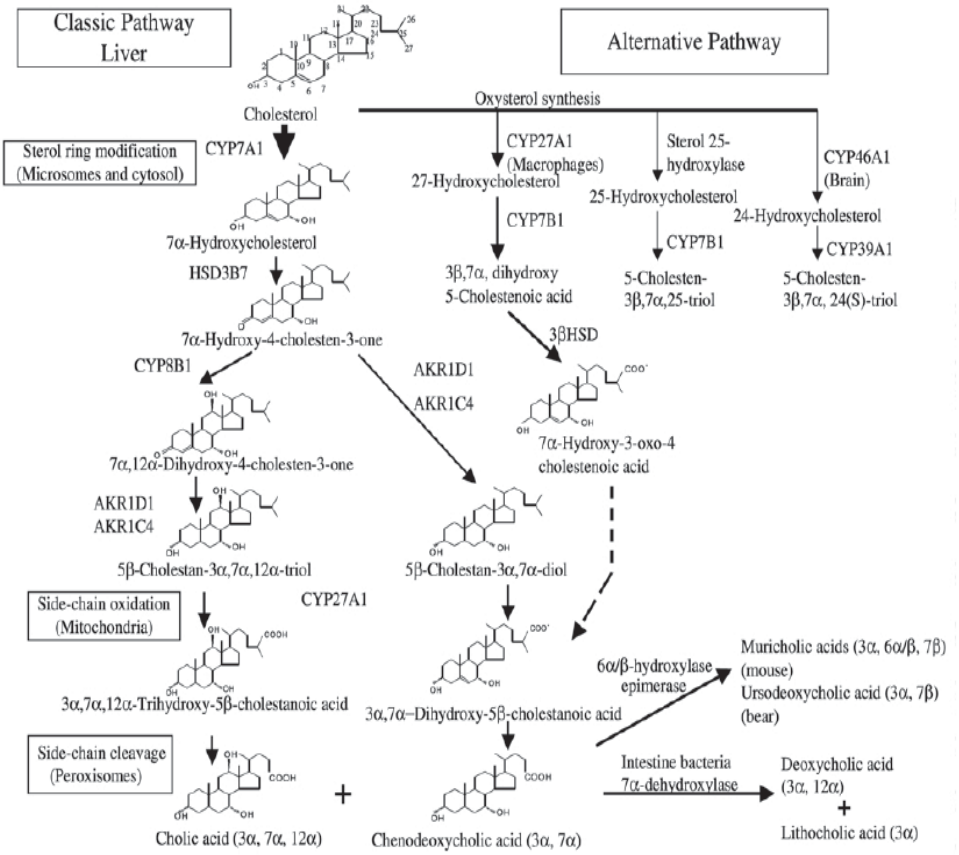
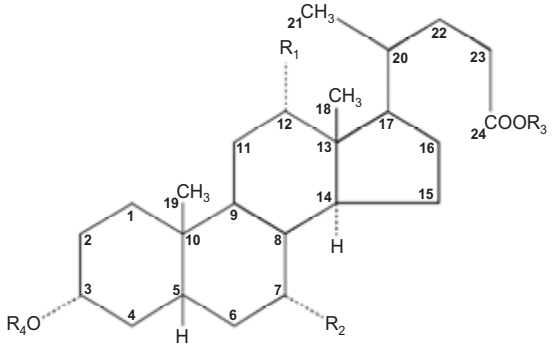


Figure 1.1 Bile acids biosynthetic pathways. Only major regulatory steps and enzymes are shown (4).

These modifications precede the oxidation and cleavage of the side-chain. On the contrary, in the alternative pathway, side-chain oxidation precedes steroid ring modification. The immediate products of the BA synthetic pathways are CA and CDCA, referred to as primary bile acids. While the classic pathway of bile acids biosynthesis is limited to hepatocytes, the alternative pathway exists in all tissues. In figure 1.1 is reported a

schematic version of these complex metabolic pathways.
After their biosynthesis from cholesterol, BAs are conjugated with either glycine or taurine to increase the solubility for secretion into the bile. Furthermore, conjugation with glycine or taurine lowers their pKa and therefore renders conjugated BAs fully ionized at physiological pH. The result is that conjugated BAs are membrane impermeable and high BA concentrations can persist in bile and intestine.
During the enterohepatic circulation, the structure of primary BAs is altered by the intestinal microflora into a variety of metabolites (5). Bacterial biotransformation regards both the side chain and the nucleus of BAs. The known BAs biotransformations include: deconjugation on the side chain and elimination of hydroxyl group at C-7 on the nucleus. Thus conjugated CA is converted to deoxycholic acid (DCA) and similarly the transformation of conjugated CDCA results in the formation of lithocholic acid (LCA). Because DCA and LCA are formed from primary BAs, they are also called "secondary bile acids". In figure 1.2 a schematic representation of the BA molecule and its most common substituents at different position is reported.



	R_1	R_2
Cholic acid	OH	OH
Chenodeoxycholic acid	H	OH
Deoxycholic acid	OH	H
Lithocholic acid	H	H

R_3= Taurine, glycine or hydrogen
R_4= Sulphuric acid, glucuronic acid or hydrogen

Figure 1.2 Schematic representation of the bile acid molecule. The most common substituents for positions C-3, C-7, C-12 and C-24 are reported.

In the hepatocyte, secondary BAs undergo different fates. DCA is conjugated with glycine or taurine and secreted into bile with the primary BAs. Because LCA is poorly water soluble and rather toxic to cells, it is conjugated with glycine or taurine and sulfated at the C-3 position.

These "double conjugates" (sulfolithocholylglycine and sulfolithocholyltaurine) are excreted into bile but, in contrast to the conjugates of primary BAs and of DCA, they are not efficiently absorbed from the small intestine. As a result, they are quickly eliminated from the organism by biliary excretion (6, 7).

1.1.2 The enterohepatic circulation and bile acids transporter

Conservation of the pool size of BAs within the enterohepatic circulation requires the coordinate action of several transporter proteins expressed at the apical and basolateral membranes of liver and intestinal epithelial cells. Enterohepatic circulation of bile acids is fundamentally composed of two major processes: secretion from the liver and absorption from the intestine. After biosynthesis, conjugated BAs are secreted into the canalicular space between hepatocytes (Figure 1.3) via two ATP-binding cassette transporters: bile acids export pump (BSEP), that carries monovalent bile acids and multidrug resistance protein 2 (MRP2), that transports divalent bile salts, such as sulfated bile salt conjugates (8). Canalicular BA secretion represents the rate-limiting step in bile formation. Whereas BA concentrations within the hepatocyte are in the micromolar range, canalicular BA concentrations are more than 1,000-fold higher, necessitating active transport across the canalicular hepatocyte membrane. Thus, the BAs are secreted against steep concentration gradients in the gallbladder and are driven by ATP hydrolysis. In the gallbladder, most of the bile salt pool is stored during the fasting state. After ingestion of a meal, the presence of fats and proteins in the duodenum stimulates the release of endogenous neuro-hormones such as cholecystokinin (CCK). The release of CCK is correlated with contraction of the gallbladder (9). The end result is that

gallbladder bile is delivered to the duodenum, where the BAs carry out their role in emulsifying dietary lipids.

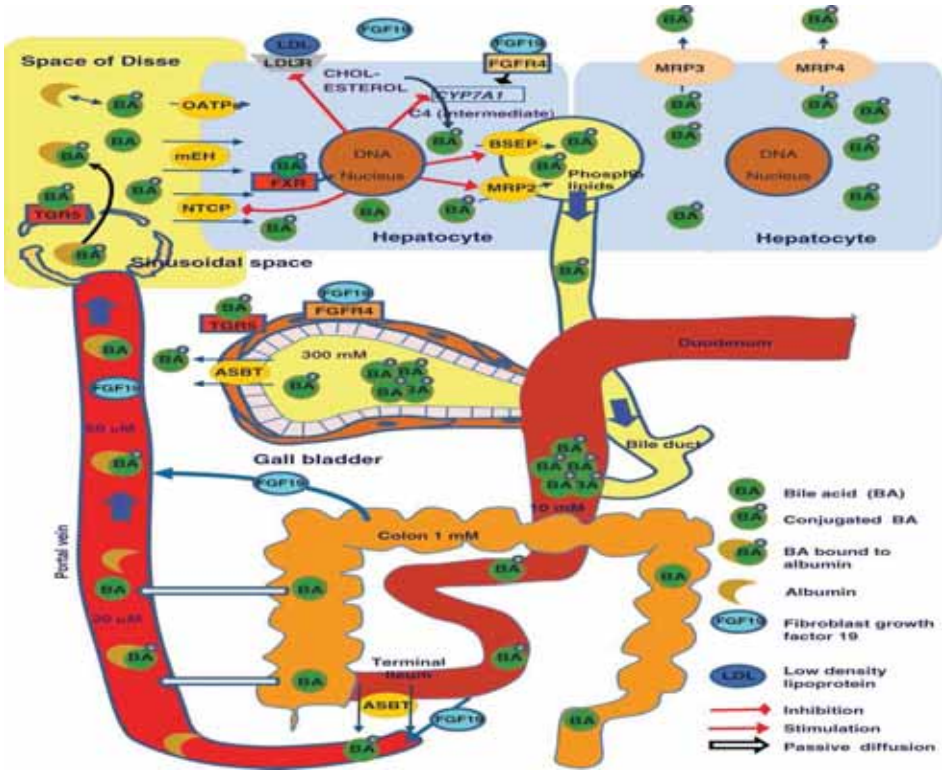


Figure 1.3 Enterohepatic circulation of bile acids (10).

The second important process of enterohepatic circulation is the absorption of BAs from the intestine. Conjugated BAs enter in the enterocytes by active uptake via a Na^+-dependent mechanism. The apical sodium-dependent bile salt transporter (ASBT) (11) imports BAs coupled to Na^+ absorption in a 1:2 stoichiometry (12). Unconjugated BAs that are formed by bacterial biotransformation for their elimination in faeces are membrane permeable and thus the absorption from the colon is a passive mechanism. This non-carried mediated mechanism accounts for a small fraction of intestinal BA conservation (13).

In the enterocytes, intracellular carrier proteins mediate the BAs transport from the apical membrane towards the basolateralal pole. Ileal bile acid binding proteins (I-BABP) have been proposed to act as the putative BA carriers in the enterocytes (14-17).

The basolater efflux of BAs from enterocytes into the portal vein is mediated by the organic solute transporter α and β (OSTα-OSTβ) (18). OST-mediated transport requires the formation of heterodimer in which the two subunits α and β interact to generate a functional transporter. OSTα and OSTβ have been shown to transport taurine and glycine conjugated bile acids in addition to other sterols (prostaglandin E2, estrone-3-sulfate and dehydroepiandrosterone sulphate).

Once in the portal vein, BAs associate with albumin and reach the space of Disse through the large fenestrea (large pores) of the liver sinusoids. After their dissociation from albumin, unconjugated BAs can traverse cell membranes by passive diffusion, while taurine or glycine conjugated BAs require an active transport mechanism for cellular uptake (19). Hepatic uptake of BAs occurs against a 5- to 10-fold concentration gradient between the portal blood plasma and the hepatocyte cytosol and is mediated by both sodium dependent and sodium-independent mechanisms (8, 20). Na^+ dependent uptake activity has been shown to be predominantly mediated by Na^+ dependent taurocholic cotransporting polypeptide (NTCP) (8, 20, 21). BAs uptake via NTCP is unidirectional with a sodium-to-taurocholate stoichiometry of 2:1, i.e., cotransport of two Na^+ with one taurocholate molecule, and it is driven by a Na^+ gradient that is maintained by the activity of Na^+/K^+ ATPase (8, 21). A further protein involved with NTCP in sodium-dependent bile salt uptake across the basolateral hepatocyte membrane is the enzyme microsomal epoxide hydrolase (mEH) (22). The contributions of mEH and NTCP to BAs transport in humans depend on the substrate specificities of these two transport proteins. While NTCP is able to transport taurocholate, its ability to transport glycocholate is significantly reduced. In contrast mEH is shown to transport preferentially glycocholate (23). Sodium-independent hepatocellular uptake of bile salts is mediated by several

members of the family of Organic Anion Transporting Polypeptides (OATPs). OATPs are multispecific transporter systems that mediate, in addition to conjugated and unconjugated BAs, hepatocellular uptake of a vast variety of other amphipathic organic compounds, including bromosulfophthalein (BSP), bilirubin, hormones and numerous drugs (21, 24). Human liver expresses four OATPs (OATP-A, OATP-B, OATP-C, and OATP8), although all with different expression levels and significance for BAs uptake (21, 24, 25).

The intracellular transport of BAs from the basolateral to the canalicular membrane of the hepatocytes has been proposed to be mediated by two distinct processes: intracellular trafficking and vesicle-mediated transport (8, 21). The intracellular trafficking represents the major intracellular transport process in physiological conditions, whereas the vesicle-mediated pathway is involved at high bile salt concentrations to prevent efflux from the cytosol back into the blood (21). Intracellular trafficking is mediated by intracellular binding proteins and it is driven by basolateral to canalicular BAs concentration gradient (21, 26).

1.1.3 Physiological function of bile acids

The first and most important BAs function is the elimination of cholesterol, through either the degradation of cholesterol to bile acids or the solubilisation of cholesterol in mixed micelles for the elimination via the fecal route. In the small intestine, BAs promote dietary lipid absorption by solubilizing lipids and their digestion products as mixed micelles. BAs are the most important organic solutes of bile end their secretion into bile represents the major driving force for bile flow. Another BAs function is related to the stimulation of biliary phospholipid secretion into bile. The presence of phospholipids in bile results in a greater fraction of bile acids existing in the form of mixed micelles, characterized by a low critical micelle concentration (CMC). A low monomer concentration of BAs in bile is important to prevent cell damage (27, 28).

1.1.4 Bile acids as regulatory molecules

Besides their roles in dietary lipid absorption and cholesterol homeostasis, BAs are increasingly recognized as important signalling molecules in liver, bile ducts and the intestine. The most important role of BAs as regulatory molecule is their implication in the BA homeostasis. The maintenance of BA homeostasis is mediated by several mechanisms, which involve on one side the feedback regulation of BA biosynthesis and on the other side the expression of hepatic and intestinal transporters. In the liver BAs decrease their own biosynthesis through an elaborate feedback inhibitory circuit, which starts by BA-mediated activation of farnesoid X receptor (FXR) (29). The activation of FXR induces the expression of a suppressor protein termed SHP (small heterodimer partner). SHP is an atypical nuclear hormone receptor that lacks of DNA binding domain, but display its regulatory function when bound to other transcriptional factors. SHP interacts with two nuclear receptors, the hepatic nuclear factor 4 (HNF4) and the liver receptor homolog-1 (LRH-1). The formation of this heterodimer complex results in the inhibition of the transcription of CYP7A1 (10, 30-32). A schematic representation of negative feedback regulation of bile acid and cholesterol biosynthesis is shown in figure 1.4.

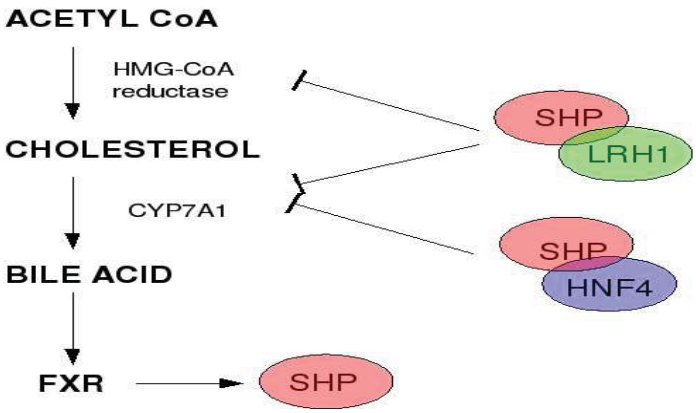


Figure 1.4 Negative feedback regulation of bile acid and cholesterol biosynthesis. Only the rate-limiting enzyme of biosynthesis are reported.

In the intestine, the activation of FXR by BAs induces also the expression of an intestinal hormone, fibroblast growth factor 19 (FGF19). Stimulation of hepatic FGF receptor 4 (FGFR4) by FGF19 leads to receptor dimerization, autophosphorylation and activation of mitogen-activated protein kinase (MAPK) pathways resulting in the repression of CYP7A1 (30).

Other factors, such as endotoxin, tumor necrosis factor (TNF) and Interleukin-1 (IL-1) are involved in cell signalling pathways that strongly decrease CYP7A mRNA level and activity (31).

In addition to the regulation of synthesis FXR plays a central role also in the flux of BAs through the regulation of the expression of hepatic and intestinal BA transporters (Figure 1.5). In order to protect the cell from the accumulations of toxic levels of BAs, FXR stimulates BAs efflux and inhibits their uptake in a concerted manner. In particular, in the liver FXR induces the expression of BSEP (33) and MRP2 (34) to increase the BAs export. At the same time the over-expression of SHP, forming heterodimeric complex with transcription factor (RAR/RXR), inhibits the transcription of NTCP and thereby the basolateral uptake of BAs (35). Similarly, in the enterocytes the activation of FXR induces the expression of transporter involved in the basolater efflux of BAs, i.e. the organic solute OSTα and β (36) and in the intracellular carrier, the IBABP (37). In addition, FXR is responsible for reducing ASBT expression level and consequently the apical uptake of BAs (38). Thus, in the enterocytes FXR acts in prevention of intracellular accumulation of BAs with inhibition of genes involved in the intestinal absorption and activation of transcription of genes encoding BAs export transpoters.

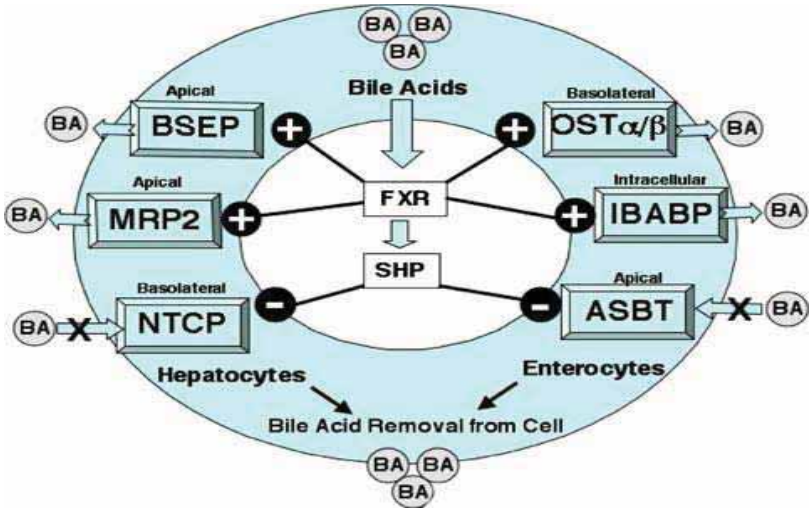


Figure 1.5 Bile acid induced regulation of hepatic and intestinal bile acid transporters.

Besides its roles in BA homeostasis, there are indications that BAs regulate gluconeogenesis, glycogen synthesis, insulin sensitivity (39) and lipid homeostasis mainly through activation of FXR (40, 41). BAs can also modulate energy metabolism increasing cAMP production. BAs bind TGR5, a G protein-coupled receptor (GPCR) which responds to BAs by inducing receptor internalization, activation of MAPK pathways, and cAMP production (42, 43). It has also been observed that administration of BAs to mice increased energy expenditure in brown adipose tissue and prevented development of obesity and insulin resistance (44).

1.2 The intracellular bile acid binding proteins

It has been proposed that specific carrier proteins might mediate the intracellular transport of cytotoxic BAs within hepatocytes and ileocytes during the enterohepatic circulation. These carrier proteins are the Bile Acid Binding Proteins (BABP) belonging to the Fatty Acid Binding Proteins (FABP) family. These are abundantly expressed 14-15 KDa proteins that bind with high affinity hydrophobic ligands, such as fatty acids, eicosanoids and other lipids (45, 46). Despite the considerable

differences in their primary structure, the tertiary structure of all FABPs is highly conserved. The common structural feature is a 10-stranded β-barrel, made of two orthogonal antiparallel 5-stranded sheets that form the "clam" shaped binding cavity (47). The opening of this clam, considered the portal domain, is framed on one side with the N-terminal helix-turn-helix domain, a common structural motif of all FABPs. Based on their primary structure and the first tissue from which they were isolated, FABPs have been classified into several types including heart, liver, intestinal, adipocyte, myelin, ileal and epidermal (48, 49). The distinctive pattern of tissue expression of the different FABP types, as well as the co expression of different FABPs in a single cell type, suggests that each FABP has a specific function in cellular metabolism (50).

1.2.1 FABPs and BABPs in the hepatocytes

In the liver, two paralogous group of FABPs have been described: liver-FABP (L-FABP), extensively characterized in mammals, and liver-BABP (initially called liver basic-FABP) that has been described in several non mammalian vertebrates, but has not yet been found in mammalian liver (51). The degree of amino acid identity between two subgroups range from 39% to 42%. Mammalian L-FABP, which binds a broad range of ligand molecules (acyl-CoAs, heme, squalene, bilirubin and certain eicosanoids), is the only FABP that forms a complex with two fatty acid molecules at the same time (Figure 1.6), whereas all the other members of the FABP family have a single fatty-acid-binding site (52-54). Recently NMR binding experiments indicated clearly that hL-FABP is able to bind not only fatty acid, but also glychochenodeoxycholic acid, one of two most abundant physiological BAs. It has been suggested that hL-FABP is involved in the physiological regulation of fatty acid and bile acid metabolism (51). Therefore this protein has been proposed to be the hepatic cytosolic bile acid carrier in humans.

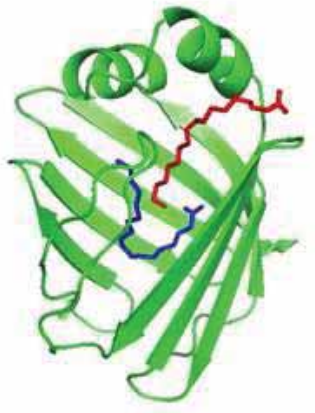


Figure 1.6 Three dimensional crystal structure of rat liver FABP in complex with two molecules of oleate (PDB 1LFO). The first molecule of fatty acid is coordinated in a bent conformation with its carboxylate shielded from free solvent (blue molecule). The second fatty acid adopts a rather linear shape with the acyl chain in the cavity interior and the carboxylate more exposed to the solvent (red molecule).

All the non-mammalian liver BABPs studied so far have shown to bind BAs with a stoichiometry of binding of two BA molecules per protein molecule. L-BABPs have been identified in birds (55-57), amphibians (58-60), reptiles (61), and fish (62-67). X-ray crystallography and NMR spectroscopy have been used to experimentally determine the three-dimensional structure of some BABPs in their apo and holo forms complexed with BAs. The best-characterized member among them is chicken BABP (cL-BABP), able to bind cholic (CA), chenodeoxycholic acid (CDA) and their glycine conjugated (68-70). The 3D structures of cL-BABP in complex with CA and CDA, resolved by X-ray and NMR spectroscopy, respectively, are reported in figure 1.7. Several NMR studies have shown that cL-BABP binds two molecules of BA with high affinity and high cooperativity. Furthermore the binding event is accompanied by a global structural rearrangement, possibly triggered by the protonation of a buried histidine, which suggests an allosteric mechanism at the basis of the cooperativity (70-72). This mechanism implies the presence of a dynamic ensemble of conformations in the apo

form. The ligand has thus the capability of selecting one existing conformation shifting the equilibrium towards the most stable holo form.

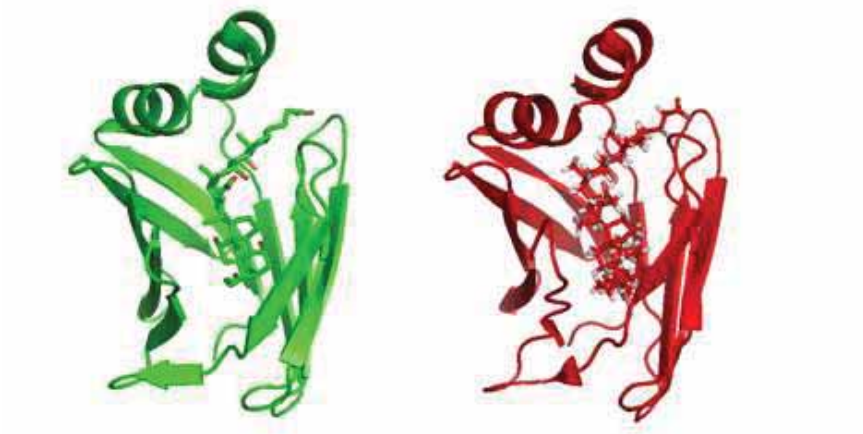


Figure 1.7 Three dimensional structure of cL-BABP in complex with two molecules of cholic acid (green, PDB code:1TW4) and chenodeoxycholic acid (red, PDB code:2JN3).

1.2.2 BABPs in the enterocytes

Intracellular transport of BAs, in enterocytes, is mediated by ileal BABPs (I-BABP). Differently from liver BABPs, these proteins have been identified in several mammalian species, and recently in the ileum of non-mammalian vertebrates (51, 73). The interaction of human I-BABP (hI-BABP) with different BAs has been extensively characterized. hI-BABP binds two molecules of GCA with low intrinsic affinity but an extraordinarily high degree of positive cooperativity. In this case the Hill coefficient, a parameter commonly used to estimate the cooperativity, is very high (n_H = 1.94), where 2 is the maximum value for an extremely positively cooperative two-site system (15). It has been reported that the binding cooperativity of hI-BABP with several physiologically BAs (CA, CDA, DCA and their glycine or taurine conjugated) is governed by the pattern of steroid rings hydroxylation and not by presence or type of side-chain conjugation (74). In addition to the strong positive cooperativity, hI-BABP has been found to exhibit a high degree of site selectivity in its interaction with GCA and GCDA. When hI-BABP is complexed with

either GCA or GCDA alone, the ligands bind to both sites. When the protein is mixed with an equimolar mixture of the two BAs, GCDA binds exclusively to one site (named site 1) and GCA to the other (named site 2). This remarkable selectivity is governed by the presence or absence of a single hydroxyl group at the C-12 position of the steroid nucleus (75). Furthermore it has been shown that cooperativity and site selectivity are not linked in the protein. Mutations that results in diminished binding cooperativity (N61A, Q99A, E110A and W49Y) do not show any substantial loss in site selectivity, and the only mutation that results in the loss of site selectivity (Q51A) exhibits no decrease in binding cooperativity (76). The characteristics of positive-binding cooperativity and site selectivity in hI-BABP report on its biological function. In the human body, approximately 55% of bile salts have three and approximately 35% of bile salts have two hydroxyl groups on their steroid ring system. Because of the relatively low intrinsic affinities of I-BABP for both GCA and GCDA, at a low bile salt concentration, a sizable fraction of bile salts remain unbound (passing through the enterocytes as monomers). When the bile salt concentrations increases, the strong cooperativity allows that the fraction of unbound glycholate decreases as the BA:protein ratio increases, thus protecting the cells from bile salt toxicity (e.g., apoptosis) On the other hand, the site selectively ensures that the majority of hI-BABP complexes *in vivo* are heterotypic: they contain both glycocholic acid (GCA) and glychochenodeoxycholic acid (GCDA) . Thus, with the combination of site selectivity and positive-binding cooperativity, nature engineered hI-BABP to be a highly efficient regulator that may function as a “regulated sponge” able to recognize structurally diverse BAs that exist in the human body (15, 76). A structural model of the heterotypic complex obtained by docking the ligands into the protein using NOE-derived inter-proton distance restraints has been described by Toke *et al*. (76) and is shown in figure 1.8.

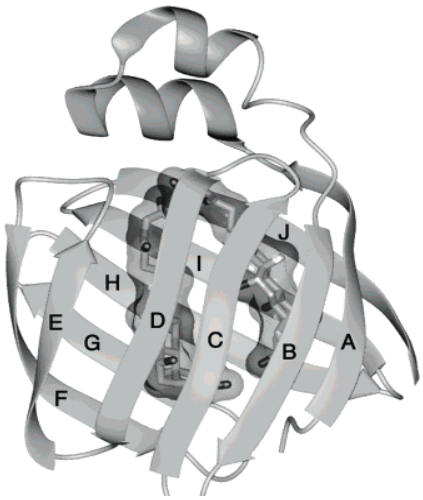


Figure 1.8 Homology model of hI-BABP with GCA and GCDA docked by restrained energy minimization.

Interestingly, the model reported for the ternary complex of hI-BABP with GCA and GCDA shows an apparent discrepancy with the previous reported NMR structures of human and porcine I-BABP in complex with BA (16, 17). The structure of human I-BABP has been solved in complex with cholytaurine (PDB code 1O1V), while the porcine protein in complex with GCA (PDB code 1EIO). In both cases the proteins bound inside the cavity a single BA molecule. The different binding stoichiometry could be explained by the temperature dependence of ligand exchange rates. Indeed both structures were determined at temperatures above 30°C, where one of the two glycocholate binding sites may have been undetected because of NMR exchange broadening (15).
Recently, calorimetry and ESI-MS studies of rabbit I-BABP (rI-BABP), highly homologous to the hI-BABP, have demonstrated that the protein is able to bind three BA molecules. Two molecules are bound inside the protein cavity, while a third binding site is hypothesised on the hydrophobic molecular surface. Besides, the observation of high affinity binding of the rI-BABP for taurine and glycine-conjugated DCA, appears to correlate with its biological function. Indeed DCA is by far the most

abundant ligand (90%) in the bile acid pool of rabbits (77). Thus, the difference in ligand binding stoichiometry may correlate with species-specific differences in ligand abundance in the diet.

An example of multiple binding has been found also in the non-mammalian zebrafish I-BABP (zI-BABP). The X-ray structere of zI-BABP has been solved in the apo and holo form, complexed with cholic acid. The structure of the holo protein shows two cholate ligands bound in the internal cavity and several cholate molecules bound on hydrophobic patches on the surface of the protein (Figure 1.9). Isothermal Titration Calorimetry (ITC) has confirmed the presence of multiple ligands (73).

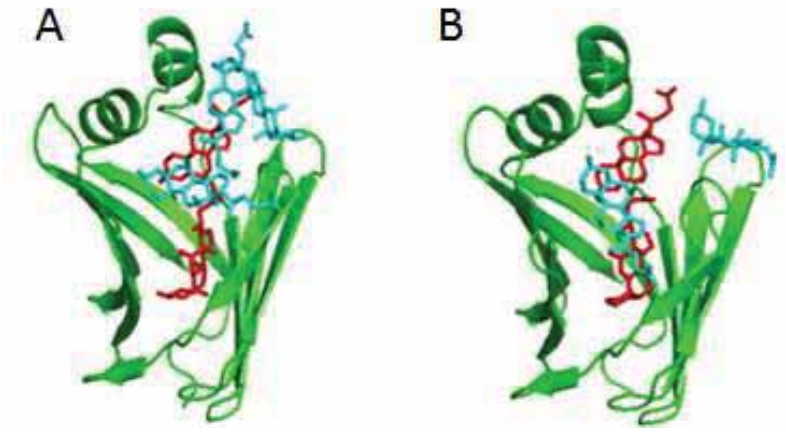


Figure 1.9 Crystal structure of zebrafish I-BABP complexed with cholate solved from two crystal forms. (a) Ribbon representation of the holo protein of crystal form A (PDB code 3ELZ). (b) Ribbon representation of the holo protein of crystal form B (PDB code 3EM0). Internal cholate molecules are coloured in red, external molecules are coloured in cyan.

1.3 Magnetic resonance imaging and gadolinium contrast agent

Clinical magnetic resonance imaging (MRI) relies on the magnetic properties of ^{1}H, as one of the most abundant naturally occurring nuclei in the human body. Spatial information about the distribution of magnetic nuclei in the body is achieved by exposing them to an inhomogeneous magnetic field that varies linearly over the body, forming a so-called *magnetic field gradient*. The magnetic field gradient causes identical

nuclei to precess at different Larmor frequencies, which are proportional to the field strength. The larger the gradient strength the larger the frequency range, and thereby the characteristically outstanding contrast between various anatomic structures in the human body is obtained by MRI.

The term contrast defines the relative difference in intensities between two adjacent regions within an examined object. The contrast of an MR image is the result of various contributing intrinsic parameters, such as the proton longitudinal relaxation times (T_1), the proton transverse relaxation times (T_2) and the proton density, and extrinsic parameters, such as the type of pulse sequence, the timing parameters of the pulse sequence and the strength of magnetic field. The diagnosis of several pathologies by MRI requires the involvement of contrast agents (CAs) that can enhance the difference between normal and diseased tissues by modifying their intrinsic parameters. The ability of a CA to influence the relaxation times of the surrounding water protons is related to its intrinsic imaging property, i.e. the relaxivity (r_1 and r_2, refer to longitudinal and transverse relaxivity, respectively). The relaxivity is defined as the relaxation enhancement of water protons produced by 1 mmol per liter of CA (expressed in s^{-1} $mmol^{-1}$ l). The observed relaxation rate, upon addition of a paramagnetic CA, is given by the sum of the diamagnetic and paramagnetic contributions (Equation 1.1). The water relaxation rates are linearly dependent on the concentration of the CA, with a slope given by the relaxivity (Equation 1.2).

$$(1/T_i)_{obs} = (1/T_i)_{diamagnetic} + (1/T_i)_{paramagnetic} \qquad i = 1,2 \quad (1.1)$$

$$(1/T_i)_{obs} = (1/T_i)_{diamagnetic} + r_i[CA] \qquad i = 1,2 \quad (1.2)$$

where:

$(1/T_i)_{obs} = (R_i)_{obs}$ = global relaxation rate

$(1/T_i)_{diamagnetic}$ = relaxation rate before addition of the CA

$(1/T_i)_{paramagnetic}$ = relaxation rate due to presence of CA

r_i = the relaxivity

[CA] = concentration of contrast agent

A higher relaxivity means more efficient proton relaxation rate enhancement, and hence, compounds with high relaxivity can be detected at lower doses, or provide greater contrast at equivalent doses. In this line, the relaxivity has to be the primary requisite in the design of new CAs. The relaxivity depends essentially on the size and chemical structure of a paramagnetic CA molecule and on the accessibility of water molecules to the paramagnetic centre (78). Among these factors, the most important parameter influencing relaxivity is the tumbling rate of the CA complex. The molecular motion of the complex can be slowed down by the formation of covalent or non covalent conjugates between the paramagnetic complex and slowly moving substrates, such as proteins (79), micelles (80), polyaminoacids (81), polysaccharides (82) or dendrimes (83). In these supramolecular systems, the tumbling rate of the paramagnetic complex decreases with a concomitant increase in the relaxivity.

The commonest and most frequently used agents for obtaining clinical imaging are low-molecular-weight gadolinium (Gd) chelates. The use of paramagnetic complex based on the presence of Gd ion is due to its high paramagnetism (seven unpaired electrons) and its favourable properties in terms of electronic relaxation (slow electronic relaxation time). Since free Gd ions are strongly toxic in the human body, they must be chelate by compounds that form very stable chelates (84). For clinical application, the Gd complex, such as all CAs, should fulfil several other requirements, such as low toxicity (high thermodynamic and kinetic stability), rapid excretion after administration, good water solubility and low osmotic potential of the solutions clinically used (85).

1.3.1 Bile acids-derived contrast agents for MRI of the liver

The contrast agents used for MRI of the liver were designed to improve the discrimination and diagnosis of focal hepatic lesion. In the past, paramagnetic extracellular CAs (e.g. Gd-DTPA or Gd-DOTA) were used to improve the differential diagnosis, but after intravenous administration

they were rapidly distributed in the interstitial space and eliminated almost exclusively through the kidney (86). Hepatospecific CAs were subsequently developed, where a lipophilic residue was conjugated to the Gd-chelate (Figure 1.10). These "lipophilic complex" are the CAs currently used in clinical imaging of the liver and include gadobenate dimeglumine (Gd-BOPTA; MultiHance, Bracco Imaging, Milan, Italy), and gadolinium ethoxybenzyl diethylenetriamine pentaacetic acid (Gd-EOB-DTPA; Primovist, Bayer-Schering Pharma, Berlin, Germany).

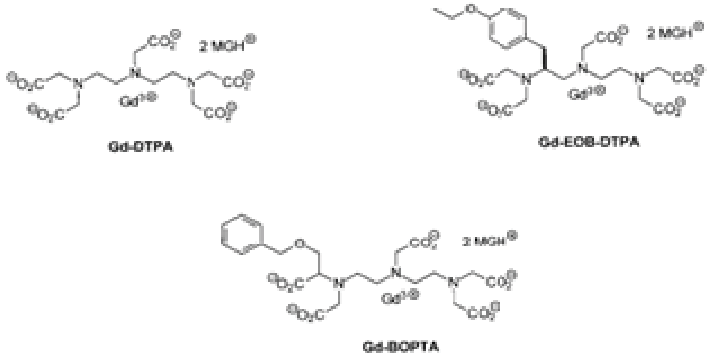


Figure 1.10 Chemical structure of different CAs. Gd-EOB-DTPA (87) and Gd-BOPTA (88) contain a Gd-DTPA complex linked to a different lipophilic residue.

The lipophilic complexes are efficiently internalized in the hepatocytes, but they are rapidly eliminated by biliary excretion, although the primary route of elimination is renal (89). Similarly to bile acids enterohepatic circulation, the excretion of Gd-BOPTA and Gd-EOB-DTPA into bile is mediated by active transport in which MRP proteins are involved (90). The active uptake of CAs through the basolateral membrane of hepatocytes is driven by OATPs. Interestingly, it was shown that OATPs are not expressed in the basolateral membrane of some hepatoma cell lines, thus making these lipophilic complexes diagnostic of specific hepatic malignancies (91). In this line, several Gd complexes conjugated to different bile acids (e.g. CA, CDA, DCA) were prepared and investigated as possible hepatospecific CAs. Among the most promising

compounds, some molecules, in which the Gd complex was linked to the position 3 of the steroid moiety of the bile acid, showed high biliary elimination as well as good tolerability (92, 93). The first of these compound (Figure 1.11) to be under clinical development has been the gadocoletic acid trisodium salt (laboratory code B22956/1).

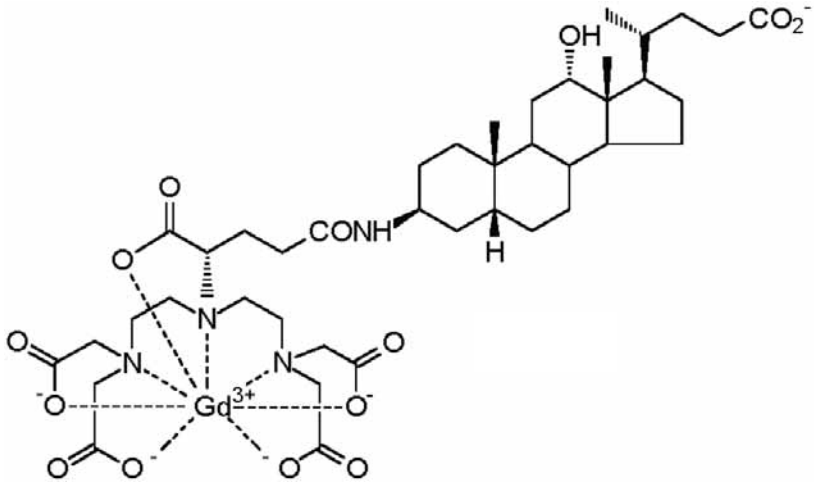


Figure 1.11 Chemical structure of B22956/1. The Gd-DTPA complex is linked to deoxycholic acid (DCA) through a flexible spacer.

1.3.2 L-BABP as possible CA-carrier

To further improve MRI performance it is crucial that CAs are developed with optimal organ specificity. This will not only result in a better diagnostic efficiency but also in a reduction of the amount of the agent administered. The organ targeting is based on the understanding of the mechanism involved in the interaction of the agent with proteins, as well as the membrane transporters or carrier proteins. Because the pharmacokinetics of each compound plays a fundamental role in its contrast enhancement effects, the understanding of the structure of a CA when bound to a macromolecule is a necessary step to set the basis for the rational design of more efficient hepatospecific CAs. While a few studies have addressed the cellular uptake and excretion of bile acid-derived CAs

in the liver, little is known about the mechanism of intracellular transport. Recently we have investigated (94) the interaction between a model BABP protein and some adducts in which Gd-DTPA was linked to different bile acids. From a preliminary analysis based on binding affinity and relaxivity, the better complex could be selected (Figure 1.12). The adduct between chicken liver BABP and Gd-DTPA-conjugated to 5β-cholanoic acid (named Gd-1) exhibited high relaxivity and the highest affinity toward the protein.

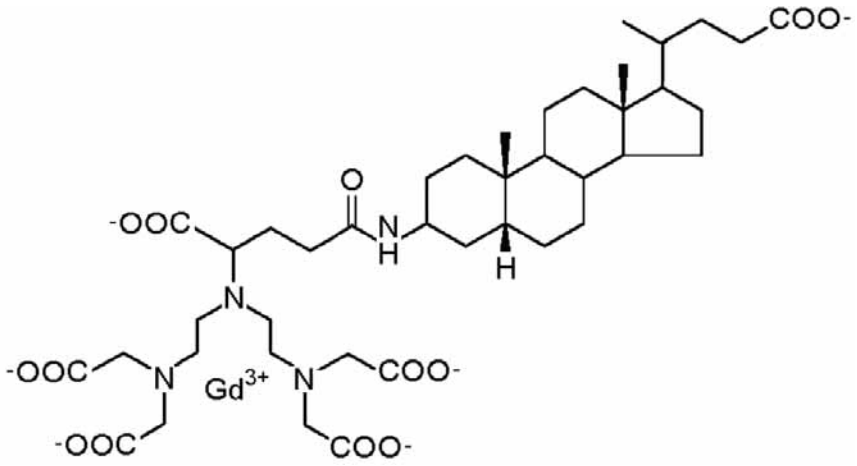


Figure 1.12 Structure of Gd-DTPA conjugated to 5β-cholanoic acid used to address the interaction with BABP.

We found, on the basis of a wide variety of NMR experiments, that Gd-1 is able to bind the intracellular carrier protein with a protein:ligand stoichiometry of 1:1 and reasonably good affinity (in the micromolar range). The portal region of the protein constitutes the binding site for the bile acid derived CA with the Gd-DTPA moiety of the ligand that remains outside the protein cavity.

Chapter 2: Structure determination of protein-ligand complexes by NMR

Protein–ligand interactions play an essential role in many biological processes such as signal transduction and cell regulation and represent important targets for engineering and design efforts. In general, the biological function of a protein depends on its interaction with ligand molecules. Accordingly, it is clear that understanding biological functions requires a precise knowledge about the structure of protein–ligand complexes. Nuclear magnetic resonance (NMR) spectroscopy for protein structure determination is now firmly established as an alternative method beside X-ray crystallography. NMR can give a detailed picture of the three-dimensional structure of a protein in solution at atomic resolution. Furthermore, NMR allows studying conformational dynamics, real time enzyme kinetics, and very week to tight protein-protein and protein-ligand interactions.

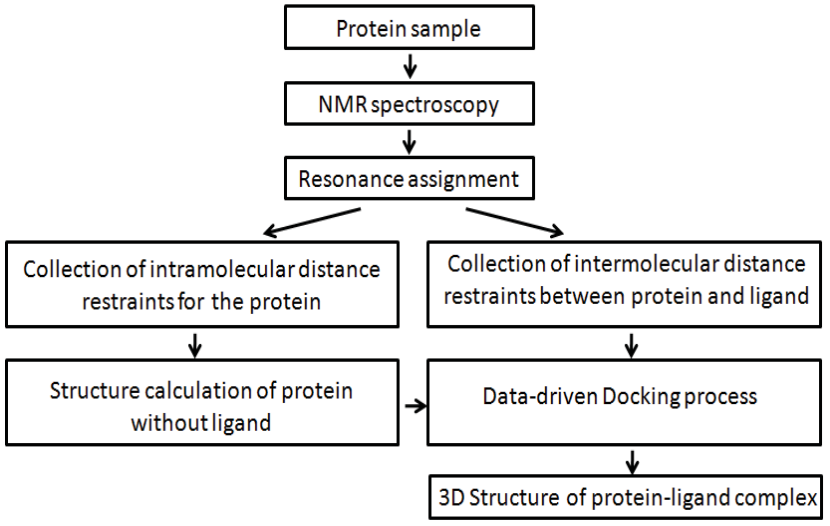
The traditional NMR approach to solving protein-ligand complexes requires the collection of intramolecular distances of the protein (nuclear Overhauser effect (NOE) distances) and a reasonable number of intermolecular NOE distances between the protein and ligands. In full structure determinations of protein–ligand complexes, intermolecular and intra-ligand NOEs can usually be treated in the same manner as NOEs within the protein. Standard *ab initio* simulated annealing based on Cartesian coordinate or torsion-angle dynamics (e.g. CNS (95)) usually work satisfactorily. On the other hand, given the intrinsically poor chemical shift dispersion of the methine and methylene proton resonances of some small aliphatic ligands (such as bile acids, studied in this thesis), it is often difficult or impossible to derive many intermolecular restraints or to obtain proton ligand assignments directly from the experimental data. A strategy is using ambiguous intermolecular restraints derived from other NMR experiments that can supplement the initial restraints list of intermolecular restraints in order to obtain a preliminary protein-ligand structure by guided docking approach. At this point, starting from the preliminary structure model, it is possible to identify further protein-ligand interactions. The latters can help in the interpretation of unassigned

protein-ligand NOEs and enrich the restraints list. Then, using an interactive procedure, the structural models calculated by docking approach can lead to the identification of a reasonable number of intermolecular restraints and subsequently to the structural determination of the complex with high accuracy.

In the last decade, several computational approaches have been used in order to help in predicting protein-protein and protein-ligand interactions and to solve protein–ligand structures. Among the current docking methods, HADDOCK (High Ambiguity-Driven DOCKing) distinguishes itself from others by the use of experimental and/or bioinformatics data to drive the docking process. In particular HADDOCK allows obtaining high resolution structures of the protein-ligand complex starting from the known 3D structure of the protein and incorporating information on distance constraints between the protein and the ligand obtained experimentally by NMR. Thus ambiguous intermolecular restraints, such as chemical shift perturbation (CSP), ^{15}N relaxation measurement data, and unambiguous intermolecular restraints such as intermolecular NOEs (96, 97) can be used to drive the docking. Furthermore, if the structure of the protein undergoes structural rearrangement upon formation of a complex with a partner, significant improvement of the docking approach is achieved if the structure of the holo protein (without ligand) is used as starting structure instead of the one corresponding to the apo protein or modeled structure (98).

In this line, the structure determination of protein complexes by NMR in combination with docking approaches, involves several steps. These are summarized in Scheme 2.1.

This chapter describes the methods, the NMR experiments and the software used in my research project for determining the three-dimensional structures of a protein-ligand complex following the procedures outlined in the scheme.



Scheme 2.1 Flow chart of the procedure for deriving 3D protein-ligand structures using experimental NMR data in conjunction with docking approach.

2.1 Protein Sample preparation

2.1.1 Expression and purification of BABPs

Structural studies require high-level expression systems and efficient purification protocols. *Escherichia Coli* is the one of the most successful host systems for high-level production of both prokaryotic and eukaryotic proteins. Moreover, expression in E.coli is often the most favoured method because of the ease of expression and high label incorporation into the recombinant protein.

In my thesis work I performed the expression and purification of two bile acid binding proteins, the chicken liver BABP and the chicken ileal BABP and some mutants of the latter. For all the proteins, wild type and mutants, the protocols used are the same, with slight differences.

For unlabelled proteins, a 10 ml starter culture of the best expression colony was grown in LB medium shacking overnight at 37°C. The culture was used to inoculate 1L of LB, then incubated in a shaker at 37°C till an OD_{600} value of 0.6 was reached. Protein expression was induced by the

addition of IPTG (Isopropyl β-D-thiogalactoside). The cells were allowed to grow overnight at the best temperature. After harvesting, the cells were lysed by sonication and the supernatant, containing the soluble protein fraction, was loaded on the first purification column. The ^{15}N or ^{15}N-^{13}C labelled proteins were expressed in M9 minimal media, containing $^{15}NH_4Cl$ for uniformly ^{15}N labelled samples or $^{15}NH_4Cl$/^{13}C-glucose for uniformly ^{15}N/^{13}C labelled samples using the same expression protocols of unlabelled proteins.
The purification started with an anion exchange chromatography followed by a size exclusion chromatography with a final delipidation step. The isolated proteins were transferred, by dialysis, in a buffer solution optimal for NMR measurements and finally processed by liophylization and stored at 4°C.
Protein concentrations were estimated from the extinction coefficients at 280 nm as calculated using the ProtParam tool (http://expasy.ch/protparam) of the Swiss Institute of Bioinformatics.

2.1.2 Sample quality control

In the case of labelled protein samples, mass spectrometry analysis was performed to estimate the ^{15}N and/or ^{13}C isotope inclusion percentage. In addition every unlabelled protein was checked by ^{1}H mono-dimensional NMR experiment. For single or double-labelled proteins basic two-dimensional heteronuclear NMR experiments were performed to confirm the isotope inclusion and to evaluate the quality of the protein preparation.

2.1.3 Sample preparation for NMR analysis

For NMR analysis liophylized samples were prepared by suspending the freeze-dried protein in 90% H2O/10% D2O or 100% D2O. NMR experiments were performed on 0.4-0.8 mM samples depending on experimental requirements. The NMR experiments were recorded on a

Bruker Avance III 600 spectrometer, operating at 600.13 MHz proton Larmor frequency, equipped with a triple resonance TCI cryoprobe, incorporating gradients in the z -axis. The experimental temperature was set to 298 K, unless otherwise specified.

2.2 NMR Spectroscopy

2.2.1 The NMR phenomenon

NMR methods for the study of protein structures depend on the variation of the resonance frequency of a nuclear spin, in an external magnetic field, with the different chemical structure, conformation, and solvent environment. Three of the four most abundant elements in proteins, H, C, and N, have naturally occurring isotopes with nuclear spin 1/2 which are suitable for high-resolution NMR experiments in solution. The proton ^{1}H has the highest natural abundance (99.98%) whereas both ^{13}C and ^{15}N have low natural abundance (1.11% and 0.37% respectively). The lower abundance of ^{13}C and ^{15}N is routinely overcome by over expression of proteins in isotope-labelled media.

The nuclear spin gives rise to a magnetic dipole moment:

$$\mu = \gamma I \qquad (2.1)$$

where μ is magnetic dipole moment, γ is gyromagnetic ratio of the nucleus and I is spin angular momentum. If such an atomic nucleus is placed into external magnetic field B0, it has energy associated with the interaction between its magnetic moment and the external field:

$$E = - \mu B_0 = - \gamma I B_0 \qquad (2.2)$$

In the absence of a magnetic field a collection of nuclear spins can adopt all possible spatial orientations. When placed in a magnetic field magnetic moments of spin 1/2 nuclei assume two orientations (Figure 2.1), either in the same direction or in the opposed direction with respect to the magnetic field. At thermal equilibrium, the population distribution for a ½ spin nucleus, can be calculated by using the Boltzmann distribution

relationship:

$$\frac{N_{upper}}{N_{lower}} = e^{-\frac{\Delta E}{KT}} \qquad (2.3)$$

where N is the population in each state, T is the absolute temperature in Kelvin and k is the Boltzmann constant.

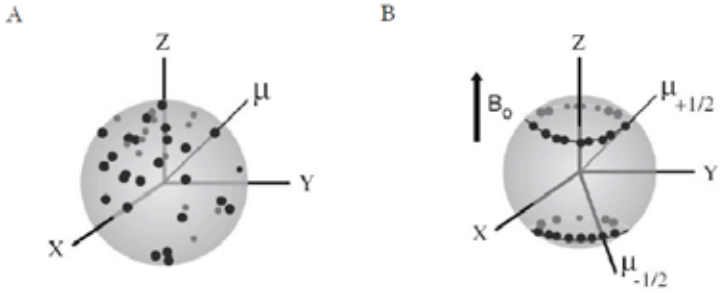


Figure 2.1 Orientation of nuclear magnetic dipoles in the absence of magnetic field (A) and in presence of a magnetic field Bo along the z-axis (B) (99).

Since the lower level is slightly more populated than the upper level, at the thermal equilibrium the sample has a net magnetization M_0 (Figure 2.2) in the direction of the magnetic field B_0 (along the z-axis). Because the magnetization aligned along the z-axis is undetectable, in order to induce nuclear magnetic resonance, the equilibrium condition of M_0 has to be perturbed. This is achieved by applying an additional magnetic field (B_1) along the x axis, which rotates the magnetization into the x,y plane perpendicular to B_0.

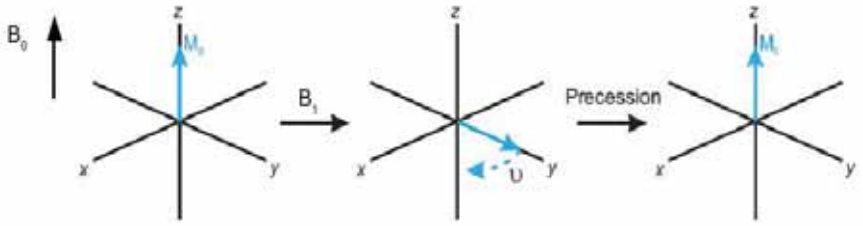


Figure 2.2 The equilibrium condition of the magnetization aligned along the z-axis (M_0) is perturbed by B_1 applied along the x axis. The magnetization rotated into the xy plane precesses at the Larmor frequency. In the NMR experiment the precession of the magnetization vector in the xy plane is detected by a coil.

The transverse magnetization in the xy plane precesses under the influence of the static magnetic field B_0 at the Larmor frequency (υ_0 or resonance frequency). The Larmor frequency is described by the following relationship:

$$\nu_0 = -\frac{\gamma B_0}{2\pi} \qquad \text{or} \qquad \omega_0 = \gamma B_0 \qquad (2.4)$$

where ν_0 refers to the frequency expressed in Hertz (Hz), while ω_0 refers to the frequency in rad s^{-1}.

The precession of the magnetization vector in the xy plane induces an electric current in a detection coil, which is the primary observation in a NMR experiment.

Since the system tends to return to its equilibrium condition with the magnetization orientated parallel to B_0, the acquired signal decreases during time: this is the so-called free induction decay (FID). The NMR spectrum is then obtained by Fourier transformation (FT) of this data (Figure 2.3).

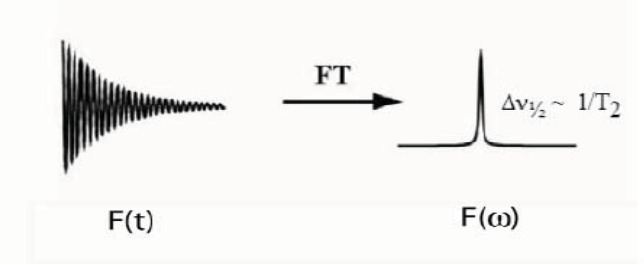


Figure 2.3 The Fourier transformation of the NMR signal. NMR signal is free-induction decay (FID) which is a function of time, F(t). The Fourier transformation is used to translate the function from the time domain into the frequency domain, F(ω).

In a NMR spectrum, each resonance (frequency line) represents one or more nuclei. As stated before, the NMR frequency of a nucleus in a molecule is determined by its gyromagnetic ratio γ and by the strength of the applied magnetic field B_0 (Equation 2.4). However not all protons, nor

all ^{13}C nuclei, have identical resonance frequencies since ν depends on the chemical environment of nucleus. This effect is called chemical shift (δ) and is generally defined as the difference between the resonance frequency of the nucleus of interest (ν) and that of a reference nucleus (ν_{ref}):

$$\delta = \frac{\nu - \nu_{ref}}{\nu_{ref}} \qquad (2.5)$$

The frequency difference ν-ν_{ref} is divided by ν_{ref} so that δ is a molecular property, independent of the magnetic field used to measure it.
The chemical shift is a fundamental parameter in protein NMR as it gives separately detectable signals for the hundreds of protons (or other NMR nuclei) that can therefore be distinguished and assigned. This allows the identification of signals from individual atoms even in complex biological macromolecules. Thus from the resulting spectra it is possible obtain information on molecular structure, conformation and dynamics.

2.2.2 1D NMR experiments

The typical scheme of a 1D experiments is shown in figure 2.4.

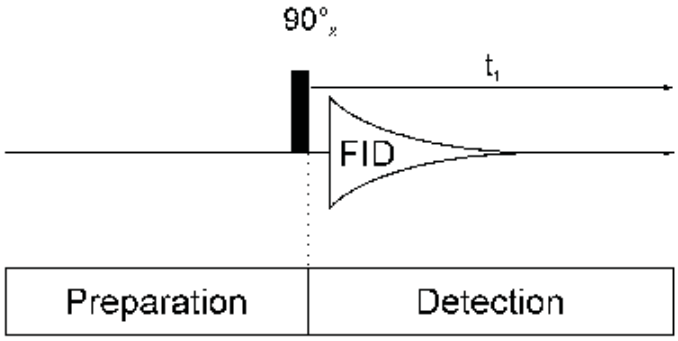


Figure 2.4 Scheme of a 1D NMR experiment

Each 1D NMR experiment consists of two sections: preparation and detection. Preparation is the time needed before the system reaches

equilibrium when immersed in the static magnetic field B_0. After the equilibrium magnetization (M_0) has been established, a radiofrequency pulse is applied. The 90° pulse (applied along the x axis) rotates the magnetization onto the perpendicular axes. After this pulse each spin precesses with its own Larmor frequency around the z axis and induces a signal in the receiver coil. During detection the resulting signal is recorded. A receiver, located in a direction perpendicular to the vector B_0 direction, acquires the signal with an intensity that is proportional to the magnetization in the xy plane. The behaviour of the magnetization decay (FID) is detected during the time T_2, which is the necessary time so that the transverse magnetization reaches the equilibrium state.

The 1D NMR spectra of biopolymers are dramatically crowded and complex, as it can be seen in figure 2.5, where the 1D ^{1}H-NMR spectrum of a 14000 Da protein is reported. The large overlap in the proton resonances exclude any chance for a detailed structural study. To overcome the natural limitation of the 1D NMR spectroscopy, multidimensional 2D, 3D and 4D NMR experiments have been developed in the last twenty years. A brief description of the 2D and 3D NMR spectroscopy is reported below.

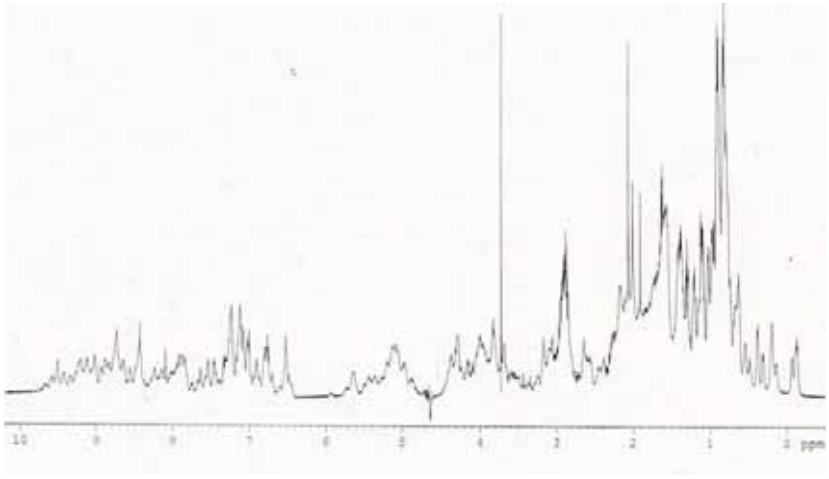


Figure 2.5 1D ^{1}H-NMR spectrum of chicken L-BABP recorded at 600 MHz..

2.2.3 2D NMR experiments

A 2D NMR general scheme experiment is shown in figure 2.6. It includes four successive time periods: preparation, evolution, mixing and detection. As well as in the 1D NMR experiment, in the preparation step the spin system reaches the equilibrium state with the external magnetic field B_0. The preparation ends when one or more pulses are applied and the system is led to a non-equilibrium state. The evolution time t_1, is the period where the system spontaneously evolves according to the individual precessional frequencies. The evolution time is an incrementable delay and a characteristic feature of all 2D pulse sequences. Data acquisition includes a large number of spectra that are acquired as follows: the first time the value of t_1 is set close to zero and the first spectrum is acquired. The second time, t_1 is increased of Δt and another spectrum is acquired. This process is repeated until there is enough data for analysis using a 2D Fourier transform.

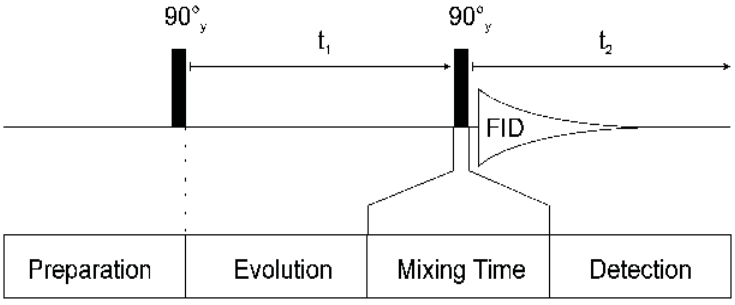


Figure 2.6 Scheme of a 2D NMR experiment

During the detection period a FID is recorded for each value of t_1. In this context the real time variable that defines the FID itself is given by the symbol t_2. The pulse sequences are designed so that the signals detected in t_2 are modulated in amplitude or phase as a function of the delay t_1. The different correlations between the spins are realised during the mixing time. Therefore the frequencies of the signals present during t_1 can be measured indirectly through their effect on the real-time signals detected

directly in t_2. Once the complete data set has been acquired, a Fourier transform with respect to t_1 and t_2 is used to produce a two-dimensional spectrum that is function of two frequencies variables F_1 and F_2. The various 2D experiments differ in the type of signals that is present during t_1 and in the interaction employed during the mixing time to transmit information from t_1 to t_2.

The majority of homonuclear experiments, in which are observable the correlations of the same isotope (usually ^1H) have a common structure: the diagonal peaks correspond to those of the conventional one dimensional spectrum, while the off-diagonal peaks, or cross peaks, contain information about the connections between resonances on the diagonal. The nature of these connections depends on the kind of two dimensional experiment being carried out. The experiment can be set up to observe the scalar correlations through bonds (J-coupled) between the resonances of the nuclei separated by two or three bonds, (e.g. COSY and TOCSY), or dipolar connections (through-space), between the resonances of nuclei which are close together in space (e.g. NOESY). Beside the 2D homonuclear experiments, several experiments involving magnetization transfer between heteronuclei has been developed. One of the most popular is the Heteronuclear Single Quantum Correlation (HSQC) experiment which has a wide diffusion in protein NMR. In the HSQC experiment the magnetization is transferred from the proton nuclei to the heteronuclei (X= ^{15}N or ^{13}C) and then back to ^1H for detection. This step is very important to enhance the sensitivity of the experiment since ^{15}N and ^{13}C heteronuclei have 300 and 32 fold less sensitivity, respectively, than protons. The HSQC experiment (Figure 2.7) starts with an INEPT (Insensitive Nuclei Enhanced by Polarization Transfer) sequence that transfers the magnetization from the proton to less sensitive spin directly attached via one bond ($^1J_{HX}$) scalar couplings.

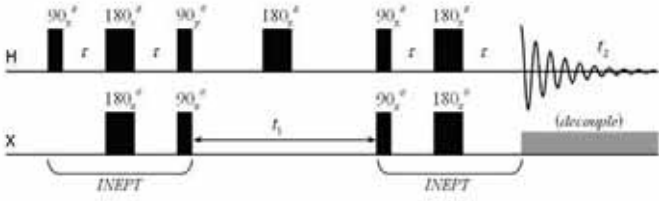


Figure 2.7 Pulse sequence diagram for HSQC experiment.

This first step is followed by the evolution period during which the magnetization evolves under the heteronucleus chemical shift. In the middle of the evolution period a 180° pulse is applied at the proton frequency that refocus the evolution due to J-coupling. A second INEPT step, also called reverse INEPT, transfers back to the proton the magnetization which carries now the modulation due the heteronucleus chemical shift. At the end of this second INEPT the signal is acquired and recorded. During the acquisition time a decoupling pulse sequence is applied to the heteronucleus that remove the coupling and allows to obtain a single signal for every J-coupled pair of spins. The global appearance of an HSQC spectrum is a series of dot signals characterized by two frequency coordinates, i.e., proton and heteronucleus. In figure 2.8 is reported an HSQC spectrum acquired on a ^{15}N labelled protein.

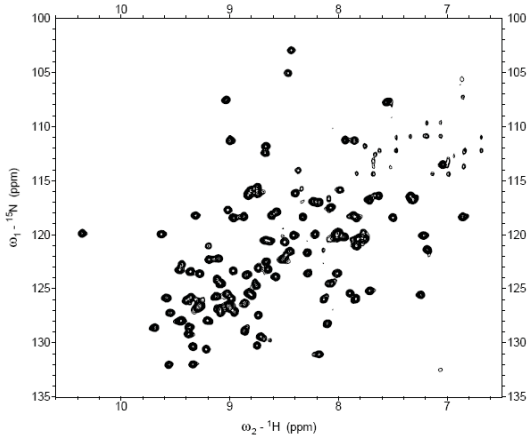


Figure 2.8 Example of an ^{1}H-^{15}N HSQC spectrum.

2.2.4 3D NMR experiments

A general scheme for a 3D NMR experiment is reported in figure 2.9.

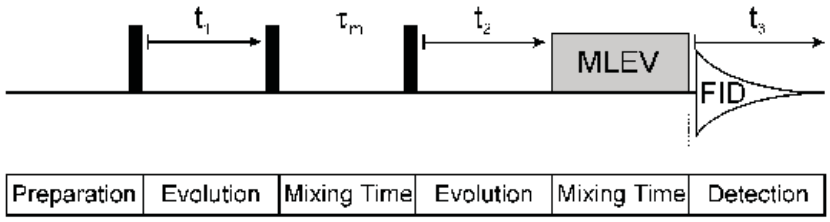


Figure 2.9 Scheme of a 3D NMR experiment

A three dimensional NMR experiment can easily set up based on a two dimensional scheme inserting an additional indirect evolution time and second mixing period before the direct data acquisition. Therefore, in a 3D experiment the signal, which is monitored during the acquisition time t_3, is function of two evolution times, t_1 and t_2, which are incremented independently. The signal is a matrix with three axes corresponding to t_1, t_2 and t_3. After Fourier transformation, the resulting 3D spectrum is a cube with three different frequency dimensions F_1, F_2, F_3. There are two principal classes of 3D experiments:

- 3D double-resonance experiments, which consist of two 2D experiments, one after another (e.g. 3D NOESY-HSQC, HCCH-TOCSY);
- 3D triple-resonance experiments, which consist of a 2D heteronuclear experiments which is then extended into a different third dimension (e.g. 3D CBCANH, HNCO).

2.3 Resonance assignment

In order to define the three dimensional structure of a protein or determine how it interacts with ligands, it is necessary to assign the NMR resonances. The goal of resonance assignment is to uniquely associate each nuclear spin of the sample molecule with a chemical shift observed in the MNR spectrum.

The conventional assignment strategy for unlabelled proteins (100) makes use of a combination of COSY/TOCSY and NOESY spectra. The first two experiments, based on the scalar coupling, are used to assign intra residue correlations between protons of same residue. A NOESY experiment is used to establish sequential correlations between protons of adjacent residues. However in 2D spectra of proteins with more than about 80 amino acids resonance overlap becomes so severe, that it is often impossible to identify all sequential correlations. The introduction of three-dimensional NMR experiments and the availability of $^{15}N^{13}C$ labelled proteins allow assigning the proton, nitrogen and carbon chemical shifts and to determine their structures in solution. The heteronuclear triple-resonance experiments (101) correlate backbone ($^{1}H_N$, ^{15}N, $^{1}H_\alpha$, $^{13}C_\alpha$ and ^{13}CO) and side chain spins (^{1}H and ^{13}C) by an efficient transfer of magnetization. The magnetization is transferred via one bond (^{1}J) or two bond (^{2}J) scalar couplings. Also, since the ^{1}J- and ^{2}J-couplings are generally larger than line width of nuclei under consideration, the coherence transfer via these couplings remain efficient for relatively large molecules with short transverse relaxation times and can compete with loss of magnetisation as a result of relaxation during the pulse sequence. Moreover ^{1}J- and ^{2}J-couplings are largely independent of conformation.

2.3.1 3D NMR Experiments for the assignment of backbone resonances

The first step in a structure determination by NMR spectroscopy of a doubly labelled protein is the assignment of the backbone resonances. In table 2.1 are summarized the triple resonance experiments used for backbone and side-chain resonance assignment performed in this thesis, with the corresponding observed correlations and the scalar couplings utilized for coherence transfer. The nomenclature for these triple-resonance experiments reflect the magnetization transfer pathway of the experiment. Nuclei that are involved in the magnetization transfers define the name of the experiment while the names of nuclei which are used only

for transfer and whose frequencies are not detected are bracketed.

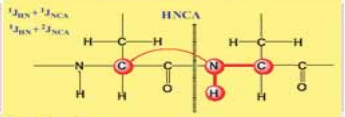
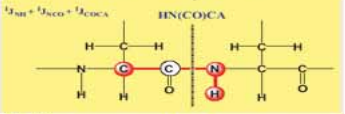
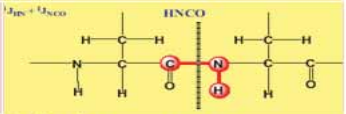
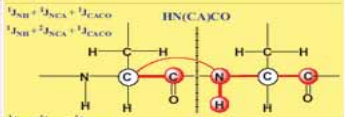
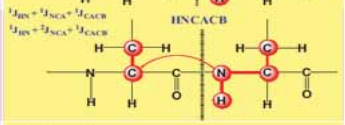
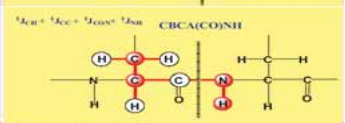
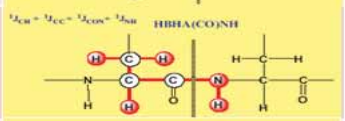
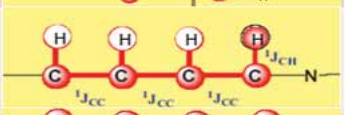
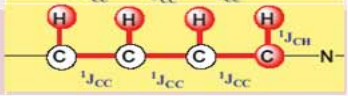
EXPERIMENT	CORRELATIONS OBSERVED	MAGNETIZATION TRANSFER
HNCA	$H_iN_iCa_i$ $H_iN_iCa_{i-1}$	
HN(CO)CA	$H_iN_iCa_{i-1}$	
HNCO	$H_iN_iCO_{i-1}$	
HN(CA)CO	$H_iN_iCO_i$ $H_iN_iCO_{i-1}$	
HNCACB	$H_iN_iCa_i$ $H_iN_iCa_{i-1}$ $H_iN_iCb_i$ $H_iN_iCb_{i-1}$	
CBCA(CO)NH	$H_iN_iCa_{i-1}$ $H_iN_iCb_{i-1}$	
HBHA(CO)NH	$H_iN_iHa_{i-1}$ e $H_iN_iHb_{i-1}$	
(H)CCH-TOCSY	C_iH_i – all C_i	
H(C)CH-TOCSY	C_iH_i – all H_i	

Table 2.1 Triple resonance experiments used in this work for resonance assignment.

The backbone assignment is achieved by combining pairs of heteronuclear NMR experiments, one of which provides intra-residual correlations between backbone resonance frequencies and the other one

inter-residual correlations of the same set of frequencies.

The most important pair consists of the HNCA (101-103) and the HN(CO)CA (102, 104) experiment. In the HNCA experiment, the coherence is transferred from the amide proton ($^{1}H_N$) to the directly attached nitrogen atom (^{15}N), which is measured as the first spectral dimension. Then the magnetization is transferred to the ^{13}C alpha nucleus ($^{13}C\alpha$) which is measured as second dimension. Afterwards, the magnetization is transferred back the same way to the $^{1}H_N$, which is measured as the third (direct) dimension. Since the coupling constant $^{15}N(i)$-$^{13}C\alpha(i)$ is very similar to the one between $^{15}N(i)$-$^{13}C\alpha(i-1)$, two correlations will be obtained for the amide proton, one with its own Cα ($^{13}C\alpha(i)$) and other one (less intense if the delay has been fitted to the first coupling) with the Cα of the previous residue ($^{13}C\alpha(i-1)$). The resulting NMR spectrum consists of two peaks with same $^{1}H_N$, ^{15}N frequencies (of residue i) but different $^{13}C\alpha$ frequency, one of residue i and the other one of preceding residue (i-1).

The HN(CO)CA experiment correlates the resonances of $^{1}H_N$ and ^{15}N of the residue i only with $^{13}C\alpha$ of the preceding residue. In the HN(CO)CA experiment the coherence transfer pathway starts from $^{1}H_N$, it is passed to ^{15}N and then to the carbonyl spin (^{13}CO), which suppresses the intra-residual pathway and keeps only the inter-residual one. In this experiment the frequency of the ^{13}CO spin itself is not recorded but is only used to transfer the coherence to the $^{13}C\alpha(i-1)$. After chemical shift evolution of the $^{13}C\alpha$, the magnetization is transferred back via ^{13}CO to ^{15}N and $^{1}H_N$ for detection. The HN(CO)CA spectrum shows, instead of two peaks, only one with the ^{1}H, ^{15}N frequencies of amino acid i and the $^{13}C\alpha$ frequency of the preceding residue. The combination of HNCA and HN(CO)CA experiments, therefore, allows one to distinguish between the intra-residual and the inter-residual $^{13}C\alpha$ chemical shifts. A schematic representation of HNCA/HN(CO)CA spectra and the observable correlations is reported in figure 2.10.

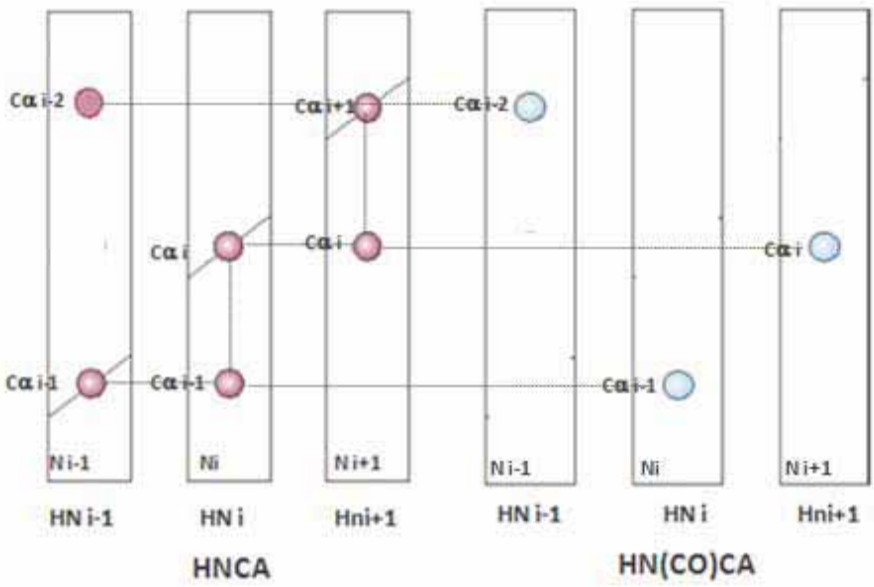


Figure 2.10 Series of strips from HNCA (left) and HN(CO)CA (right). The x and y-axes represent the $^{1}H_{N}$ and $^{13}C\alpha$ frequencies respectively. The ^{15}N chemical shift for each plane (third dimension) is noted at the bottom of the strips. In the HNCA spectrum are shown the correlations observed between the amide proton and the $^{13}C\alpha$ of the same residue and the preceding amino acid $^{13}C\alpha$(i-1). In the HN(CO)CA are observable only the correlations of $^{1}H_{N}$ with the alpha carbons of preceding residue $^{13}C\alpha$(i-1).

A second pair of experiments that give additional information to complete backbone assignment consists of HNCO/HN(CA)CO experiments. The HNCO (101, 102, 105) experiment correlates the chemical shifts of $^{1}H_{N}$ and ^{15}N with the resonance of the ^{13}CO of the same residue. Magnetization is passed from ^{1}H to ^{15}N and then selectively to the carbonyl ^{13}C via $^{1}J_{NCO}$ coupling. Magnetization is then passed back along the reversed path for detection. The HNCO yields exclusively a single peak with ^{1}H ^{15}N frequencies of residue i and the frequency of ^{13}CO of i-1 residue. Besides the HN(CA)CO (106) provides intra-residue correlations between the $^{1}H_{N,}$ ^{15}N and ^{13}CO chemical shift using the one-bond ^{15}N-$^{13}C\alpha$ and $^{13}C\alpha$-^{13}CO J-couplings ($^{1}J_{NC\alpha}$ and $^{1}J_{C\alpha CO}$ respectively) to transfer the coherence. In addition, this experiment can also provide

sequential connectivities from the ^{15}N resonance to the ^{13}CO of the preceding residue via inter-residue two bond ^{15}N-$^{13}C\alpha$ J-coupling ($^{2}J_{NC\alpha}$). Thus for each NH group, two carbonyl groups are observed in the spectrum. Because the $^{1}J_{NC\alpha}$ is stronger than $^{2}J_{NC\alpha}$, the peak corresponding to ^{13}CO of residue i is more intense than the peak with the ^{13}CO frequency of the preceding residue. A disadvantage of the HNCA/HN(CO)CA and HNCO/HN(CA)CO pairs is the small chemical shift dispersion of the $^{13}C\alpha$ and the ^{13}CO spins, which can cause significant resonance overlap, leading to ambiguous assignments. To obtain additional information, the $^{13}C\beta$ spin can be used. A combination of HNCACB and CBCA(CO)NH experiments, allows to correlate the $^{13}C\beta$ chemical shift with the backbone resonances. In addition, the $^{13}C\alpha$ and $^{13}C\beta$ chemical shifts provide information on the type of amino acid that can also be employed to position sequentially connected fragments within the amino acid sequence. The HNCACB experiment transfers magnetization from the amide proton via the ^{15}N to the $^{13}C\alpha$ and further to the $^{13}C\beta$ spins (107). The resulting spectrum contains the correlations between $^{1}H_N$ and ^{15}N of amino acid i and the $\alpha-$ and $\beta-$carbons of amino acids i and i–1 (Figure 2.11).

The CBCA(CO)NH experiment (108) correlates both the $^{13}C\alpha$ and $^{13}C\beta$ resonances of an residue with the $^{1}H_N$, ^{15}N chemical shifts of the following residue. As well as the HN(CO)CA experiment, inter-residue correlations are established by transferring coherence via the intervening ^{13}CO spins. The CBCA(CO)NH spectrum, therefore, contains the same information of HN(CO)CA but with the addition of the $^{13}C\beta$ chemical shifts. In figure 2.11 are shown the correlations observed in the HNCACB/CBCA(CO)NH spectra.

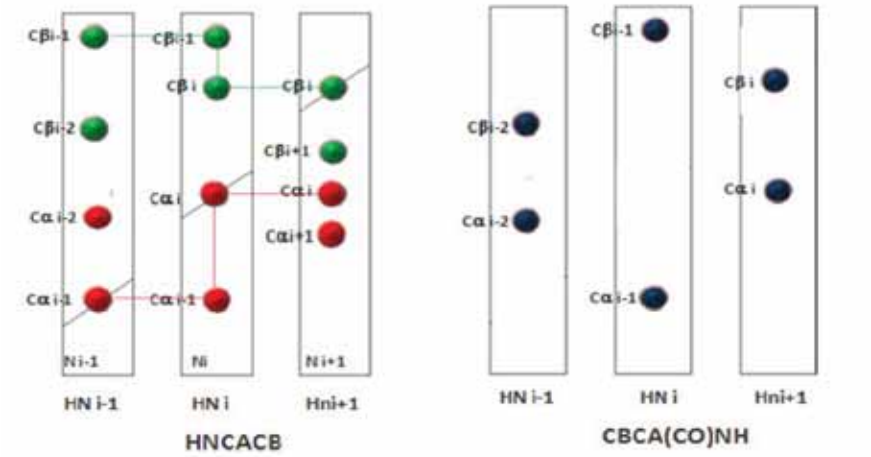


Figure 2.11 The combination of HNCACB (left spectrum) and CBCA(CO)NH (right spectrum) allows to distinguish sequential (i-1) and intra-residual (i) peaks. Moreover in the HNCACB it is possible to distinguish between Cα and Cβ because of their positive (red) or negative (green) signals.

2.3.2 3D NMR Experiments for the assignment of side-chain resonances

The NMR experiments used for the side chain resonances assignment are experiments that correlate the side chain resonances with the already known resonances of backbone nuclei (^{15}N, ^{13}Cα and ^{13}Cβ).

The first step in the side chain assignment is the identification of alpha and beta proton resonances (^{1}Hα and ^{1}Hβ respectively). An experiment that provides connectivity to ^{1}Hα and ^{1}Hβ is the HBHA(CO)NH (109), which correlate the ^{1}Hα and ^{1}Hβ resonances of amino acid i–1 with the backbone amide nitrogen and amide proton of amino acid i. The magnetization is transferred from ^{1}Hα to ^{13}Cα and from ^{1}Hβ to ^{13}Cβ, respectively, and then from ^{13}Cβ to ^{13}Cα. From here it is transferred first to ^{13}CO, then to ^{15}N of following residue and finally to $^{1}H_N$ for detection. In the resulting three dimensional spectrum it is possible to identify, per each amide group (with known backbone assignment), the resonances of ^{1}Hα and ^{1}Hβ of the preceding amino acid.

From knowledge of ^{13}Cα, ^{13}Cβ, ^{1}Hα and ^{1}Hβ resonances it is then

possible to complete the side chain assignment. The full assignment is achieved by combining two 3D double-resonance experiments, the (H)CCH-TOCSY and H(C)CH-TOCSY (110, 111). These experiments are used to assign aliphatic ^{1}H,^{13}C spin systems within residues, and to link them to the sequentially assigned backbone resonances. The (H)CCH-TOCSY spectrum correlates an aliphatic ^{13}C-^{1}H pair with all ^{13}C nuclei of the same side chain. In the spectrum two of the three dimensions correspond to ^{13}C and the other one to ^{1}H. Starting from a spin system with $^{13}C\alpha$ and $^{1}H\alpha$ (or $^{13}C\beta$ and $^{1}H\beta$) resonances assigned, it is feasible to identify the chemical shifts of all aliphatic ^{13}C belonging to that side chain. The related H(C)CH-TOCSY experiment correlates an aliphatic ^{13}C-^{1}H pair with all aliphatic ^{1}H nuclei of the same side chain. The H(C)CH-TOCSY spectrum, in which two of the three dimensions correspond to protons and the third one to carbon, is used in the assignment of all aliphatic ^{1}H nuclei. In figure 2.12 is reported a schematic example of correlations observable in the (H)CCH-TOCSY (A) and H(C)CH-TOCSY (B) spectra.

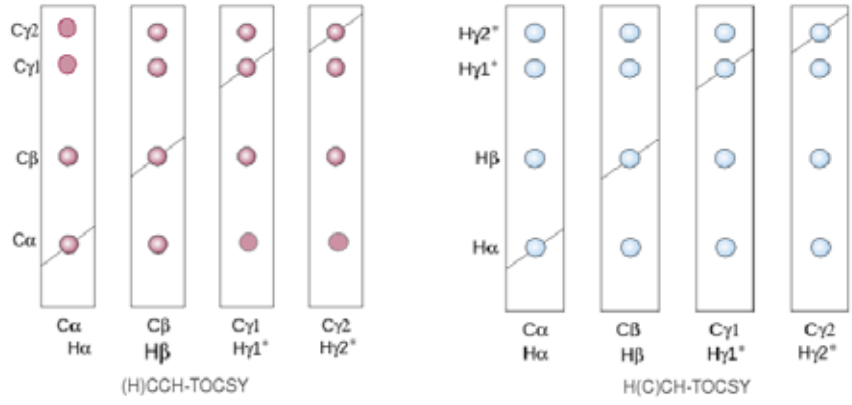


Figure 2.12 The combination of (H)CCH-TOCSY (left spectrum) and H(C)CH-TOCSY (right spectrum) allows to identify all side-chain resonances. In figure is shown an example of strips observable for a valine residue. From the knowledge of $^{13}C\alpha$, $^{13}C\beta$, $^{1}H\alpha$ and $^{1}H\beta$ chemical shifts, it is possible to assign the resonances of $^{13}C\gamma_1$ and $^{13}C\gamma_2$ in the (H)CCH-TOCSY and the resonances of methyl protons $^{1}H\gamma_1$ and $^{1}H\gamma_2$ in the H(C)CH-TOCSY.

2.3.3 Structural information obtained by the assignment of resonances

Once the sequential assignment is complete, the secondary structure of the protein can be determined from the secondary-structure specific chemical shifts. A commonly utilized method is the CSI protocol (112). The CSI method relies on the Cα, Cβ, CO and ^{1}Hα shifts. The protocol uses a two-step digital filtering process. In the first step, a chemical-shift index (CSI) of -1 or 0 or +1 is assigned to each residue based on the chemical shift of a particular nucleus relative to an appropriate random-coil value. In the second step, secondary structural elements are identified by a clustering of identical chemical-shift indices. With this CSI protocol, helices are indicated by a cluster of chemical-shift indices having value +1 for Cα and CO, and -1 for ^{1}Hα. In contrast, β-strands are indicated by a cluster of chemical-shift indices having value -1 for Cα and CO, and +1 for ^{1}Hα and Cβ. An example of CSI output is shown in figure 2.13.

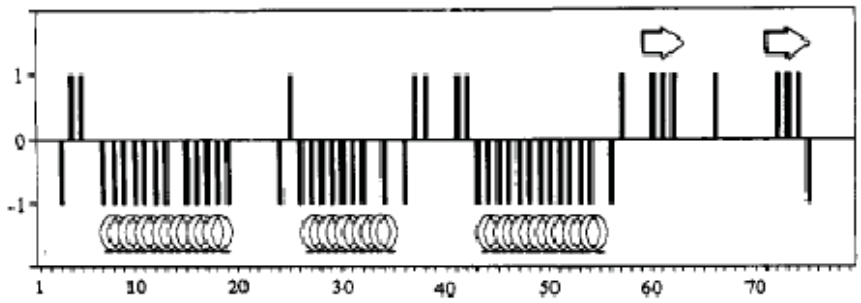


Figure 2.13 Histogram representation of the chemical-shift index obtained for ^{1}Hα proton of an α/β protein. Arrows represent β-strands and coils represent α-helices.

The experimental chemical shift of ^{13}Cα, ^{13}Cβ, ^{13}CO, ^{15}N, $^{1}H_N$ and ^{1}Hα can be also used for backbone dihedral angles determination by TALOS+ (113). TALOS+ compares the similarity in amino acid sequence (residue triplets) and chemical shifts of the protein under study with proteins of known structure in its database in order to make quantitative predictions for the protein backbone angles phi and psi. The structural information contained in dihedral angles can be subsequently used in the

protein structure calculation. Moreover the experimental chemical shifts can also be used as input for a structure prediction using CS-ROSETTA software (114). CS-Rosetta protocol matches the experimental chemical shifts of the backbone to a structural database to identify protein fragments with similar chemical shifts. Selected protein fragments are then used as input for a fragment assembly procedure to make a reliable prediction of the 3D structure. The output models are then re-evaluated by adding a chemical shift term to the all atom energy score. The chemical shift term is basically a term that compares the original chemical shifts with the back-calculated chemical shifts of a generated model (using SPARTA). Finally, the "new" lowest energy model is compared to the other generated models (Figure 2.14). If the difference between the models (evaluated by RMSD on Cα) and the all atom score shows convergence, the lowest energy model is deemed successful and accepted.

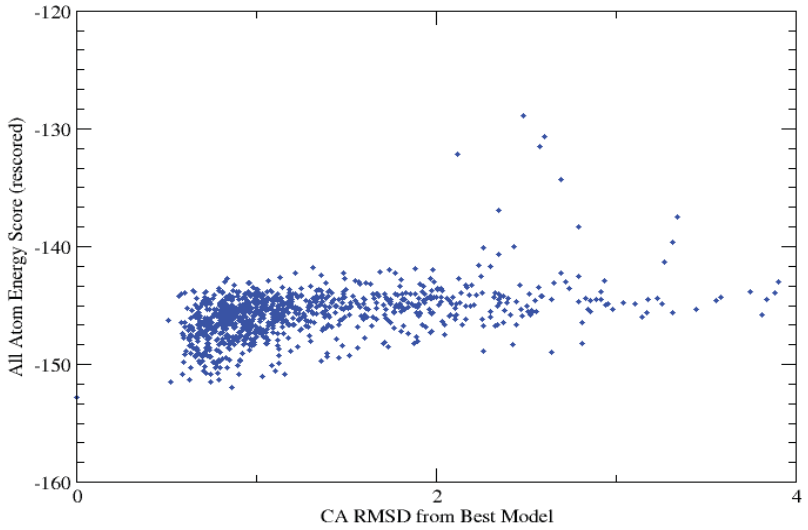


Figure 2.14 Plot of rescored Cs-Rosetta all-atom energy against Cα RMSD relative to the lowest-energy model. The convergence for this model shows a “funneling distribution”, suggesting that the CS-Rosettastructure generation process has converged.

2.4 NMR Experiments for the collection of intra-molecular restraints for protein structure calculation

The determination of three-dimensional protein structures in the solution state by NMR spectroscopy relies largely upon the measurement and interpretation of a large number of Nuclear Overhauser Effect (NOE) interactions between the neighbouring protons. The NOE originates from cross-relaxation between dipolarly coupled spins. The intensity of an NOE is related to the distance r between the two interacting spins as $1/ r^6$. Thus, from the NOE intensities, it is possible to derivate distances restraints, which are used as input for the structure calculation.

After the assignment of all or nearly all resonances of a protein, 3D ^{13}C- and ^{15}N- NOESY-HSQC spectra are generally used to obtain the NOEs structural information. The NOESY-HSQC pulse sequence basically consists of an homonuclear 2D NOESY pulse sequence followed by a ^1H-X HSQC pulse sequence (where X=^{13}C or ^{15}N heteronucleus). The 2D NOESY (Nuclear Overhauser Effect SpectroscopY) experiment is based on the same principles of the homonuclear 2D experiments, while a 2D HSQC (Heteronuclear Single Quantum Correlation) pulse sequence starts with the proton magnetization and ends up with proton detection. In the first part of the sequence the proton magnetization is transferred with an appropriate pulse sequence via $^1J_{XH}$ to the heteronucleus that evolves during the t_1 period. In this way, the proton spin is labelled with a modulation arising from the heteronucleus. After the time t_1 the magnetization is transferred back to the proton and detected during t_2.

In the combined 3D ^{15}N-NOESY-HSQC (115-117) NOE connectivities between the backbone amide protons (^1H_N) and the spatially neighboring protons are observed. In particular the experiment allows the detection of protons, either within the same or in other residues, that are closer than 5 Å. In this experiment, the magnetization is exchanged between all ^1H using the NOE (2D homonuclear NOESY experiment) and then is transferred to ^{15}N, via $^1J_{NH}$, and finally back to ^1H_N for detection (^1H-^{15}N HSQC experiment) (Figure 2.15). The ^{15}N NOESY-HSQC spectrum can

provide information that can be used also in the assignment procedure to identify the sequential neighbor.

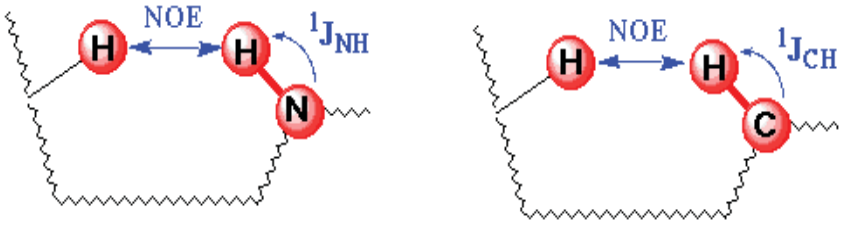


Figure 2.15 Schematic illustration of correlations observed in a 3D NOESY-HSQC. The mechanism involves ^{1}H-^{1}H NOE step followed by an heteronuclear transfer via $^1J_{XH}$.

Characteristic sequential NOEs, which are, however, secondary-structure dependent (100), can be observed in the 3D NOESY-HSQC spectra (Figure 2.16). For α-helices, sequential amide protons are in close contact, giving rise to amide-amide (^1H_N-^1H_N) NOEs. On the other hand, in β-sheets, very strong NOEs between the amide protons and the ^{1}Hα of preceding residue (^1H_N(i)-^{1}Hα(i-1)) are generally observed.

The 3D ^{13}C-NOESY-HSQC (115-117) provides NOE correlations between an aliphatic ^{13}C-^{1}H pair and all protons with an internuclear distance of less 5 Å. Similarly to the 3D ^{15}N-NOESY-HSQC experiment, the mechanism involves an initial step where the magnetization is exchanged between all hydrogens using the NOE and a second step in which the magnetization is transferred to the heteronucleus via $^1J_{CH.}$ To observe the NOEs between an aromatic ^{13}C-^{1}H pair and all protons that are close in space, it is necessary to record a second 3D ^{13}C-NOESY-HSQC, optimizing the ^{13}C frequency offset and the value of the^{1}H-^{13}C couplings used during the pulse sequence.

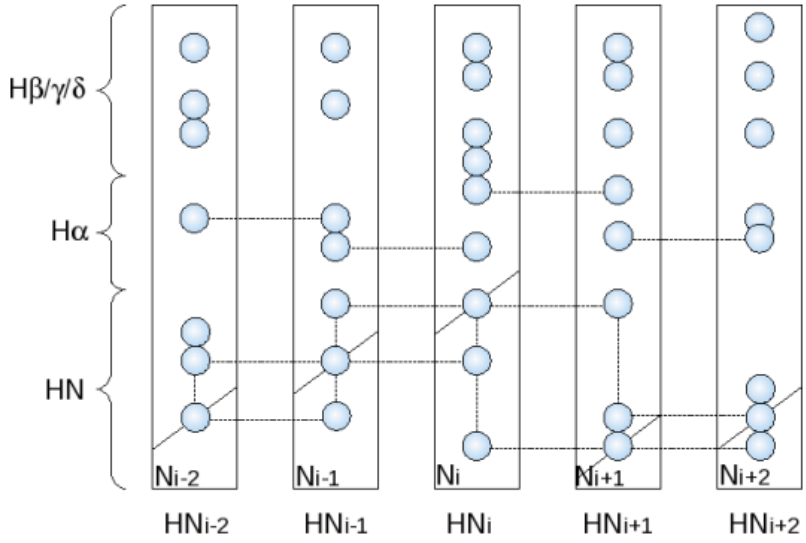


Figure 2.16 The ^{15}N-NOESY-HSQC provides the NOEs correlation between an amide proton and all ^{1}H which are within of 5Å of the $^{1}H_{N.}$ The strips show the observable NOEs for sequential amide groups. The lines indicate the observable $^{1}H_{N}$-$^{1}H_{N}$ and $^{1}H_{N}$(i)-^{1}H□(i-1) correlations.

2.5 NMR Experiments for the collection of inter-molecular restraints between protein and ligands

Knowledge about the binding site and the distances of nuclei of two binders is necessary to obtain an high resolution structure of a complex. Several NMR approaches can be used for the identification of protein and ligand atoms involved in the interaction and their intermolecular distances.

In my thesis work, the NMR approaches used to collect inter-molecular information can be divided into two classes, depending on whether i) protein resonances or ii) both protein and ligand resonances are observed.

i. The observation of protein-ligand binding relies on the detection of chemical shift changes of protein resonances upon titration of the protein with increasing amounts of ligand and it is called chemical shift perturbation (CSP). If a ligand containing a paramagnetic ion is

employed, broadening of protein signals can provide information on intermolecular restraints. Indeed protons located 15 to 20 Å away from the paramagnetic ion can experience an increase in their relaxation rates. This effect is called paramagnetic relaxation enhancement (PRE).

ii. This second approach relies on the observation of intermolecular cross-relaxation effects between the protein and the ligand. The identification of intermolecular NOEs requires the assignment of both protein and ligand resonances and provides the most important information for the structure determination.

2.5.1 NMR Experiments for the identification of the protein binding site

The most used NMR method for studying protein-ligand interactions is based on chemical shift changes resulting from complex formation. The HSQC experiments allow the detection of ligand binding over a wide range of K_D values. The HSQC spectra can be performed on a ^{15}N or on a ^{13}C labelled protein. In the first case, the ^{1}H-^{15}N HSQC is an experiment designed to correlate proton chemical shifts with the chemical shifts of the ^{15}N atoms to which they are attached. In the resulting NMR spectrum, each signal corresponds to an amide group (NH group) of each amino acid residue. Thus a ^{1}H-^{15}N HSQC spectrum is like a "fingerprinting" of a protein, where the number of peaks in the spectrum is indeed close to the number of residues in the protein (NH groups of the backbone plus NH_2 groups of the side chains, such as Asparagine and Glutamine, and protons attached to nitrogen of aromatic residues such as Tryptophan and Histidine).

The binding-induced shift can derive from two distinct phenomena:

- upon ligand addition chemical shifts of the protein residues at the binding interface will be perturbed by the proximity of the partner;
- the addition of the ligand may induce structural rearrangement in the protein, even at distant sites.

Thus the analysis of the shift changes can give information on the binding interface and on the conformational changes of the protein, even though it not always possible to disentangle the two effects. The identification of residues involved only in the binding interface, can be better achieved through a ^{1}H-^{13}C HSQC titration experiment, as $^{13}C\alpha$ shift changes should be less sensitive to conformational changes with respect to NH signals.

The protein binding site can also be easily identified when the ligand contains a paramagnetic metal ion. Indeed the presence of unpaired electrons on the paramagnetic centre greatly affects longitudinal and transverse relaxation rates (R_1 and R_2 respectively) of neighbouring nuclei due to spin-electron dipolar interactions. Thus NMR lines become broader on decreasing the distance between the metal and the observed nucleus. The comparison of 2D ^{1}H-^{15}N HSQC spectra of the ^{15}N labelled protein in the absence and in presence of the ligand bearing the paramagnetic ion, allows the quantification of intensity changes of affected resonances. As relaxation rate enhancement effects depend on the metal-to nucleus distance, it is possible to calculate such distance according to the following equation:

$$\Delta R_{1,2} = f K_{dip} / r^6 \qquad (2.6)$$

where $\Delta R_{1,2}$ is the difference in relaxation rates measured in the absence and in presence of the paramagnetic ligand, f is the occupancy factor of the binding site and K_{dip} is the sum of the Salomon and Curie constants (118) and depends on operating frequencies, on the electronic relaxation times of the metal ion, on the lifetime of the protein-ligand adduct and on the overall rotational correlation. Once an estimate of the proportionality constant is given the metal-nucleus distances can be obtained.

2.5.2 NMR filtered experiments for the assignment of ligand resonances

Assignment of the ^{1}H resonances of unlabelled small ligand when bound to a macromolecule cannot be performed in a simple 2D homonuclear NOESY experiment due to the presence of a large number of cross peaks

resulting from the dipole-dipole interaction of both protein and ligand protons. The use of isotope-filtered experiments in combination with appropriate labelling strategies, allows distinguishing intramolecular NOEs within the bound ligand or protein and protein-ligand intermolecular NOEs. It is often more straightforward and cost-effective incorporate isotope labels into the protein than into the ligand. Thus, in the example of a binary complex consisting of a $^{15}N^{13}C$-labelled protein and an unlabelled molecule (ligand), three kinds of NOE cross peaks can be observed (Figure 2.17):

- intramolecular NOEs of protein (^{1}H-(^{13}C; ^{15}N) ↔ ^{1}H-(^{13}C; ^{15}N));
- intramolecular NOEs of ligand (^{1}H-(^{12}C; ^{14}N) ↔^{1}H-(^{12}C; ^{14}N));
- intermolecular NOEs between protein and ligand (^{1}H-(^{13}C; ^{15}N) ↔^{1}H-(^{12}C; ^{14}N))

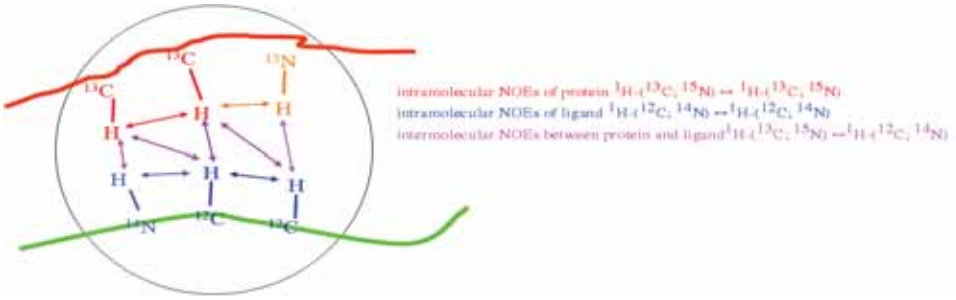


Figure 2.17 Scheme of the intramolecular and intermolecular NOEs that can be observed in a complex between a labelled protein and unlabelled ligand.

The connectivity of protons to the heteronuclei ^{13}C and ^{15}N can be used to separate intra- and intermolecular NOEs of protein and ligand. Two different experiments are used: **Isotope-filtered experiments** that remove coherences of heteronucleus-bound protons and detect only ^{1}H signals attached to ^{12}C and ^{14}N nuclei; **Isotope-edited (isotope-separated) experiments** that select the magnetization of ^{1}H bound to ^{13}C and ^{15}N heteronuclei and remove the ligand signals, i.e. the ^{1}H signals attached to ^{12}C and ^{14}N nuclei. Examples of isotope-edited experiments are the 3D NOESY-HSQC, in which the magnetization of ^{1}H bound to heteronucleus

is selected to detect intramolecular protein NOEs.
In order to identify only the signals of unlabelled ligand, two different experimental strategies have been proposed for filtering. The original idea behind filtering methods comes from Ernst and co-workers, who introduced the concept of purge filter (or low-pass J filter) to allow the suppression of coherences of heteronucleus-attached protons on the basis of J-coupling evolution (119). A second method, originally developed by Wuthrich and co-workers (120, 121) makes use of a spin echo difference scheme, whereby ^{1}H-^{13}C or ^{1}H-^{15}N pairs are selected against on the basis of the difference in sign of magnetization originating from protons that either are or are not one-bond coupled to ^{13}C or ^{15}N spins. In figure 2.18 is reported the pulse scheme of isotope filtering with a low pass J filter element that rejects the magnetization of protons attached to the heteronucleus.

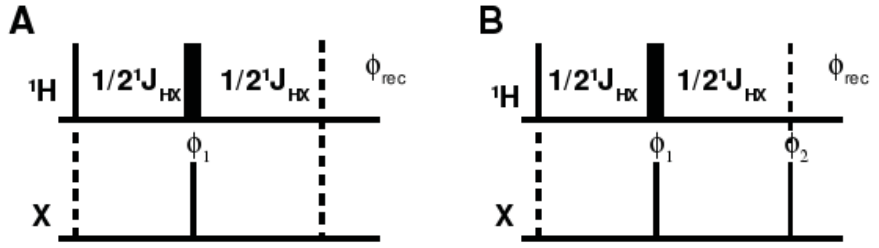


Figure 2.18 Pulse sequences with purge filter elements. A) basic low pass J filter with 90°(X) purge pulse. For optimal performance the delay is set to $1/2J_{HX}$. B) composite purge filter produced by concatenation of two purge filter elements.

The pulse sequence starts with a 90° proton pulse followed by a delay optimized on the one bond H-X J-coupling constant ($1/(2^{1}J_{HX})$). During this delay, the proton magnetization (coupled to the heteronucleus) acquires an anti-phase character with respect to its directly attached ^{13}C or ^{15}N spin. At this point a 90° purge pulse on the heteronucleus X converts the anti-phase term (H-{X}) into multiple quantum magnetization that is unobservable, while leaving the in-phase magnetization of the non-X-bound protons unaffected. The 180° pulse on the protons refocuses the

proton chemical shift evolution that occurs during the delay. To achieve better suppression of anti-phase magnetization that escapes conversion to heteronuclear multiple quantum magnetization (as a result of pulse imperfections), the phase ϕ of the X-nucleus 90° pulse is cycled x, -x. To improve the level of suppression, a second purge can be applied after a further $1/(2^1J_{HX})$ delay. The resulting composite filter is more tolerant to slight variation in J-couplings. This is especially useful for the acquisition of ^{13}C-filtered ^{1}H spectra where there is a spread of $^1J_{HC}$ value ($^1J_{HCali}$=120-145 Hz, $^1J_{HCaro}$=160-220Hz). In this concatenation approach, the two purge filter elements can be tuned for different values of $^1J_{HC}$ and thus much more efficient suppression can be achieved over a wider range of $^1J_{HC}$.

In the past several different implementations of this approach have been proposed (122). In this thesis, we performed doubly-filtered [F_1-C/N, F_2-C/N] NOESY experiments (123) with heteronuclear filter acting in both F_1 and F_2 for observation of exclusively intramolecular distances within the unlabelled bound ligand, The doubly-filtered [F_1-C/N,F_2-C/N] NOESY uses the low-pass J-filter principle, and permits to remove simultaneously the coherence of ^{15}N- and ^{13}C-bound protons by converting it to undetectable multiple-quantum coherence. Moreover the uniform suppression of proton resonances attached to the ^{13}C heteronucleus is achieved optimizing the suppression of aliphatic resonances in the F_1 dimension and aromatic resonances in the F_2 dimension. In this experiment the NOEs originating from the ^{15}N^{13}C-labelled protein are filtered out in both dimensions and thus the resulting spectrum contains only the intramolecular NOEs of ligand (^{1}H-(^{12}C; ^{14}N) $\leftrightarrow$^{1}H-(^{12}C; ^{14}N)). In figure 2.19 is reported a comparison between a 2D NOESY experiment recorded without any filtering and with [F_1-C/N, F_2-C/N] filtering.

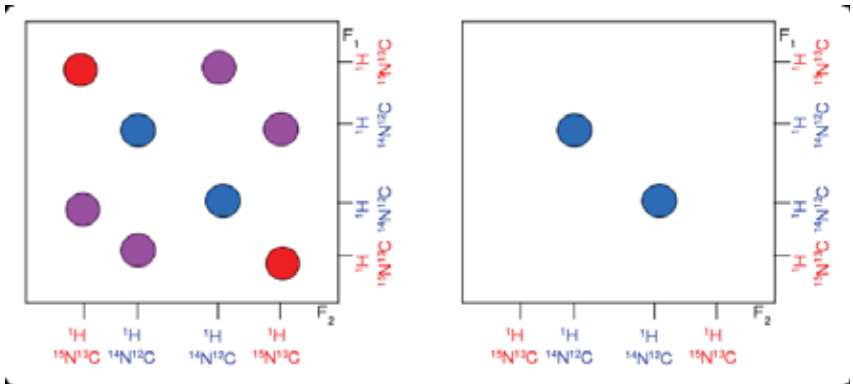


Figure 2.19 Observable peaks in a 2D NOESY experiment (left) and in doubly-filtered [F_1-C/N, F_2-C/N] NOESY experiment (right). In the 2D NOESY spectrum intramolecular NOEs of the protein (red circles), intramolecular NOEs of the ligand (blue circles) and intermolecular NOEs between protein and ligand (pink circles) are schematically drawn. In the doubly-filtered NOESY experiment all ^{1}H signals attached to $^{13}C^{15}N$ nuclei are rejected and thus only the intramolecular NOEs of ligand are observed.

A drawback of this experiment is that signals belonging to the free ligand can be observed in the spectrum. The result is that a more crowded spectrum is obtained and the assignment of bound-ligand resonances becomes more complicated. An approach to detect only proton signals of the unlabelled ligand experiencing NOEs with labelled protein (bound ligand protons) is the acquisition of filtered isotope experiments for observing intermolecular NOEs (next section).

2.5.3 NMR filtered experiments for the detection of intermolecular distances

In order to observe only intermolecular NOEs between the labelled protein and an unlabelled bound ligand, isotope-filtering has to be applied in the one dimension, and a isotope-editing in the other dimension (Figure 2.20). In the dimension where filtering is applied are detected only proton signals from unlabeled protons, while in the dimension where editing is applied the only proton signals that are directly detected are from labelled protons. Thus, the resulting spectrum contains only the intermolecular NOEs between labelled protein and unlabelled ligand (^{1}H-(^{13}C; ^{15}N) $\leftrightarrow$ ^{1}H-(^{12}C; ^{14}N)).

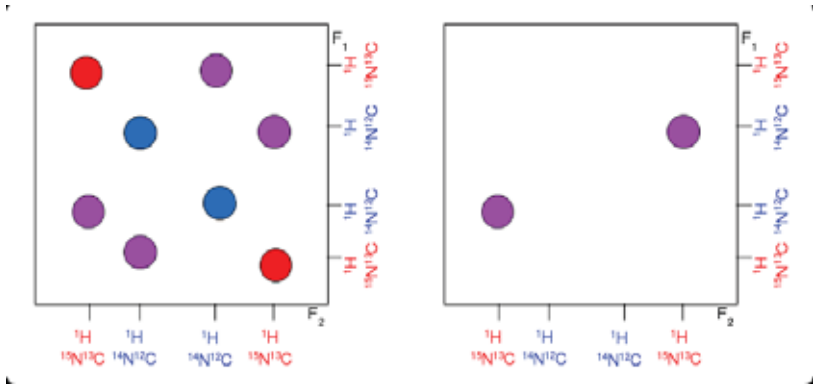


Figure 2.20 Observable peaks before and after the application of an isotope-filter in one dimension (in this example F_1) and an isotope-editing in the other dimension (F_2). The peaks are coloured as reported in the previous figure. In F_1 the ^1H signals attached to $^{13}C^{15}N$ nuclei are removed (observable only ^1H signals of ligand), while in F_2 are select the signal of labelled protein and the ^1H signals of ligand are removed. Thus the resulting spectrum will show only the intermolecular NOEs between protein and ligand.

In a simple 2D experiment the unambiguous assignment of protein protons is difficult due to high overlap, thus it is often beneficial to conduct the isotope-editing as a multidimensional experiment (and in this case it is referred as isotope-separation), exploiting the chemical shift dispersion of the heteronucleus to increase the spectral resolution for the labelled protein. This can be done using hybrid schemes combining purge filters for selecting non-X bound proton signals in one proton dimension with sequences for recording 3D isotope-separated NOESY spectra.

The identification of intermolecular NOEs in this study has been performed recording a 3D F_1-[^{13}C]-filtered,F_2-[^{13}C]-separated,F_3-[^{13}C]-edited NOESY-HSQC spectrum in D_2O. The pulse sequences of the experiment select for magnetization transferred from protons on the unlabelled component to proximal protons of the labelled molecule. Differently from 2D double filtered NOESY used to observe intramolecular NOEs of ligand, in this experiment the purge filter element is a WURST pulse (124) (Figure 2.21).

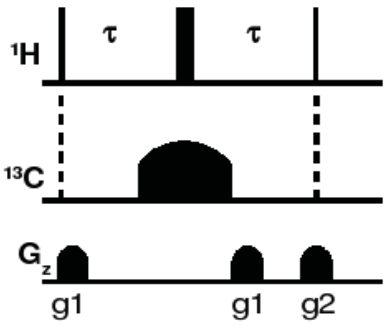


Figure 2.21 WURST-based purge filter element.

In this filter element the conventional ^{13}C inversion pulse is replaced by a pulse whose frequency is swept adiabatically during the filter delay, such that inversion of carbon magnetization is inverted at different times for carbon of different chemical shift. During this period (2τ) the desired magnetization from protons not scalarly coupled to ^{13}C is stored along the z-axis. At the end of the 2τ period, the gradient g2 dephases the antiphase magnetization arising from ^{13}C-bound protons and thus only the protons attached to ^{12}C evolve during t_1. The filtered magnetization is then subjected to a NOESY pulse sequence, followed by an HSQC sequence, which is used to correlate the protons which are "excited" by the NOESY mixing to ^{13}C through their ^{1}H-^{13}C scalar couplings. After the chemical shift evolution of ^{13}C in t_2, the protons directly attached to it are detected.
In F_1 dimension of the resulting experiment only the signals of the unlabelled ligand protons are observable, while in the other two dimensions only the proton (F_3) and carbon (F_2) signals of the labelled protein are observable.

As underlined in the previous section (2.5.2), filtered experiments that combine filter and isotope elements can also be used to detect only the signals of bound ligand.
Only resonances of protons bound to ^{12}C and exhibiting NOE with protein can be observed in the Fourier-transformed first fid of several experiments F_1-edited, F_3-filtered 3D HMQC-NOESY recorded with

different mixing-time (125). Differently from 3D F_1-filtered, F_2-separated, F_3-edited NOESY-HSQC, the sequence of 3D F_1-edited, F_3-filtered HMQC-NOESY experiment starts with a heteronuclear multiple quantum coherences block (in which are observable the correlations between directly-bonded ^{1}H and X-heteronuclei) and the succeeding ^{1}H-^{1}H NOE mixing time is followed by a ^{13}C double purging filter. During the mixing time, the magnetization is transferred between dipolarly coupled spins and only those NOEs corresponding to the transfer of magnetization from ^{13}C-bound protons to ^{12}C-bound protons are selected during the subsequent purging scheme. By recording spectra incrementing the mixing time period, it is possible to observe the growth of NOEs signals resulting from intermolecular interaction. The superimposition of different transformed first fid allows establishing the assignment of protons ligand (Figure 2.22).

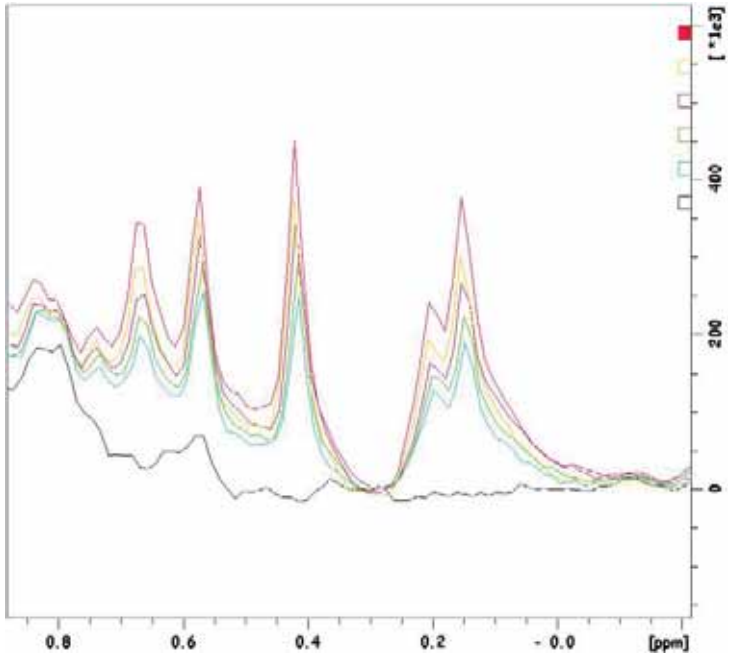


Figure 2.22 Fourier transformed first fid of different 3D F_1-edited, F_3-filtered HMQC-NOESY experiments recorded increasing the mixing time. In this superimposition the growth of the proton signals of ligand that exhibited NOEs only with the protein is observable. In black the experiment (reference spectrum) recorded with mixing time zero.

At this stage, once established the assignment of some proton resonances of the ligand and ^{13}C-^{1}H pairs of almost all the labelled protein, it is possible to assign the intermolecular NOEs observed in the 3D F_1-filtered, F_2-separated, F_3-edited NOESY-HSQC spectra and obtain the restraints needed for the structural determination of protein-ligands complex.

2.6 Proteins structure calculation

In de novo 3D structure determinations of proteins in solution by NMR spectroscopy, the key conformational data are upper distance limits derived from NOEs. In order to extract distance constraints from a NOESY spectrum, its cross peaks have to be assigned, that is, the pairs of interacting atoms that give rise to the observed cross peak have to be identified. Because of the limited accuracy of chemical shift values and signals overlap, many NOESY cross peaks cannot be attributed unambiguously to a single unique spin pair. In addition to the problem of resonance and peak position, considerable difficulties may arise from spectral artefacts, noise and from the absence of expected signals because of fast relaxation or conformational exchange. In principle it is possible to solve such ambiguities during the structure calculation process using an automatic interactive procedure. Several structure calculation softwares have been supplemented with the automated NOESY assignment routine. In my thesis work the structure calculation has been performed using the program CYANA (126), which contains the CANDID (127) algorithm for automated NOE assignment. Figure 2.23 reports a general scheme showing the iterative approach of automated combined NOESY assignment and structure calculation used in the CYANA software.

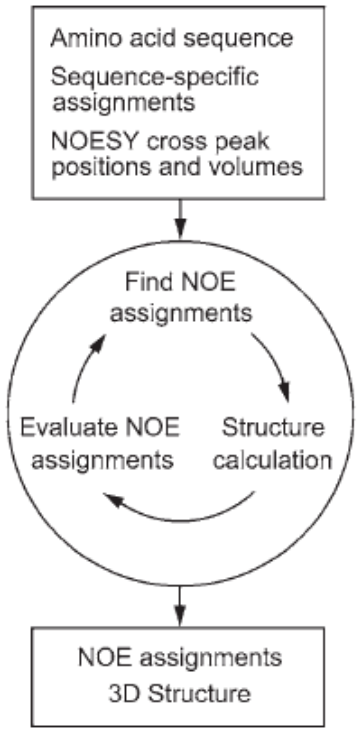


Figure 2.23 General scheme of an automated combined NOESY assignment and structure calculation.

The automated CANDID method proceeds in iterative cycles (seven cycles), each consisting of ambiguous NOE assignment followed by a structure calculation with the CYANA torsion angle dynamics algorithm. Between subsequent cycles, information is transferred exclusively through the intermediary 3D structures, so that the protein molecular structure obtained in a given cycle is used to guide further NOE assignments in the following cycle. Otherwise, the same input data are used for all cycles, i.e. the amino acid sequence of the protein, a complete chemical shift assignment list and one or several lists containing the positions and volumes of cross peaks collected in 3D dimensional NOESY spectra.

A CANDID cycle starts by generating for each NOESY cross peak an initial assignment list from the fit of chemical shift values with the peak positions, using a user-defined tolerance range. Subsequently, for each cross peak these initial assignments are weighted with respect to several criteria and initial assignments with low overall score are then discarded. The filtering criteria include the agreement between the values of the chemical shift list and the peak position, the presence of symmetry-related cross peaks, the compatibility with covalent polypeptide structure and, most importantly, element, the self-consistency within the entire NOE network. The second and subsequent cycles differ from the first cycle in the use of additional selection criteria for cross peaks and NOE assignments that are based on assessments relative to the protein three-dimensional structure from the preceding cycle. For each cross peak, the retained assignments are interpreted in the form of an upper distance limit derived from the cross peak volume. In order to reduce deleterious effects on the resulting structure from erroneous distance constraints that may pass this filtering step, long-range distance constraints are incorporated into "combined distance constraints". The distance constraints are then included in the input for the structure calculation with the CYANA torsion angle dynamics algorithm. In the first cycle, network anchoring and constraints combination have a dominant impact for the successful generation of the correct polypeptide fold, since structure-based criteria cannot be applied yet. Network-anchoring exploits the observation that the correctly assigned constraints form a self-consistent subset in any network of distance constraints is sufficiently dense for the determination of a protein three-dimensional structure. Network-anchoring thus evaluates the self-consistency of NOE assignments independently of knowledge of the three dimensional protein structure.

Constraint combination consists of generating virtual distance constraints with combined assignments from different, generally unrelated, cross peaks. The basic property of ambiguous distance constraints, namely that the constraints will be fulfilled by the correct structure whenever at least

one of its assignments is correct, regardless of the presence of additional erroneous assignments, implies that such combined constraints have a lower probability of being erroneous with respect to the corresponding original constraints. In the final CANDID cycle (cycle 7), an additional filtering step ensures that all NOEs have either unique assignments to a single pair of hydrogen atoms, or are eliminated from the input for the structure calculation.

Each CANDID cycle is completed by a structure calculation using the CYANA torsion angle dynamics algorithm, which uses as input the list of distance constraints from CANDID and possibly additional conformational constrains such as dihedral angle constraints.

In the program CYANA, the structure calculation is based on molecular dynamic simulated annealing (SA) protocol performed in torsion angle space. A general feature of this protocol is that it uses a "target function" that measures how well the calculated structure fits the experimental data and the chemical information; the lower this function, the better the agreement. The chemical information is defined in the force field that contains terms such as bond length, bond angles, van der Waals interactions. Therefore, the calculation of the 3D structure is formulated in CYANA as a minimization problem for a target function that measures the agreement between a structure and the given set of constraints.

The minimization algorithm in CYANA is based on simulated annealing by molecular dynamic simulation in torsion angle space. Torsion angle dynamics consists in molecular dynamic simulation using only torsion angles as degree of freedom. The degree of freedom of torsion angles dynamics is the rotation about covalent bonds, so the number of degrees of freedom is reduced respect to the Cartesian coordinates. Compared to the minimization of a target function, the molecular dynamics simulation comprises also the kinetic energy that allows overcoming barriers of potential, reducing the probability of encountering a local minimum.

The potential energy landscape of a protein is complex and studded with many local minima. Because the temperature, i.e. the kinetic energy,

determines the maximal height of energy barriers that can be overcome in a molecular dynamics simulation, the temperature schedule is important for the success and efficiency of a simulated annealing calculation. The standard simulated annealing protocol in CYANA starts from a conformation with all torsion angles treated as independent and consists of five stages. An initial minimization step, including initially only distance constraints between atoms up to 3 residues apart along the sequence, is carried out in order to reduce high-energy interactions that could disturb the torsion angle dynamics algorithm. In the second stage (high-temperature phase) a torsion angle dynamics calculation at constant high temperature is performed, then followed by torsion angle dynamics calculation with slow cooling to temperature close to zero. Up to this point the hydrogen atoms have not been included in the calculation. They are included in this fourth stage of low temperature phase to check for steric overlap. The final stage consists of a minimization of 1000 conjugate gradient steps. With the CYANA torsion angle dynamics algorithm it is possible to efficiently calculate protein structures on the basis of NMR data.

The structures of proteins calculated in this thesis work were obtained using the standard torsion angle dynamics simulated annealing protocol. Hundred trial structures were generated in each cycle and the 20 structures with the lowest target function values were chosen to represent the solution structure. The correctness of the resulting 3D structure has been assessed using the CYANA output criteria; these criteria are as follow:

- average target function for cycle 1 below 250 Å̊
- average target function for cycle 7 below 10 Å̊
- less than 20% unassigned NOEs
- less than 20% discarded long- range NOEs
- RMSD for cycle 1 below 3Å̊
- RMSD between cycles 1 and 7 below 3Å̊

For the assessment of protein structures quality, two different tools were used, the i-CING (128) and the PSVD (129) server. I-CING is part of the Common Interface for NMR structure Generation (CING) package and integrates the data generated by PROCHECK, WHAT_IF, Wattos, ShiftX. The analysis with the i-CING suite provides information on the stereochemical quality of the structures as well as their agreement with experimental restraints. i-CING is particularly informative on possible errors at the "per-residue" level using a very straightforward colouring scheme for visualisation of all its assessments (peaks, restraints, residues, assignments). This is called ROG coding, where:

- Red means there is something wrong;
- Orange means there might be something wrong;
- Green means it is ok.

The other validation suite used was PSVS. Also PSVS checks stereochemical parameters as well as the agreement with experimental data. The latter integrates analyses from several widely used structure quality evaluation tools, including PROCHECK, but also RPF, MolProbity, Verify3D, Prosa II and the PDB validation software. PSVS provides standard constraint analyses, statistics on the PDB validation goodness-of-fit between structures and experimental data, and knowledge-based structure quality scores in standardized format suitable for database integration. Global quality measures are reported as Z scores, based on calibration with a set of high-resolution X-ray crystal structures.

2.7 Data driven docking (HADDOCK)

As stated initially, the docking protocol used to obtain high resolution structure of the protein-ligand complex has been performed by HADDOCK (*High Ambiguity Driven biomolecular DOCKing)* approach The docking protocol is implemented in CNS (95) for structure calculations and makes use of python scripts derived from ARIA for automation (130).

The docking protocol requires the PDB files of protein and ligands as input together with ambiguous and unambiguous distance restraints. The starting holo protein structures were the 10 best structures (lowest target function) of the final bundle obtained by CYANA calculation. The starting coordinates of ligands were obtained with the program SMILE (131), while the topology and parameter files were calculated using PRODRG server (132). Ambiguous interaction restraints (AIR), that provide information on the interaction interface, were derived from NMR experiments, such as chemical shift perturbation and filtered experiments. In Haddock the protein residues directly affected by ligand interaction are defined as "active residues", while neighbours of active residues are defined as "passive residues". For every active residue, a single AIR is defined between that residue and all active and passive residues on the binder molecule. The distance of an AIR between any atom m of active residue i of the protein A (*mi*A) and any atom n of both active and passive residues k (*N*res in total) of the molecule B (*nk*B) (and inversely for molecule B) is calculated using the equation:

$$d_{\mathrm{iAB}}^{\mathrm{eff}} = \left(\sum_{m_{\mathrm{iA}}=1}^{N_{\mathrm{atoms}}} \sum_{k=1}^{N_{\mathrm{res}}\mathrm{B}} \sum_{n_{\mathrm{kB}}=1}^{N_{\mathrm{atoms}}} \frac{1}{d_{m_{\mathrm{iA}} n_{\mathrm{kB}}}^{6}} \right)^{(-1/6)} \qquad (2.7)$$

where d_{iAB}^{eff} is the effective distance between the two molecules A and B, *N*atoms indicates all atoms of a given residue and *N*res the sum of active and passive residues for a given protein. A $1/r^6$ sum averaging is used, not by analogy to NOE restraints, but because this mimics the attractive part of a Lennard-Jones potential and ensures that the AIRs are satisfied as soon as any two atoms of the two molecules are in contact. The definition of passive residues ensures that residues that are at the interface but are not detected by NMR experiments, are still able to satisfy the AIR restraints.

Intermolecular NOE-derived distances obtained from the analysis of NMR isotope-filtered experiments were introduced in the docking. Since

initially only a small number of assigned intermolecular distance restraints can be used, we performed several docking calculations to identify and assign new intermolecular NOEs, in order to improve the accuracy of final protein-ligand complex structure.

To solve the structure of the protein in complex with a paramagnetic ligand, additional ambiguous and unambiguous restraints can be added. In particular, the paramagnetic attenuation of protein amide resonances due to the presence of paramagnetic ion are used as AIRs, while intermolecular metal-nucleus distances derived by paramagnetic relaxation measurement are introduced as unambiguous intermolecular restraints to drive the docking.

To obtain a high resolution structure of the protein-ligand complex with a docking approach, additional restraints derived from the holo CYANA structure are used as input in order to ensure the consistency with the all experimental data, i.e. that the protein structure in the complex satisfies the experimentally derived intramolecular distances. Specifically, we used dihedral angles restraints derived by TALOS to maintain the secondary structure of the protein, and also, as unambiguous distances, all intramolecular NOEs distances used to calculate the final ensemble of holo protein structure.

The HADDOCK docking protocol consists of three stages:

- randomization of orientations and rigid body energy minimization (EM);
- semi rigid simulated annealing in torsion angle space (TAD-SA);
- final refinement in Cartesian space with explicit solvent.

In the randomization stage, the protein and ligands are positioned at 15 Å from each other in space and each one is randomly rotated around its centre of mass. Rigid body EM is then performed: first, four cycles of orientational optimization are performed in which each binder in turn is allowed to rotate to minimize the intermolecular energy function. Then both translations and rotations are allowed, and the protein-ligand molecules are docked by rigid body EM. For each starting structure

combination, this step can be repeated a number of times. Generally, a large number of complex conformations are calculated in this stage (in the order of 1000 to 4000 solutions). The best solutions in terms of intermolecular energies are then select for the next step.

In the second stage, three different simulated annealing are performed, during which increasing amounts of flexibility are introduced. That permits that the amino acids at the interface (side chains and backbone) are allowed to move to optimize the interface packing. In the first simulated annealing, the binders are considered as rigid bodies and their respective orientation is optimized. In the second simulated annealing only the side chains of protein at the interface are allowed to move. In the third simulated annealing both side chains and backbone at the interface are allowed to move to allow for some conformational rearrangements. The resulting structures are then subjected to a minimization steps.

The final stage consists of a final water refinement in an explicit solvent layer (8Å for water). Although no real significant structural changes occur during the water refinement stage, the side chains and backbone at the interface are free to move. That is useful for the improvement of the energetics of the interface.

The final structures are clustered using the pairwise backbone RMSD at the interface and scored using a combination of energy terms (E_{elec}, E_{vdw}, E_{AIR}) and their buried surface area. The final ensemble of protein-ligand complex is represented by the best (lowest) HADDOCK score structures of the best (lowest) HADDOCK score cluster.

The structures of protein-ligand complexes calculated in this thesis work were obtained using the standard HADDOCK protocol, but defining the entire protein as fully flexible. This option allows the use of HADDOCK, in combination with a reasonable number of unambiguous interaction restraints, for structure calculations of complexes as “classical” NOE-based structure determination approach. Thus, the final bundle of protein-ligand complex obtained can be considered as NMR solution structure and not as a structural model.

Chapter 3: Solution S tructure of the Supramolecular Adduct between a Liver Cytosolic Bile Acid Binding Protein and a Bile Acid-Based Gadolinium(III)-Chelate, a Potential Hepatospecific Magnetic Resonance Imaging Contrast Agent

Short presentation of publication 1

Magnetic resonance imaging (MRI) is rapidly emerging as imaging tool for detection and characterization of liver lesions. To increase the imaging contrast between normal and abnormal tissue, hepatospecific contrast agents (CA) are commonly used. One of the most active areas of contrast agent development concerns the design of new **liver**-specific **contrast agents**. The tissue specific targeting of CAs is based on the understanding of the mechanism involved in their cellular uptake and interaction with carriers proteins.

Recently we have shown that a bile acid-derived Gadolinium contrast agent (Gd-DTPA-conjugated to 5β-cholanoic acid and called Gd-1) is efficiently internalized in hepatocytes and is able to bind to a liver bile acid binding protein (L-BABP) in vitro. In particular the protein is able to bind a single ligand molecule of Gd-1 with a low dissociation constant (20 μM). The result prompted us to study the structural details of the complex with the purpose of identifying the features of the interaction that only can help in the rational design of new hepatospecific CAs.

In this study, the NMR solution structure of the complex has been solved using the data-driven docking software HADDOCK. This approach allows obtaining high resolution structures of protein-ligand complexes starting from the structure of the protein and incorporating distance restraints between the protein and the ligand obtained experimentally by NMR.
As L-BABP undergoes a substantial structural rearrangement upon Gd-1 binding, a significant improvement of the docking approach is achieved if the structure of the holo protein is used as starting structure. The first step in the structure determination of the holo protein requires the complete assignment of all the protein resonances in order to proceed to the identification of intramolecular protein distance restraints. The final structure ensemble of L-BABP was determined using the software CYANA. This ensemble has been subsequently used for the docking calculation. To guide the docking we used both ambiguous (AIR) and unambiguous intermolecular distances. The AIR

were obtained from chemical shift perturbations (CSPs) of protein amide resonances upon addition of ligand and from paramagnetic attenuation of amide signals due to the presence of Gd ion. Unambiguous interaction restraints were further introduced in the docking with the aim of obtaining an accurate orientation of the ligand. These restraints include both intermolecular distances between the protein and the steroid moiety of the ligand (bile acid) and distances between the protein and the metal ion. The identification of intermolecular restrains between the protein and the steroid moiety was performed analyzing isotope edited/filtered NOESY experiments. Only 14 intermolecular restraints have been identify unambiguously, due to the difficulty of assigning the highly overlapped proton ligand resonances. The intermolecular restraints between the protein and the metal ion were calculated on the basis of paramagnetic relaxation enhancement of amide proton resonances due to the presence of the paramagnetic Gd ion. Indeed, paramagnetic ion induces an enhancement of longitudinal relaxation rate of protein nuclei, and this effect is distance-dependent.
The HADDOCK structure calculation led to a final ensemble of seven structures. In the obtained complex the ligand is bound at the portal region of the protein, with the steroid moiety buried inside the hydrophobic cavity and the Gd-chelating unit, conjugated to the ring A of the steroid moiety, located outside the protein cavity.

The comparison of L-BABP-Gd-1 complex with the other available structures of holo BABPs (in complex with bile acids) underlines some important differences:

- Gd-1 accesses the protein cavity in a reverse orientation with respect to bile acids. Indeed the latter are bound with ring A of the steroid moiety buried in the protein pocket and the carboxylic tail exposed to the solvent.
- Conformational differences between the proteins in complex with natural or synthetic ligands are mainly observed in the portal region of the protein.

In conclusion, the solution structure of L-BABP in complex with the bile acids-derived contrast agent underlines the capability of this protein scaffold to adapt

itself to different ligands. In this line, the L-BABP can be considered as a good carrier for hepatospecific contrast agents.
The protein ligand interactions observed in the determined structure will contribute to the rational design and synthesis of new and more efficient hepatospecific contrast agents.

My contribution to this work consisted in production of the protein samples, in the assignment of protein resonances and in the analysis of NMR experiments for the detection of intermolecular distances restraints used in the docking approach.

Solution Structure of the Supramolecular Adduct between a Liver Cytosolic Bile Acid Binding Protein and a Bile Acid-Based Gadolinium(III)-Chelate, a Potential Hepatospecific Magnetic Resonance Imaging Contrast Agent

Simona Tomaselli,[†] Serena Zanzoni,[‡] Laura Ragona,[†] Eliana Gianolio,[§] Silvio Aime,[§] Michael Assfalg,*[,‡] and Henriette Molinari[‡]

ISMAC-CNR, Via Bassini 15, 20133 Milano, Italy, Dipartimento Scientifico e Tecnologico, Università degli Studi di Verona, 37134 Verona, Italy, Dipartimento di Chimica, Università degli Studi di Torino, Via Pietro Giuria 7, Torino, Italy

Received July 7, 2008

Bile acid-conjugated gadolinium chelates were shown to display promising features for the development of hepatospecific constrast agents for magnetic resonance imaging (MRI). The study of the pharmacokinetics of these compounds should address their possible interaction with the bile acid protein transporters. We have previously shown that a 5β-cholanoic acid-based contrast agent is efficiently internalized in hepatocytes and is able to bind to a liver bile acid binding protein (BABP) in vitro. Here we report the solution structure of the adduct between a BABP and a gadolinium chelate/bile acid conjugate. The identification of unambiguous intermolecular distance restraints was possible through 3D edited/filtered NOESY-HSQC experiments, together with distance information derived from paramagnetic relaxation enhancements. These intermolecular contacts were used for the structure determination of the complex, using the data-driven docking software HADDOCK. The obtained results represent the starting point for the design of new and more efficient MRI contrast agents.

Introduction

Magnetic resonance imaging (MRI[a]) is now a commonly used diagnostic tool as it offers a powerful way to map structure and function in soft tissues by sampling the amount, flow, and environment of water protons in vivo. The intrinsic contrast can be increased by the use of contrast agents (CAs) in both clinical and experimental settings. A large part of the CAs developed earlier were composed of gadolinium(III) [Gd(III)] complexes[1,2] because of their favorable electronic and magnetic properties. In recent years, the use of gadolinium complexes of simple polyaminopolycarboxylic ligands (e.g., diethylenetriaminepentaacetic acid (DTPA) or 1,4,7,10-tetraazacyclododecane-1,4,7,10-tetraacetic acid (DOTA)) as extracellular MRI contrast agents has become more and more widespread.[3−8] At the same time, the search for complexes that are taken up by cells, particularly in the liver, has been quite active.[9−13] Gd(III) complexes in which a lipophilic residue was introduced in the basic unit of DTPA have been developed as hepatospecific agents (e.g., Gd-BOPTA,[14] Gd-EOB-DTPA[15]). Compared to the extracellular contrast agents, these "lipophilic" complexes are taken up to a variable extent by hepatocytes in different animal species and in humans,[16] although they appear to be internalized by a passive diffusion process.[17]

More recently, the interest has grown in CAs that enter the cells by means of active transport mechanisms. A variety of contrast agents have been developed for contrast-enhanced MRI of the liver, which are designed to overcome the limitations of unspecific tissue uptake by extracellular low molecular weight gadolinium chelates.[18] Active molecular transport in hepatocytes may be conveniently realized by exploiting the enterohepatic circulation machinery. The latter is an efficient recycling mechanism by which bile salts, which are synthesized in the liver from cholesterol, enter the intestine in the form of bile, where they are absorbed by the enterocytes and transferred back to the liver through the portal vein. The key steps at both the hepatocytes and the enterocytes are mediated by a receptor system, an intracellular protein carrier, and an exit system.[19] On the basis of the knowledge of bile salt transport, a number of potential CAs were synthesized by conjugating bile acid moieties to gadolinium chelating units such as DTPA or DOTA and tested in terms of pharmacokinetic properties and toxicology.[6] Among the most promising compounds, a conjugate in which the Gd(III) complex is linked to cholic acid through the 3α hydroxyl group seems to enter hepatocytes using the Na^+/taurocholate transporter. Because it has been shown that the Na^+/taurocholate transporter is not expressed in the basolateral membrane of some human hepatoma cell lines,[20] the discovery of a contrast agent that is transported into hepatocytes by such a transporter could, in principle, allow clear-cut diagnoses of specific hepatic diseases (e.g., hepatocellular carcinoma). Noticeably, two related Gd-DTPA-based conjugates of cholanoic (Gd-1) and deoxycholic (Gd-2) acid (Scheme 1) were shown to display differential cellular uptake between human healthy model cells and hepatoma cell lines.[21]

The study of the mechanisms involved in the transport of the compound by the hepatocytes is crucial for the definition of the drug's pharmacokinetics as well as to explain the relaxivity properties that are at the basis of the magnetic resonance imaging technique. While a few studies have ad-

* To whom correspondence should be addressed. Phone +39 0458027949. Fax +39 0458027929. E-mail: michael.assfalg@univr.it.

[†] ISMAC-CNR.

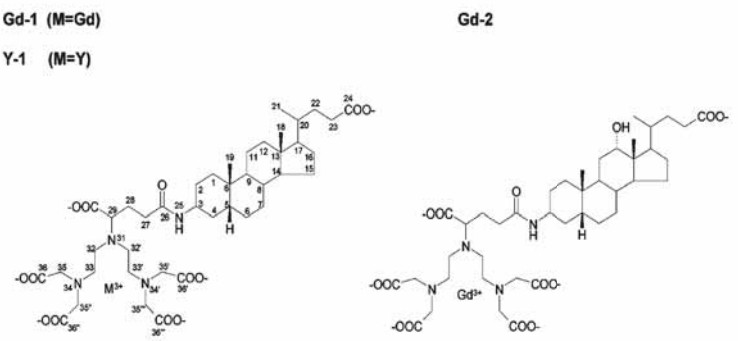
[‡] Dipartimento Scientifico e Tecnologico, Università degli Studi di Verona.

[§] Dipartimento di Chimica, Università degli Studi di Torino.

[a] Abbreviations: MRI, magnetic resonance imaging; CA, contrast agent; DTPA, diethylenetriaminepentaacetic acid; DOTA, 1,4,7,10-tetraazacyclododecane-1,4,7,10-tetraacetic acid; BABP, bile acid binding proteins; cl-BABP, chicken liver bile acid binding proteins; CHA, cholic acid; CDA, chenodeoxycholic acid; NOESY, nuclear Overhauser effect spectroscopy; HSQC, heteronuclear single quantum coherence; NOE, nuclear Overhauser effect; rmsd, root-mean-square deviation; AIR, ambiguous interaction restraint; CSP, chemical shift perturbation.

10.1021/jm800820b CCC: $40.75
Published on Web 10/22/2008

Scheme 1. Structure of the Gd-1, Gd-2, and Y-1 Ligands



dressed the cellular uptake and excretion of bile salt-CAs, as outlined above, little is known about the mechanisms of intracellular transport. Small proteins belonging to the family of the intracellular lipid binding proteins are found to be the major bile salt carriers both in the liver and the ileum[22] and are called bile acid binding proteins (BABPs). It is therefore legitimate to consider that bile acid-based molecules such as the above-described CAs may interact with these transporters inside the cell. Chicken liver BABP (cl-BABP) is the most characterized liver BABP. Crystallographic[23] and NMR[24] studies described its fold in terms of a β-clam structure with a two-helical lid and established its ability to bind two bile salt molecules. NMR dynamic and interaction studies further pointed out the relevant role of protein mobility in ligand binding and probable allosteric regulation.[25,26]

Recently, we have reported on the in vitro interaction between a number of related CAs and chicken liver BABP.[21] By use of relaxometric measurements and high resolution NMR, we observed that the protein is able to bind a single ligand molecule with a significantly low dissociation constant, down to 20 μM as observed for compound Gd-1, despite the presence of the non-native bulky chelate moiety linked to the bile acid scaffold. These results prompted us to study the structural details of the above protein−drug interaction to set the basis for the rational design of more efficient hepatospecific CAs. We describe here the NMR solution costructure of chicken liver BABP and compound Gd-1 (cl-BABP/Gd-1), determined by using both paramagnetic structural restraints as well as intra- and intermolecular restraints obtained from experiments on a diamagnetic analogue of the Gd-1 compound. This structure represents the first structure of a BABP with a synthetic drug and is here compared to known complexes with putative native ligands.

Results and Discussion

Protein Resonance Assignments and Protein Fold Determination of Holo cl-BABP. Because the strong paramagnetism of Gd-1 quenches or attenuates all atom resonances in its spatial proximity, NMR experiments for protein resonances assignment were done using the diamagnetic analog Y-1 (Scheme 1), where the gadolinium ion is substituted by an yttrium ion. The two cations are isoelectronic, they have the same charge (+3) and very similar ionic radii (1.053 Å, Gd^{3+} and 1.019 Å, Y^{3+}).[27] It was further observed that chelating compounds such as DOTA and DTPA coordinate the two ions with the same tricapped trigonal prism coordination geometry.[28−30] This substitution is therefore not expected to introduce significant perturbations in the molecular structure and solution dynamics of the complex. The resonance assignment of signals from backbone and side chains atoms of holo cl-BABP was performed as described in the Experimental Section, using standard 3D heteronuclear multidimensional NMR experiments based on through-bond scalar connectivities[31] recorded on a sample of ^{15}N and ^{13}C labeled protein in complex with unlabeled Y-1. The assignment is complete except for amide resonances of M73, N86, K95, and E99, which could not be univocally assigned due to signal overlap. Assignment tables have been deposited in the BioMagResBank.

The secondary structure elements of holo cl-BABP, as derived from chemical shift data with the software TALOS, include 10 antiparallel β-strands and two α-helices. The strands comprise residues 5−12 (strand A), 37−42 (B), 46−53 (C), 57−63 (D), 68−72 (E), 76−85 (F), 88−93 (G), 95−103 (H), 105−112 (I), and 116−123 (J). The α-helices are composed of residues 14−21 (H1) and 26−32 (H2).

The success of docking procedures for the structural determination of protein−ligand complexes is related to the availability of accurate starting protein tertiary structures. Particularly, when the protein undergoes substantial structural rearrangement upon ligand binding, significant improvement of the final complex structure is achieved if the structure of the holo protein is used instead of the one corresponding to the apoprotein. The structures of cl-BABP in complex with cholic acid (CHA) and chenodeoxycholic acid (CDA) were determined previously by X-ray crystallography[23] and NMR spectroscopy,[24] respectively, and appeared to be very similar to each other. However, the structure of BABP when bound to Gd-1 could not be expected to be the same as the above structures, as inferred from the different HSQC spectra of the respective protein complexes. Therefore, the starting "holo" structure for the successive docking calculation was de novo determined. Intramolecular restraints used to determine the protein fold included dihedral angles derived from TALOS and interproton distances obtained from a NOESY-[^1H,^{15}N]-HSQC spectrum as well as NOESY-[^1H,^{13}C]-HSQC spectra optimized for the aliphatic or the aromatic side chains. Structural calculations were performed using CYANA 2.1,[32] implementing an efficient protocol for structure calculation/automated NOE assignment. The 20 structures with the lowest target function were selected to represent the NMR solution structure of the holo protein. The structures were determined with very high precision, as evident from the rmsd values of 0.52 ± 0.12 Å for the backbone, and 1.19 ± 0.10 Å for all heavy atoms. The quality of the structures appeared very good, with 98% of the residues occupying the

Table 1. Statistics for the Final Structural Ensemble Obtained for the Complex BABP/Gd-1

average rmsd (Å)	
⟨rmsd⟩ backbone (C_α, N^H, and C′)[a]	0.52 ± 0.12
⟨rmsd⟩ heavy atoms	1.19 ± 0.10
experimental restraints	
intraresidue NOEs	491
sequential NOEs ($\|i - j\| = 1$)	583
medium-range NOEs ($1 \leq \|i - j\| \leq 4$)	346
long-range NOEs ($\|i - j\| \geq 5$)	767
total number of distance restraints	2187
manually assigned NOEs	150
TALOS dihedral angle restraints	176
experimental restraints violations	
distances with violations[b] > 0.3 Å	3
distances with violations[b] > 0.5 Å	2
dihedral angles with violations[c] > 5°	18
Ramachandran plot (%)	
residues in most favored regions	98
residues in generously allowed regions	1.1
residues in disallowed regions	0.9

[a] Root mean square deviations of atomic coordinates were calculated over residues 1−125 using MOLMOL. [b] Distance restraint violations observed in at least one of the 20 structures. [c] Dihedral angle violations observed in at least one of the 20 structures.

most favored regions of the Ramachandran torsion angle space. The overall statistics for the final structural ensemble is reported in Table 1.

Ligand Resonance Assignment and Experimental Intermolecular Structural Restraints. The stoichiometry of binding of Gd-1 to cl-BABP was established in our previous work as 1:1 based on high resolution NMR as well as relaxometric analyses.[21] The structure determination of the complex was based on the structure obtained for the holo protein as well as the introduction of intermolecular restraints in structural calculations. The chosen structure determination procedure allows for the introduction of ambiguous interaction restraints (AIRs). The latter were obtained from chemical shift perturbations (CSPs) and paramagnetic-induced intensity attenuations of protein amide resonances observed upon ligand binding and measured in a previous work.[21] These experiments did not require ligand resonances assignments. The identified interacting residues were defined as active according to the software's definition, while the adjacent residues were defined as passive (Table 2).

For the observation of unambiguous intermolecular contacts, we analyzed samples composed of a $^{15}N,^{13}C$-labeled protein bound with unlabeled ligand and used editing and filtering techniques to identify the resonances from the two molecular species. Most NMR experiments for the determination of intermolecular restraints were recorded using the diamagnetic analogue Y-1. A 4-fold excess of ligand was used in the experiments because of a measured dissociation constant of about 20 μM and because, from the inspection of the titration curves, the protein backbone amide chemical shift perturbations produced by ligand binding appear to reach a plateau at a protein:ligand ratio of about 1:4. The identification of intermolecular distance restraints was made difficult by the intrinsically poor chemical shift dispersion of protons belonging to the steroid moiety of Y-1, as well as the complex second-order spectra of the metal ion chelating unit. However, the methyl resonances of the steroid moiety, C18, C19, and C21, appear as high-intensity signals at relatively high fields and represented a good

Table 2. Active and Passive Residues of cl-BABP Used to Define the Ambiguous Interaction Restraints within HADDOCK (in Parenthesis Are the Experimental Methods Used to Define the Residue as Active)

active	passive
A10 (CSP[a])	Y9, Q11
E15 (CSP)	Y14, E16
L23* (CSP, attenuation[b], NOE[c])	A22
E25 (attenuation)	P24
D26 (CSP, attenuation)	L27
D33 (attenuation)	R32
T53 (attenuation)	K52
R55 (attenuation)	P54
T57 (attenuation)	V58
T72 (attenuation)	T71
M73 (NOE)	
D74* (NOE)	
G75* (attenuation)	K76

[a] CSP: chemical shift perturbation. [b] Attenuation: signal intensity attenuation in the presence of the paramagnetic ligand. [c] NOE = filtered NOESY data. [d] residues used both as ambiguous and unambiguous restraints.

entry point for the assignment of the bound molecule. Indeed, the methyl resonances from both the free and bound molecules, displaying slow chemical exchange compared to the NMR time scale, are identified from the 2D F_2-[$^{15}N,^{13}C$]-filtered-[$^1H,^1H$]-NOESY spectrum (Figure 1A). Exchange peaks in the spectrum allowed us to unambiguously assign the methyl resonances of the bound molecules starting from the known assignments of the free ligand. The above cross-peaks were confirmed as exchange peaks by running a series of ROESY spectra with varying mixing times because they always displayed the same sign as the diagonal (Figure 1B). In addition to the methyl groups, other unambiguously assigned resonances corresponded to H1, H2, H3, and H4 of the first steroid ring, and to H27, H28, and H29 of the linker between the steroid ring A and the chelating moiety (see Scheme 1 for numbering).

Despite the nonoptimal binding affinity, several intermolecular NOEs were observed in a 3D F_1-[^{13}C]-filtered, F_2-[^{13}C]-separated, F_3-[^{13}C]-edited NOESY-HSQC spectrum. Fourteen intermolecular NOEs were assigned (listed in Table 3), and some of them are shown in Figure 2. They were subsequently used as unambiguous interaction restraints in the complex structure calculation. However, because the filtered experiments do not provide signals corresponding to known distances, no calibration of the peak volumes was performed, and an upper limit distance of 6 Å was introduced. A single intermolecular distance, namely that between A22-Hβ and Y-1-H27, could be obtained directly from cross-peak volume analysis of a NOESY-[$^1H,^{13}C$]-HSQC spectrum after calibration based on known intramolecular distances.

While observation and assignment of intermolecular NOEs allowed us to derive distance restraints between the protein and the steroid moiety of the ligand, no information was obtained from these experiments concerning the metal ion chelating group. Still, it is important to get knowledge about the relative position of this unit that is separated from the steroid scaffold by five rotatable torsion angles. Information derived from paramagnetic centers has long been introduced into structure calculations of metalloproteins, proteins modified to introduce metal chelates or radicals, as well as protein−protein complexes.[33,34] Indeed, paramagnetic centers produce perturbations of protein signals which are distance-dependent and can be of long-range effect.[35] Gadolinium(III) complexes are highly paramagnetic because of the presence of seven unpaired electrons. The high spin multiplicity is maintained also after complexation by common chelating agents,[36,37] and the spherical symmetry of the 4f orbitals explains the absence of chemical shift effects on

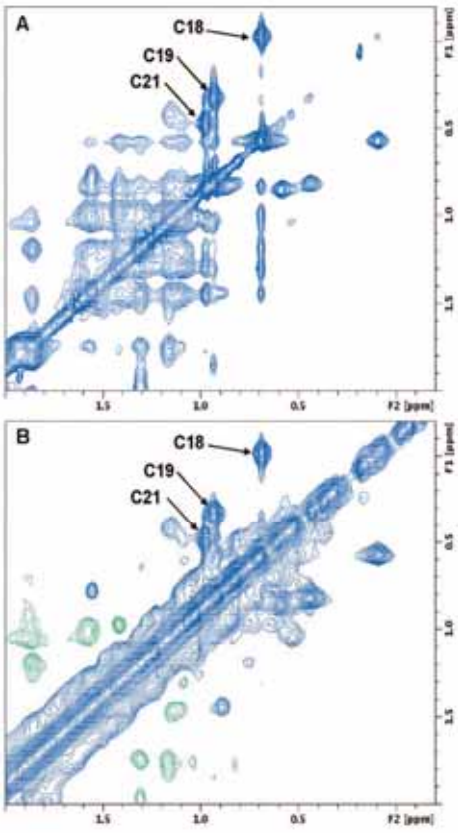


Figure 1. (A) High field region of an F_2-[^{15}N,^{13}C]-filtered-[^{1}H,^{1}H]-NOESY spectrum recorded at 600 MHz, 25 °C, pH 7.0 with 80 ms mixing time. The sample was a [^{15}N/^{13}C] cl-BABP in the presence of a 4-fold excess of [*a*]Y-1. (B) Same region of a [^{1}H,^{1}H]-ROESY experiment recorded in the same conditions as above, with a spin-lock time of 50 ms. Marked peaks are exchange peaks between the proton methyl signals of the free and the bound ligand corresponding to methyl positions C18, C19, and C21.

the signals of nearby nuclei. On the other hand, unpaired electrons markedly affect the nuclear relaxation of neighboring nuclei through dipolar mechanisms. The paramagnetic enhancement of longitudinal relaxation rates of amide protons, R_{1p}, is proportional to the inverse of the sixth power of the metal−proton distance and can thus be exploited to derive structural restraints to place the metal center in the correct position with respect to the protein, once an estimate of the proportionality constant is given. R_{1p} values were here calculated as the difference in R_1 observed between a paramagnetic and a diamagnetic protein−ligand adduct. To avoid excessive reduction in signal/noise due to the strong paramagnetism of gadolinium, the paramagnetic adduct was obtained by dilution of Gd-1 with diamagnetic Y-1. In other words, while keeping a 4-fold excess of total ligand with respect to the protein, 10% of the ligand was Gd-1 while 90% was Y-1. The relaxation rates were measured on amide protons using an inversion recovery scheme coupled to ^{15}N-editing (see Experimental Section for details).[38] Selected signal decays are shown as an example in Figure 3A. The determined R_{1p} rates are plotted as a function of the residue number in Figure 3B and mapped on the protein structure in Figure 3C. Missing values are due to problems of signal overlap or low sensitivity. R_{1p} values up to 2.68 s^{-1} were determined and converted into distances (Table 3). Prior to the rate-to-distance conversion an occupancy factor of 0.1 was introduced (eq 1, Experimental Section), taking into account the reduced paramagnetic effect on protein resonances obtained after 10-fold dilution of Gd-1 by the diamagnetic Y-1. Because the data are redundant, in order to avoid overfitting, only the distances obtained with the lowest errors were retained and introduced into structure calculations (Table 3).

Table 3. Intermolecular Unambiguous Distance Restraints Used to Calculate the Structure of cl-BABP/Gd-1

protein atom	ligand atom	experiment	R_{1p} (s^{-1})[a]	Distance (Å)[b]
L18 Hδ	H-(C18)	NOE[c]		<6.0
L18 Hδ	H-(C21)	NOE		<6.0
K19 H_N	Gd[d]	relaxation[e]	0.47 ± 0.13	15.6
A20 H_N	Gd	relaxation	0.79 ± 0.12	14.4
L21 Hδ	H-(C19)	NOE		<6.0
L21 H_N	Gd	relaxation	1.30 ± 0.22	13.2
A22 Hβ	H-(C3)	NOE		<6.0
A22 H_N	Gd	relaxation	2.07 ± 17	12.2
A22 Hβ	H-(C40)	NOE		<6.0
L23 H_N	Gd	relaxation	1.63 ± 0.19	12.7
L27 Hδ	H-(C40)	NOE		<6.0
L27 Hβ	H-(C19)	NOE		<6.0
L27 H_N	Gd	relaxation	1.65 ± 0.36	12.7
I28 H_N	Gd	relaxation	0.89 ± 0.14	14.1
K29 H_N	Gd	relaxation	0.72 ± 0.15	14.6
M30 Hε	H-(C40)	NOE		<6.0
M30 Hε	H-(C19)	NOE		<6.0
M30 Hβ	H-(C19)	NOE		<6.0
M30 Hγ	H-(C19)	NOE		<6.0
M30 H_N	Gd	relaxation	1.47 ± 0.21	13.0
A31 Hβ	H-(C19)	NOE		<6.0
A31 Hβ	H-(C18)	NOE		<6.0
A31 H_N	Gd	relaxation	0.95 ± 0.16	14.0
R32 H_N	Gd	relaxation	0.33 ± 0.15	16.6
I34 H_N	Gd	relaxation	0.38 ± 0.13	16.2
S51 H_N	Gd	relaxation	0.32 ± 0.16	16.8
T53 H_N	Gd	relaxation	0.63 ± 0.20	14.9
Q56 H_N	Gd	relaxation	2.68 ± 0.61	11.7
D74 H_N	Gd	relaxation	1.81 ± 0.20	12.5
G75 H_N	Gd	relaxation	2.38 ± 0.27	12.0
K76 H_N	Gd	relaxation	1.84 ± 0.44	12.5
K77 H_N	Gd	relaxation	1.70 ± 0.34	12.6
F113 Hβ	H-(C21)	NOE		<6.0

[a] R_{1p}: paramagnetic contribution to the longitudinal relaxation rate. [b] Distance restraints consisted in upper limits if derived from intermolecular NOEs, or in an allowed interval when derived from relaxation measurements (with upper and lower distance bounds set to ± 2.0 Å of the calculated distances). [c] NOE: filtered NOESY data. [d] Gd refers to the metal center. [e] Proton longitudinal relaxation rate measurements.

Structure Calculation and Analysis of the Protein−Ligand Adduct. The calculation of the tertiary structure of the adduct between cl-BABP and Gd-1 was performed with the popular data-driven docking software HADDOCK. Contrary to the most common implementation, in this work the whole set of protein intramolecular restraints were retained during the calculations. This approach, compared to standard protein structure determination methods, has proven successful in similar studies.[39] It is close to a full complex structure determination procedure and ensures highest accuracy of the protein structure in the complex while exploiting the better capability of the docking algorithm to search in the intermolecular conformational space compared to protein structure determination software. The starting structures for ligand docking were thus given by the ten structures determined as described above. An initial ligand conformation was produced on the basis of known crystal-

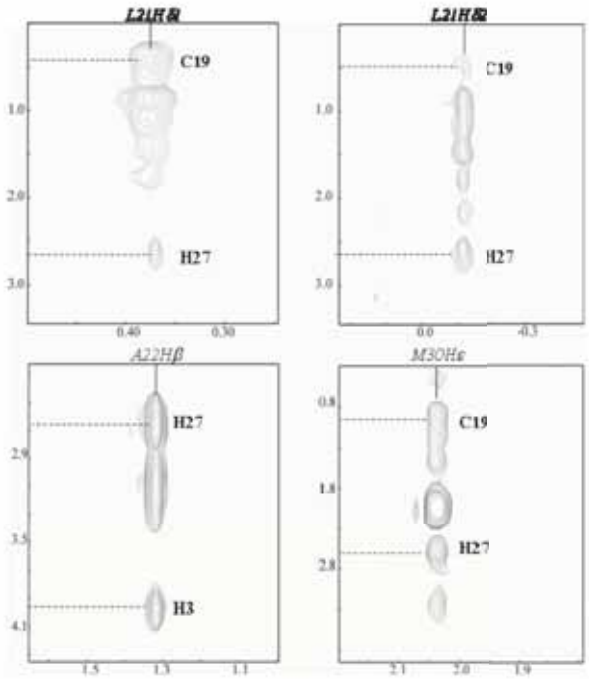


Figure 2. Selected F_1-F_3 slices of the 3D F_1-[^{13}C]-filtered, F_2-[^{13}C]-separated, F_3-[^{15}N,^{13}C]-edited NOESY-HSQC experiment showing intermolecular NOEs. The spectrum was recorded at 25 °C, 900 MHz and with 120 ms mixing time. The sample was [^{15}N/^{13}C] cl-BABP in the presence of a 4-fold excess of [a]Y-1, pH 7.0. Ligand resonances (bold) and protein resonances (italics) are indicated on the spectra.

lographic structures of Gd-DTPA complexes and steroid molecules.[40] HADDOCK ambiguous and unambiguous restraints employed for the calculation of the structure of the complex are reported in Tables 2 and 3, respectively. In many cases the interaction of a protein residue with the ligand was detected with more than one method: ambiguous intermolecular contacts were converted to unambiguous distance restraints if also observed in NOESY or paramagnetic relaxation enhancement experiments.

The HADDOCK structure calculation led to a final ensemble of seven structures displaying backbone and global pairwise rmsd values for protein atom coordinates of 0.66 ± 0.11 Å and 1.42 ± 0.21 Å, respectively, and an rmsd referred to ligand coordinates of 0.88 ± 0.30 Å, after superposition of all atoms. This highly consistent set of seven lowest energy structures is displayed in Figure 4. The family of structures is of good quality in terms of structural statistics and appears to largely satisfy both intramolecular and intermolecular experimentally derived restraints, as evidenced from Table 4. The proton−metal distances measured in the final docked structures (d^{dock}) are in good agreement with the experimental distances (d^{exp}) considering the value of 0.26 ± 0.01 Å assumed by the function

$$\sigma = \frac{1}{N}\sum_{i=1}^{N}\sqrt{(d_i^{dock} - d_i^{exp})^2}$$

with N corresponding to the number of distances and a correlation factor $R = 0.90 \pm 0.01$. The rmsd between the input and the docked protein coordinates is 1.06 ± 0.05 Å, with the largest differences located in helix I. The latter region, which is not very tightly restrained by intramolecular contacts, is more shifted in the direction of the ligand in the docked structure as a consequence of the introduction of six unambiguous intermolecular restraints (between residues 19, 20, 21, 22, and 23 and the metal ion) and one ambiguous restraint (residue 15).

The present structure indicates that the ligand binds to the "upper" part of the large hydrophobic protein cavity, a region called the portal area because it is thought to be the entry site for ligand binding. While the steroid moiety becomes practically buried inside the cavity, the bulky hydrophilic ion chelating moiety is completely at the exterior of the protein. It can be observed that the steroid moiety of the ligand, which is completely apolar due to the absence of hydroxyl groups that are present in bile salts, is surrounded by hydrophobic residues providing significant (25% of total) intermolecular interaction energy (Figure 5A). On the other hand, two positively charged residues, namely R120 and K76, establish polar contacts with the terminal carboxylate deep inside the protein cavity and the amide carbonyl at the opposite extreme, respectively, thereby stabilizing the observed ligand orientation. In a number of structures, another polar contact links an arginine residue, R55, to one carboxylate of the ion chelating group (Figure 5B). Table 5 reports a list of the intermolecular hydrophobic and polar interactions present in the complex.

Structural Comparisons with Known Complexes. High resolution structures of cl-BABP were previously determined,[23,24] showing the capability of the protein to bind two bile acid molecules inside a large hydrophobic cavity with high affinity. Both molecules of either cholic or chenodeoxycholic acid, which are putative native ligands, show an orientation in which ring A is buried within the protein pocket while the carboxylate group

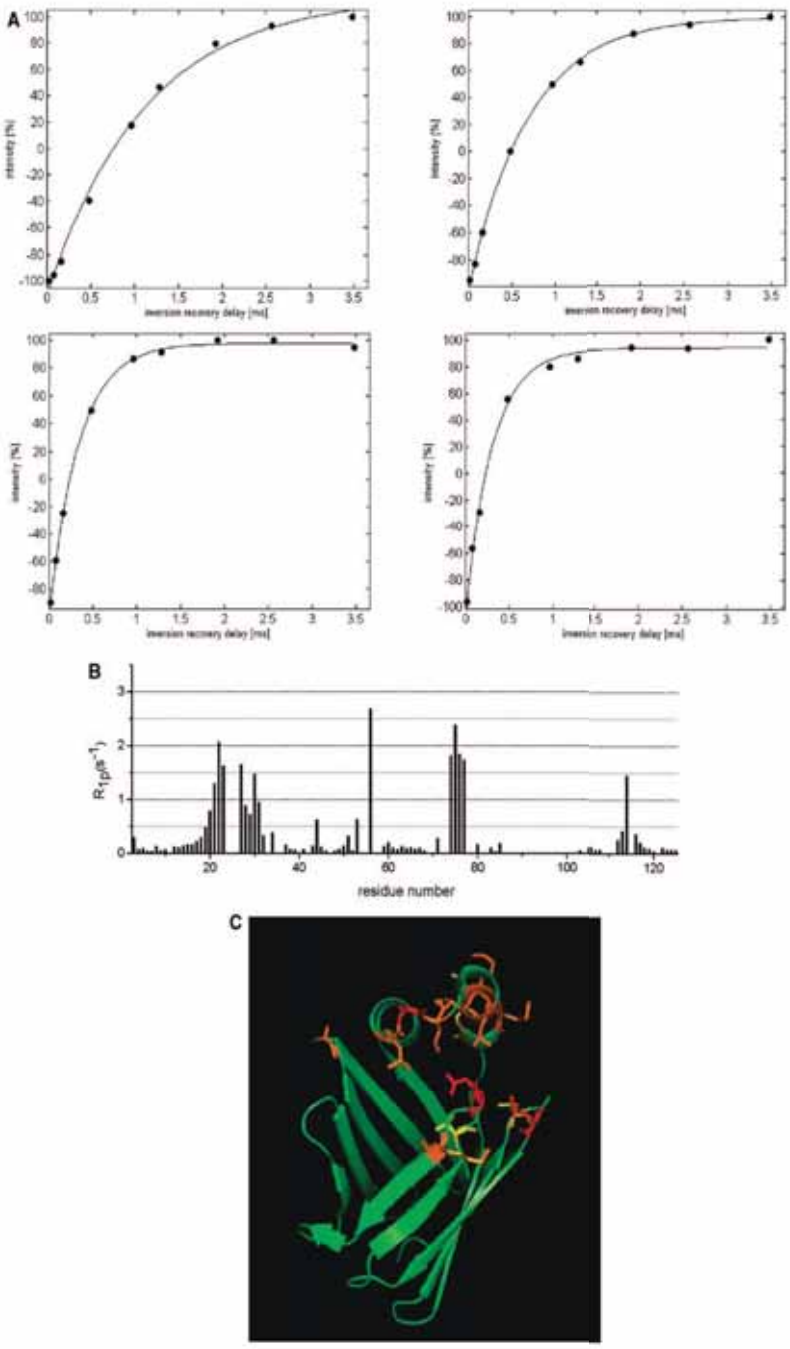


Figure 3. (A) Amide proton signal intensities as a function of increasing inversion recovery delays for residues 22 (left) and 76 (right) highlighting the different longitudinal relaxation rate in the presence of the diamagnetic (top) and paramagnetic (bottom) ligand. Fitting curves obtained with a three parameter fit of a monoexponential function are shown. (B) Paramagnetic enhancements of longitudinal proton relaxation rates as a function of the primary sequence. (C) Paramagnetic enhancements of longitudinal proton relaxation due to the presence of Gd-1 are mapped on the cl-BABP structure and color coded according to the intensity of the enhancement (red: highest attenuation).

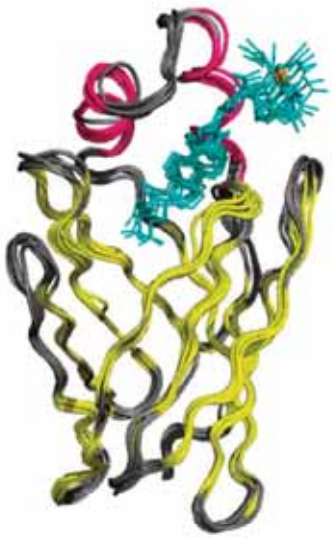


Figure 4. Lowest-energy solution structures calculated for the complex cl-BABP/Gd-1 (PDB: 2k62). The protein in shown in ribbon representation, color coded according to secondary structure (yellow:sheet, magenta:helix). The ligand is represented in cyan sticks, while the metal ion is indicated by an orange sphere.

is fairly solvent-exposed pointing toward the helix−loop−helix motif. The latter is considered the lid that allows entrance of the ligands by its opening movement. The structure determined here differs in several respects from the above complexes. It should be noted that the bulky functionalization introduced on the ligand to carry the paramagnetic center is located at C3 of steroid moiety ring A. As a consequence, Gd-1 accesses the protein cavity in a reverse orientation with respect to bile acids, with ring A pointing toward the open end of the protein pocket (Figure 6). While the protein backbone structures of both holo CHA and CDA complexes are very similar, here we observe conformational changes in correspondence of the protein open end, which exhibits a closure of the loops surrounding the cavity entrance and an opening of the helix−turn−helix lid, which controls the access to the binding cavity. The present data confirm a stoichiometry of binding of 1:1, with Gd-1 occupying mostly the subsite that binds the more external cholic acid. The structure with CHA (1tw4) indicates that five amino acids form hydrogen bonds with the two cholate molecules, namely R55, Q56, K76, H98, and Q100. By analysis of the bundle of NMR structures with CDA (2jn3), the presence of intermolecular hydrogen bonds among the ligands and the residues R55, Q56, T72, M73, D74, K76, K95, and Q100 is found. For the cl-BABP/Gd-1 complex, intermolecular hydrogen bonds with residues R55, K76, and R120 involve the two polar tails of the ligand. The fraction of electrostatic energy contributing to the stabilization of the adducts is larger for the interaction of cl-BABP with CDA compared to Gd-1 (90% and 75% of the total interaction energy, respectively). Thus, new interactions seem to allow for an alternative medium affinity binding when reverse binding is present. Further stabilization could be achieved by appropriately introducing hydroxyl groups on the steroid moiety, considering the availability of H-bond partners along a polar spine, which was recognized as highly conserved in an earlier study on fatty acid binding proteins.[41]

Table 4. Structural Statistics of the Seven Best Structures of cl-BABP/Gd-1

average rmsd (Å)	
⟨rmsd⟩ backbone(C_α, N_H, and C′)[a]	0.66 ± 0.11
⟨rmsd⟩ heavy atoms	1.42 ± 0.21
backbone rmsd (Å) with respect to mean	
flexible interface backbone	0.44 ± 0.17
all backbone	0.48 ± 0.21
experimental restrains	
number of interaction restraints (AIRs)	
total AIRs	26
number of unambigous interaction restraints	
intermolecular restraints	33
restraints violations	
total AIR violations	2
total unambigous violations[b]	1
intermolecular energies after water refinement	
E_{vdw} (kcal mol−1)	−28.2 ± 2.7
E_{elec} (kcal mol−1)	−87.2 ± 29.9
buried surface area (Å^2)	972.4 ± 42.0
rmsd from idealized covalent geometry	
bonds (Å)	0.004 ± 0.00
angles (deg)	0.6 ± 0.02
impropers (deg)	0.55 ± 0.01
dihedrals (deg)	22.4 ± 0.5
Ramachandran plot (%)	
residues in most favored regions	98.3
residues in generously allowed regions	0.8
residues in disallowed regions	0.9

[a] Root mean square deviations of atomic coordinates were calculated over residues 1−125 using MOLMOL. [b] Distance violated in more than 30% of the structures of the final bundle.

The structural comparison of cl-BABP/Gd-1 with the other available liver BABP structures in complex with their putative ligands bears some relevance with respect to the understanding of the determinants of binding in this protein family. Few other structures of holo BABPs are presently available, all of them referring to liver proteins from nonmammalian species. We calculated and compared the positional global rmsd per residue between the cl-BABP/Gd-1 structure and the other reported holo complexes from chicken, zebrafish, and axolotl liver (Figure 7). It is clear from this comparison that the protein regions 18−29 (loop connecting the two helices and helix2), 42−46 (BC loop), 55−56 (CD loop), 84−88 (G strand), and 113−115 (IJ loop) in cl-BABP/Gd-1 display the largest structural differences to all other holo structures. This observation is consistent with the high plasticity at the level of the portal area, considered relevant for ligand binding. On the other hand, large differences are observed also in the opposite side of the cavity at the level of stretch 84−88. The latter comprises some highly conserved residues, playing a structural role in protecting the cavity from solvent. At first sight, this difference might be related to the fact that, in the cl-BABP/Gd-1 structure, the ligand does not approach the bottom of the cavity, contrary to other ternary complexes. Interestingly, however, large differences are observed with the only other singly ligated protein, namely

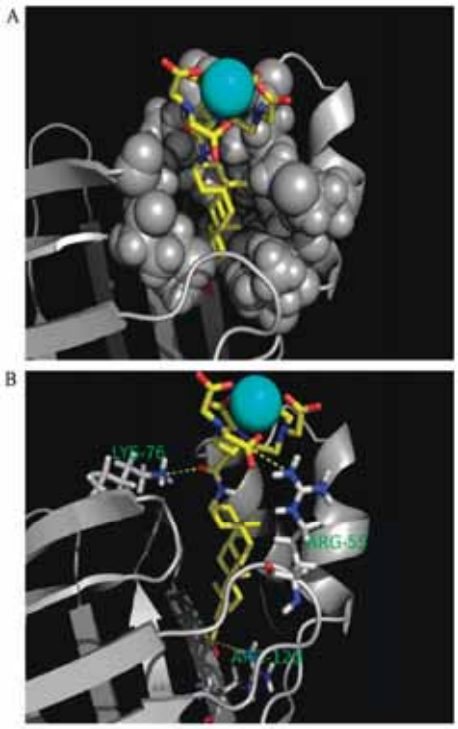


Figure 5. Hydrophobic (A) and polar (B) intermolecular contacts are highlighted on the cl-BABP/Gd-1 structure.

Table 5. Hydrophobic Interactions, H-Bonds, and Salt Bridges Present in at Least Four of the Best Seven Structures of cl-BABP/Gd-1 Calculated by HADDOCK

H-bond/salt bridges	hydrophobic interactions
K76 (F strand)	Y14 (helix I)
R120 (J strand)	F17 (helix I)
	L18 (helix I)
	L21 (helix I)
	A22 (helix1−helix II Loop)
	L23 (helix I−helix II Loop)
	M30 (helix II)
	A31 (helix II)
	L34 (helix II−B strand loop)
	M73 (EF loop)
	D74 (EF loop)
	L118 (J strand)

zebrafish liver, thus indicating that any opening motion occurring at the portal area is transmitted to the opposite side of the protein.

Conclusions

The solution structure of a liver BABP in complex with Gd-1 has been obtained using the HADDOCK procedure based on the determined NMR structure of the holo protein and both ambiguous and unambiguous intermolecular restraints derived from a variety of NMR experiments. The structure of the cl-BABP/Gd-1 complex represents the starting point for the design of new and more efficient CAs. The observations derived from structural comparison with holo proteins in complex with native ligands underline the capability of this protein scaffold to adapt

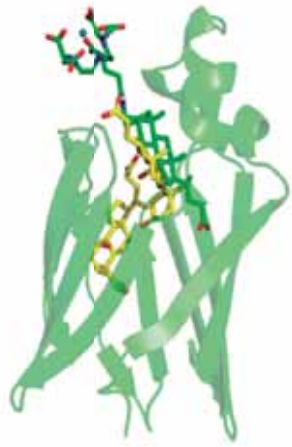


Figure 6. Structure of cl-BABP/Gd-1 (PDB: 2k62) with superimposed CDA ligands as in the ternary complex cl-BABP/CDA (PDB: 2jn3). The protein structure is shown in a green cartoon representation and the ligand Gd-1 in green sticks. The position of the CDA ligands (yellow) is obtained after global superimposition of the two complex structures.

itself to the steroid moiety, thus making this system an appropriate carrier for hepatospecific contrast agents. The analysis of the hydrogen bonding pattern in the available cl-BABP complexes suggests that a further design could take into account the introduction, on the steroid moiety, of hydroxyl groups, which play some important role in fine-tuning polar interactions with conserved polar residues, always involved in ligand binding.

Experimental Section

Protein Expression and Purification. Recombinant cl-BABP was expressed as soluble protein in *Escherichia coli* cells, purified with anion exchange and size-exclusion chromatography, and finally delipidated following already described protocols.[25] Uniform ^{15}N and ^{13}C isotope labeling was achieved using M9 minimal media containing 1 g/L $^{15}NH_4Cl$ and 4 g/L ^{13}C enriched glucose.

Ligand Synthesis. The bile acid conjugate was prepared as previously described[6] and kindly supplied by Bracco Imaging Spa (Milan).

NMR Experiments and Data Analysis. All samples were prepared in 30 mM sodium phosphate buffer, pH 7.0, containing 0.03% NaN_3 and 95%/5% H_2O/D_2O or 99% D_2O. NMR spectra were acquired at 25 °C on Bruker spectrometers Avance III 600 and Avance 900, both equipped with a 5 mm TCI cryoprobe and Z-field gradient. Data were processed with NMRPipe[42] and visualized by NMRView.[43]

The following experiments were run on the adduct between the [$^{15}N/^{13}C$]cl-BABP and unlabeled diamagnetic yttrium analogue [*u*]Y-1 (Scheme 1). A protein:ligand ratio of 1:4 was used with a final protein concentration of 0.5 mM. The sequential amino acid resonances assignment has been previously described,[21] and was here completed through the analysis of HNCACB, CBCA(CO)NH, HN(CA)CO, and HNCO triple resonance experiments collected at 600 MHz. Side chain carbon and proton resonances were assigned based on (H)CCH-TOCSY, H(C)CH-TOCSY, and HBHA(CO)NH spectra. The resonance assignments were submitted to the BioMagResBank under accession number 15854. Intramolecular NOE-based distance restraints were extracted from 3D NOESY-[$^1H,^{15}N$]-HSQC, NOESY-[$^1H,^{13}C$]-HSQC in H_2O, as well as NOESY-[$^1H,^{13}C$]-HSQC and NOESY-[$^1H,^{13}C$]-HSQC spectra in D_2O optimized for the aromatic side chains, all acquired at 600 MHz.

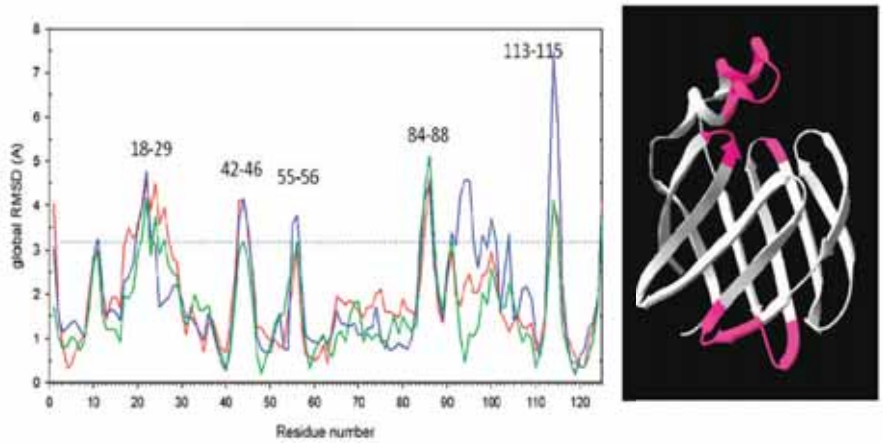


Figure 7. Left panel: average global rmsd between the backbone atoms of the cl-BABP/Gd-1 complex (PDB 2k62) and cl-BABP in complex with CDA (red, PDB 2jn3), liver zebrafish BABP/CHA (blue, PDB 2qo4), liver axolotl BABP/CHA (green PDB 2ft9), reported as a function of residue number. The rmsd values were computed for each couple of structures by using the program MOLMOL. Regions of cl-BABP/Gd-1 complex exhibiting rmsd larger than 3 Å from all the other analyzed structures are indicated in the plot in the left panel and mapped in pink on the protein structure in the right panel.

Intermolecular NOEs were obtained from a F_2-[^{15}N,^{13}C]-filtered-[^{1}H,^{1}H]-NOESY and a NOESY-[^{1}H,^{13}C]-HSQC spectra acquired at 600 MHz, as well as from a 3D F_1-[^{13}C]-filtered, F_2-[^{13}C]-separated, F_3-[^{15}N,^{13}C]-edited NOESY-HSQC spectrum acquired at 900 MHz. Mixing times of 80, 120, or 150 ms were used in all NOESY experiments. Finally, 2D [^{1}H,^{1}H]-ROESY experiments were performed with 50 and 90 ms spin-locking time. To obtain the maximum sensitivity, no filter was applied to this experiment. Pulse programs from the standard Bruker library were used.

Additional intermolecular restraints were derived from relaxation measurements on the adduct between a ^{15}N cl-BABP complexed with both the unlabeled diamagnetic Y-1 and paramagnetic Gd-1 in a P:Y-1:Gd-1 ratio of 1:4.5:0.5. In this way, it was possible to observe a fully saturated protein while keeping the concentration of the paramagnetic ligand sufficiently low, thereby avoiding bleaching out of amide signals. Amide proton longitudinal relaxation times were measured with an inversion recovery−HSQC sequence.[38] R_1 values were obtained by collecting a series of 10 spectra with delays between 0.08 and 3.48 s. The same experiments were repeated on a fully diamagnetic sample containing Y-1 with a P:L ratio of 1:5. Peak intensities were fitted to a monoexponential decaying function using three adjustable parameters within the program RELAXFIT.[44]

Protein Structure Calculation. Structure calculations were performed by simulated annealing in torsion angle space using the CYANA 2.1 package,[32] which implements an efficient protocol for structure calculation/automated NOE assignment.[45] The peak lists in CYANA format were generated by Format Converter.[46] The standard annealing protocol was used with 10000 steps of torsion angle dynamics; in each of the seven cycles, 100 structures were calculated and the 20 with the lowest target function were used in the next cycle. Then 150 NOEs were assigned manually and imposed in all stages of the protocol while all the other NOEs were automatically assigned, resulting in a total of 2187 upper distance bounds. Dihedral angle restraints were derived from TALOS.[47] In the last cycle, the 20 structures with the lowest target function were selected as the final bundle. The 10 best structures of the final bundle were used for the docking procedure.

Protein−Ligand Costructure Determination. The structure of the cl-BABP/Gd-1 adduct was obtained by data-driven docking using the software HADDOCK 2.0[48] in combination with CNS.[49] An ensemble of 10 NMR structures of the holo protein calculated by CYANA was used as input for the docking calculation. The starting coordinates of Gd-1 were obtained with the program SMILE.[50] The ligand molecule was built in complex with Yb(III), which is the metal ion most similar to Gd(III) in terms of atomic radius and mass among the ions currently defined in HADDOCK. A number of 11 freely rotatable dihedral angles were defined for the cholanoate and the linker moieties, while the DTPA residue was defined as rigid.

The definition of the "active and passive residues" was based on the experimental data collected as previously described,[24] including the chemical shift perturbations (CSPs) upon ligand binding and paramagnetic attenuation of amide resonances due to the presence of Gd(III). These data were used as ambiguous interaction restraints (AIR) to guide the docking. Unambiguous interaction restraints were further introduced in the calculations, including 14 intermolecular NOE-derived distances obtained from the analysis of the 3D edited/filtered NOESY, as well as 19 distances derived from paramagnetic relaxation enhancements. Paramagnetic contributions to longitudinal relaxation of amide protons, R_{1p}, were calculated as the difference in R_1 observed between the paramagnetic and the diamagnetic adduct and then converted into distance restraints according to the following equation:

$$R_{1p} = f \times K_{dip}/r^6 \quad (1)$$

where f is an occupancy factor and K_{dip} is the sum of the Solomon and Curie constants.[51] On the basis of estimates of the electronic relaxation times of the metal ion,[38] the lifetime of the protein−ligand adduct,[52] and the overall rotational correlation time,[25] a K_{dip} value of 2.3×10^7 s^{-1} Å^6, was determined. The occupancy factor was set to 0.1, corresponding to the paramagnetic/diamagnetic compound ratio in the sample. The errors in the relaxation rates were obtained from the fitting of the decay curves using a standard Monte Carlo approach. All the parameters used to estimate K_{dip} contain potential sources of error, but even a total uncertainty of 10% in the value of K_{dip} translates into less than 2% error in the calculated distance. However, instead of estimating the uncertainty in each distance, the approach taken was to assume that the errors from all sources can best be dealt with by adjusting the uncertainty through the distance bounds. On the basis of a number of structure calculation tests and evaluation of the penalty functions, the chosen bounds were ±2 Å.

The final upper distance limits used to calculate the holo protein structure in CYANA were used as unambiguous interaction restraints in order to ensure that the protein structure in the complex satisfied the experimentally derived intramolecular distances. Finally, TALOS restraints and hydrogen bond restraints derived from the CYANA structures were also used to calculate the structure of the complex by HADDOCK.

The HADDOCK docking protocol consists of three steps: (i) randomization of orientation and rigid body minimization, (ii) semirigid simulated annealing in torsion angle space, (iii) final refinement in Cartesian space with explicit solvent. The rigid body docking step was performed five times, with 1000 structures generated at each stage, the best 200 of which were refined in the semiflexible stage and subsequently in explicit water. Electrostatic and van der Waals terms were calculated with a 8.5 Å distance cutoff using the OPLS nonbonded parameters from the paral-lhdg5.3.pro parameter file.[53,54] Topology and parameters of the ligand were calculated using PRODRG server.[55] The dielectric constant ε was set to 10.0 and to 1.0 during explicit solvent refinement. The ligand was kept rigid during the semiflexible stage. The resulting solutions were clustered using the algorithm of Daura et al.[56] with 1 Å cutoff on the basis of the pairwise rmsd matrix for the carbon atoms of the ligand after fitting on the backbone atoms of the protein. The structures were divided in nine clusters ranked according to the HADDOCK score. We selected the best seven structures for every cluster and defined a final bundle of seven structures according to the HADDOCK score. The protein−ligand contacts were analyzed using the software LIGPLOT,[57] and the rmsd referred to ligand coordinates (carbons) was calculated after global superposition of protein backbone using the PROFIT program (Martin, A. C. R., Profit http://www.bioninf.org.uk/software/profit). The final ensemble of structures has been deposited in the PDB (2k62).

Acknowledgment. This research was supported by FIRB 2003 (grant RBNE03PX83_006) and by Cariverona Foundation. Consorzio Interuniversitario di Risonanze Magnetiche di Metalloproteine Paramagnetiche (CIRMMP) is gratefully acknowledged for measurements performed at 900 MHz. We are indebted to Alexandre Bonvin (Utrecht University) for helpful suggestions in the docking procedures and to Lucia Zetta for helpful suggestions. Clelia Cogliati is acknowledged for technical help in the assignment procedure.

References

(1) Kanematsu, M.; Kondo, H.; Goshima, S.; Kato, H.; Tsuge, U.; Hirose, Y.; Kim, M. J.; Moriyama, N. Imaging liver metastases: review and update. *Eur. J. Radiol.* **2006**, *58*, 217–228.

(2) Runge, V. M.; Pels Rijcken, T. H.; Davidoff, A.; Wells, J. W.; Stark, D. D. Contrast-enhanced MR imaging of the liver. *J. Magn. Reson. Imaging* **1994**, *4*, 281–289.

(3) Cabella, C.; Crich, S. G.; Corpillo, D.; Barge, A.; Ghirelli, C.; Bruno, E.; Lorusso, V.; Uggeri, F.; Aime, S. Cellular labeling with Gd(III) chelates: only high thermodynamic stabilities prevent the cells acting as "sponges" of Gd^{3+} ions. *Contrast Media Mol. Imaging* **2006**, *1*, 23–29.

(4) Vlasie, M. D.; Comuzzi, C.; van den Nieuwendijk, A. M.; Prudencio, M.; Overhand, M.; Ubbink, M. Long-range-distance NMR effects in a protein labeled with a lanthanide-DOTA chelate. *Chemistry* **2007**, *13*, 1715–1723.

(5) Chong, H. S.; Garmestani, K.; Bryant, L. H., Jr.; Milenic, D. E.; Overstreet, T.; Birch, N.; Le, T.; Brady, E. D.; Brechbiel, M. W. Synthesis and evaluation of novel macrocyclic and acyclic ligands as contrast enhancement agents for magnetic resonance imaging. *J. Med. Chem.* **2006**, *49*, 2055–2062.

(6) Anelli, P. L.; Lattuada, L.; Lorusso, V.; Lux, G.; Morisetti, A.; Morosini, P.; Serleti, M.; Uggeri, F. Conjugates of gadolinium complexes to bile acids as hepatocyte-directed contrast agents for magnetic resonance imaging. *J. Med. Chem.* **2004**, *47*, 3629–3641.

(7) Lu, Z. R.; Wang, X.; Parker, D. L.; Goodrich, K. C.; Buswell, H. R. Poly(L-glutamic acid) Gd(III)-DOTA conjugate with a degradable spacer for magnetic resonance imaging. *Bioconjugate Chem.* **2003**, *14*, 715–719.

(8) Curtet, C.; Maton, F.; Havet, T.; Slinkin, M.; Mishra, A.; Chatal, J. F.; Muller, R. N. Polylysine-Gd-DTPAn and polylysine-Gd-DOTAn coupled to anti-CEA F(ab′)2 fragments as potential immunocontrast agents. Relaxometry, biodistribution, and magnetic resonance imaging in nude mice grafted with human colorectal carcinoma. *Invest. Radiol.* **1998**, *33*, 752–761.

(9) Planchamp, C.; Montet, X.; Frossard, J. L.; Quadri, R.; Stieger, B.; Meier, P. J.; Ivancevic, M. K.; Vallee, J. P.; Terrier, F.; Pastor, C. M. Magnetic resonance imaging with hepatospecific contrast agents in cirrhotic rat livers. *Invest. Radiol.* **2005**, *40*, 187–194.

(10) Marinelli, E. R.; Neubeck, R.; Song, B.; Wagler, T.; Ranganathan, R. S.; Sukumaran, K.; Wedeking, P.; Nunn, A.; Runge, V.; Tweedle, M. Synthesis and evaluation of macrocyclic gadolinium chelates as hepatospecific MRI agents. *Acad. Radiol.* **2002**, *9* (1), S251–S254.

(11) Vander Elst, L.; Chapelle, F.; Laurent, S.; Muller, R. N. Stereospecific binding of MRI contrast agents to human serum albumin: the case of Gd-(S)-EOB-DTPA (Eovist) and its (*R*) isomer. *J. Biol. Inorg. Chem.* **2001**, *6*, 196–200.

(12) Pascolo, L.; Petrovic, S.; Cupelli, F.; Bruschi, C. V.; Anelli, P. L.; Lorusso, V.; Visigalli, M.; Uggeri, F.; Tiribelli, C. ABC protein transport of MRI contrast agents in canalicular rat liver plasma vesicles and yeast vacuoles. *Biochem. Biophys. Res. Commun.* **2001**, *282*, 60–66.

(13) Anelli, P. L.; Calabi, L.; de Haen, C.; Lattuada, L.; Lorusso, V.; Maiocchi, A.; Morosini, P.; Uggeri, F. Hepatocyte-directed MR contrast agents. Can we take advantage of bile acids? *Acta. Radiol. Suppl.* **1997**, *412*, 125–133.

(14) Uggeri, J.; Belletti, S.; Bussolati, O.; Dall'Asta, V.; Gazzola, G. C. Suppression of anionic amino acid transport impairs the maintenance of intracellular glutamate in Ha-ras-expressing cells. *Biochem. Biophys. Res. Commun.* **1995**, *211*, 878–884.

(15) Schmitt-Willich, H.; Brehm, M.; Ewers, C. L.; Michl, G.; Muller-Fahrnow, A.; Petrov, O.; Platzek, J.; Raduchel, B.; Sulzle, D. Synthesis and Physicochemical Characterization of a New Gadolinium Chelate: The Liver-Specific Magnetic Resonance Imaging Contrast Agent Gd-EOB-DTPA. *Inorg. Chem.* **1999**, *38*, 1134–1144.

(16) Hamm, B.; Staks, T.; Muhler, A.; Bollow, M.; Taupitz, M.; Frenzel, T.; Wolf, K. J.; Weinmann, H. J.; Lange, L. Phase I clinical evaluation of Gd-EOB-DTPA as a hepatobiliary MR contrast agent: safety, pharmacokinetics, and MR imaging. *Radiology* **1995**, *195*, 785–792.

(17) Pascolo, L.; Cupelli, F.; Anelli, P. L.; Lorusso, V.; Visigalli, M.; Uggeri, F.; Tiribelli, C. Molecular mechanisms for the hepatic uptake of magnetic resonance imaging contrast agents. *Biochem. Biophys. Res. Commun.* **1999**, *257*, 746–752.

(18) Brasch, R. C. New directions in the development of MR imaging contrast media. *Radiology* **1992**, *183*, 1–11.

(19) Dawson, P. A.; Oelkers, P. Bile acid transporters. *Curr. Opin. Lipidol.* **1995**, *6*, 109–114.

(20) von Dippe, P.; Levy, D. Reconstitution of the immunopurified 49 kDa sodium-dependent bile acid transport protein derived from hepatocyte sinusoidal plasma membranes. *J. Biol. Chem.* **1990**, *265*, 14812–14816.

(21) Assfalg, M.; Gianolio, E.; Zanzoni, S.; Tomaselli, S.; Russo, V. L.; Cabella, C.; Ragona, L.; Aime, S.; Molinari, H. NMR structural studies of the supramolecular adducts between a liver cytosolic bile acid binding protein and gadolinium(III)-chelates bearing bile acids residues: molecular determinants of the binding of a hepatospecific magnetic resonance imaging contrast agent. *J. Med. Chem.* **2007**, *50*, 5257–5268.

(22) Guariento, M.; Raimondo, D.; Assfalg, M.; Zanzoni, S.; Pesente, P.; Ragona, L.; Tramontano, A.; Molinari, H. Identification and functional characterization of the bile acid transport proteins in nonmammalian ileum and mammalian liver. *Proteins* **2008**, *70*, 462–472.

(23) Nichesola, D.; Perduca, M.; Capaldi, S.; Carrizo, M. E.; Righetti, P. G.; Monaco, H. L. Crystal structure of chicken liver basic fatty acid-binding protein complexed with cholic acid. *Biochemistry* **2004**, *43*, 14072–14079.

(24) Eliseo, T.; Ragona, L.; Catalano, M.; Assfalg, M.; Paci, M.; Zetta, L.; Molinari, H.; Cicero, D. O. Structural and dynamic determinants of ligand binding in the ternary complex of chicken liver bile acid binding protein with two bile salts revealed by NMR. *Biochemistry* **2007**, *46*, 12557–12567.

(25) Ragona, L.; Catalano, M.; Luppi, M.; Cicero, D.; Eliseo, T.; Foote, J.; Fogolari, F.; Zetta, L.; Molinari, H. NMR dynamic studies suggest that allosteric activation regulates ligand binding in chicken liver bile acid-binding protein. *J. Biol. Chem.* **2006**, *281*, 9697–9709.

(26) Tomaselli, S.; Ragona, L.; Zetta, L.; Assfalg, M.; Ferranti, P.; Longhi, R.; Bonvin, A. M.; Molinari, H. NMR-based modeling and binding studies of a ternary complex between chicken liver bile acid binding protein and bile acids. *Proteins* **2007**, *69*, 177–191.

(27) Shannon, R. D.; Previtt, C. T. Effective ionic radii in oxides and fluorides. *Acta Crystallogr., Sect. B: Struct. Sci.* **1969**, *25*, 925–946.

(28) Hsieh, W. Y.; Liu, S. Synthesis, characterization, and structures of indium In(DTPA-BA2) and yttrium Y(DTPA-BA2)(CH3OH) complexes (BA = benzylamine): models for 111In- and 90Y-labeled DTPA-biomolecule conjugates. *Inorg. Chem.* **2004**, *43*, 6006–6014.
(29) Kumar, K.; Chang, C. A.; Francesconi, L. C.; Dischino, D. D.; Malley, M. F.; Gougoutas, J. Z.; Tweedle, M. F. Synthesis, Stability, and Structure of Gadolinium(III) and Yttrium(III) Macrocyclic Poly(amino carboxylates). *Inorg. Chem.* **1994**, *33*, 3567–3575.
(30) Uggeri, F.; Aime, S.; Anelli, P. L.; Botta, M.; Brocchetta, M.; de Haeen, C.; Ermondi, G.; Grandi, M.; Paoli, P. Novel Contrast Agents for Magnetic Resonance Imaging. Synthesis and Characterization of the Ligand BOPTA and Its Ln(III) Complexes (Ln = Gd, La, Lu). X-ray Structure of Disodium (TPS-9-145337286-C-S)-[4-Carboxy-5,8,11-tris(carboxymethyl)-1-phenyl-2-oxa-5,8,11-triazatridecan-13-oato(5-)]gadolinate(2−) in a Mixture with Its Enantiomer. *Inorg. Chem.* **1995**, *34*, 633–643.
(31) Blomberg, N.; Sattler, M.; Nilges, M. 1H, 15N, and 13C resonance assignment of the PH domain from *C. elegans* UNC-89. *J. Biomol. NMR* **1999**, *15*, 269–70.
(32) Guntert, P. Automated NMR structure calculation with CYANA. *Methods Mol. Biol.* **2004**, *278*, 353–378.
(33) Kosen, P. A. Spin labeling of proteins. *Methods Enzymol.* **1989**, *177*, 86–121.
(34) Bertini, I.; Donaire, A.; Luchinat, C.; Rosato, A. Paramagnetic relaxation as a tool for solution structure determination: Clostridium pasteurianum ferredoxin as an example. *Proteins* **1997**, *29*, 348–358.
(35) Bertini, I.; Luchinat, C.; Rosato, A. The solution structure of paramagnetic metalloproteins. *Prog. Biophys. Mol. Biol.* **1996**, *66*, 43–80.
(36) Caravan, P.; Ellison, J. J.; McMurry, T. J.; Lauffer, R. B. Gadolinium(III) Chelates as MRI Contrast Agents: Structure, Dynamics, and Applications. *Chem. Rev.* **1999**, *99*, 2293–352.
(37) Aime, S.; Fasano, M.; Terreno, E. Lanthanide(III) chelates for NMR biomedical applications. *Chem. Soc. Rev.* **1998**, *27*, 19–29.
(38) Aime, S.; D'Amelio, N.; Fragai, M.; Lee, Y. M.; Luchinat, C.; Terreno, E.; Valensin, G. A paramagnetic probe to localize residues next to carboxylates on protein surfaces. *J. Biol. Inorg. Chem.* **2002**, *7*, 617–622.
(39) Assfalg, M.; Bertini, I.; Del Conte, R.; Giachetti, A.; Turano, P. Cytochrome c and organic molecules: solution structure of the *p*-aminophenol adduct. *Biochemistry* **2007**, *46*, 6232–6238.
(40) Benazeth, S.; Purans, J.; Chalbot, M. C.; Nguyen-Van-Duong, M. K.; Nicolas, L.; Keller, F.; Gaudemer, A. Temperature and pH Dependence XAFS Study of Gd(DOTA)(−) and Gd(DTPA)(2)(−) Complexes: Solid State and Solution Structures. *Inorg. Chem.* **1998**, *37*, 3667–3674.
(41) Ragona, L.; Colombo, G.; Catalano, M.; Molinari, H. Determinants of protein stability and folding: comparative analysis of beta-lactoglobulins and liver basic fatty acid binding protein. *Proteins* **2005**, *61*, 366–376.
(42) Delaglio, F.; Grzesiek, S.; Vuister, G. W.; Zhu, G.; Pfeifer, J.; Bax, A. NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J. Biomol. NMR* **1995**, *6*, 277–293.
(43) Johnson, B. A. Using NMRView to visualize and analyze the NMR spectra of macromolecules. *Methods Mol. Biol.* **2004**, *278*, 313–352.
(44) Fushman, D.; Cahill, S.; Cowburn, D. The main-chain dynamics of the dynamin pleckstrin homology (PH) domain in solution: analysis of 15N relaxation with monomer/dimer equilibration. *J. Mol. Biol.* **1997**, *266*, 173–194.
(45) Herrmann, T.; Guntert, P.; Wuthrich, K. Protein NMR structure determination with automated NOE assignment using the new software CANDID and the torsion angle dynamics algorithm DYANA. *J. Mol. Biol.* **2002**, *319*, 209–227.
(46) Fogh, R.; Ionides, J.; Ulrich, E.; Boucher, W.; Vranken, W.; Linge, J. P.; Habeck, M.; Rieping, W.; Bhat, T. N.; Westbrook, J.; Henrick, K.; Gilliland, G.; Berman, H.; Thornton, J.; Nilges, M.; Markley, J.; Laue, E. The CCPN project: an interim report on a data model for the NMR community. *Nat. Struct. Biol.* **2002**, *9*, 416–418.
(47) Cornilescu, G.; Delaglio, F.; Bax, A. Protein backbone angle restraints from searching a database for chemical shift and sequence homology. *J. Biomol. NMR* **1999**, *13*, 289–302.
(48) Dominguez, C.; Boelens, R.; Bonvin, A. M. HADDOCK: a protein−protein docking approach based on biochemical or biophysical information. *J. Am. Chem. Soc.* **2003**, *125*, 1731–1737.
(49) Brunger, A. T.; Adams, P. D.; Clore, G. M.; DeLano, W. L.; Gros, P.; Grosse-Kunstleve, R. W.; Jiang, J. S.; Kuszewski, J.; Nilges, M.; Pannu, N. S.; Read, R. J.; Rice, L. M.; Simonson, T.; Warren, G. L. Crystallography & NMR system: A new software suite for macromolecular structure determination. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **1998**, *54*, 905–921.
(50) Eufri, D.; Sironi, A. SMILE−shaded molecular imaging on low-cost equipment. *J. Mol. Graph.* **1989**, *7*, 165–169, 158.
(51) Bertini, I.; Luchinat, C. NMR of paramagnetic substances. *Coord. Chem. Rev.* **1996**, *150*, 1–296.
(52) Ayers, S. D.; Nedrow, K. L.; Gillilan, R. E.; Noy, N. Continuous nucleocytoplasmic shuttling underlies transcriptional activation of PPARgamma by FABP4. *Biochemistry* **2007**, *46*, 6744–6752.
(53) Jorgensen, W. L.; Tirado-Rives, J. The OPLS (optimized potentials for liquid simulations) potential functions for proteins, energy minimizations for crystals of cyclic peptides and crambin. *J. Am. Chem. Soc.* **1988**, *110*, 1657–1666.
(54) Linge, J. P.; Williams, M. A.; Spronk, C. A.; Bonvin, A. M.; Nilges, M. Refinement of protein structures in explicit solvent. *Proteins* **2003**, *50*, 496–506.
(55) van Aalten, D. M.; Bywater, R.; Findlay, J. B.; Hendlich, M.; Hooft, R. W.; Vriend, G. PRODRG, a program for generating molecular topologies and unique molecular descriptors from coordinates of small molecules. *J. Comput.-Aided Mol. Des.* **1996**, *10*, 255–262.
(56) Daura, X.; Gademan, K.; Jaun, B.; Seebach, D.; van Gunsteren, W. F.; Mark, A. Peptide folding: when simulation meets experiment. *Angew. Chem., Int. Ed.* **1999**, *38*, 236–240.
(57) Wallace, A. C.; Laskowski, R. A.; Thornton, J. M. LIGPLOT: a program to generate schematic diagrams of protein−ligand interactions. *Protein Eng.* **1995**, *8*, 127–134.

JM800820B

Chapter 4: Chicken ileal bile-acid-binding protein: a promising target of investigation to understand binding cooperativity across the protein family

Short presentation of publication 2

Bile acids, besides their importance in physiological processes, such as cholesterol homeostasis and enterohepatic circulation, have been recognized to be signaling molecules. Activation of nuclear receptors and cell signaling pathways by bile acids alters the expression of numerous genes encoding enzyme and proteins involved in the regulation of bile acids, lipids, glucose and energy metabolism.

This work addresses the study and proposes a mechanism of interaction between bile acids and chicken ileal bile acid binding protein (chicken I-BABP). Although several data have been reported addressing the interaction of different BABPs with bile acids, the binding mechanism is still poorly understood. Indeed BABPs from different species or tissues show high sequence and structural homology but they bind bile acid molecules with different stoichiometry, selectivity and cooperativity. In order to better understand the mechanism of ligand binding of BABPs we additionally performed a comparative analysis of binding capabilities of chicken I-BABP with homologous proteins.

The interaction of chicken I-BABP with bile acids has been studied through different approaches, including chemical shift perturbation mapping by NMR titrations and Isotermal Titration Data (ITC).
The obtained data show that the chicken I-BABP binds two molecules of glycochenodeoxycholic acid (GCDA) inside the internal cavity. In addition, in response to ligands binding, the protein undergoes a global structural rearrangement. ITC data and NMR experiments independently suggest that the interaction of chicken I-BABP with GCDA involves two consecutive binding events. In particular, the first binding event promotes a large structural rearrangement, possibly related to an allosteric mechanism. Both the used approaches pointed to the presence of a singly ligated protein at low ligand/protein molar ratios, thus suggesting the absence of binding cooperativity. Competition experiments performed on cI-BABP with GCDA and GCA (glycocholic acid) show the absence of site-selectivity for one site,

while the other binding site displays some selectivity for GCDA. A similar analysis performed on a mutant (Q51A) does not affect the ligand-specificity for GCDA.

Comparing the binding features displayed by chicken I-BABP with homologous proteins (human I-BABP, rabbit I-BABP, zebrafish I-BABP and chicken liver BABP), the following considerations can be made:

- The binding of two ligand molecules inside the internal cavity appears to be a common feature of all homologous BABPs. Chicken I-BABP does not display any evidence of additional superficial binding sites as reported for the rabbit and zebrafish proteins.
- In order to assess the binding cooperativity displayed by cI-BABP and homologues proteins, we calculated, for each protein, the change of free energy of interaction between the two internal binding sites, using reported macroscopic binding constants. The obtained values indicate that the human protein displays the largest cooperativity, the rabbit and zebrafish BABPs have intermediate cooperativity, while the chicken protein has poor energetic coupling between sites.

  The identification of substitutions that can correlate with the trends in binding cooperativity is not straightforward and the only available high-resolution structure of an ileal protein with two bile salts bound inside the cavity is that of the zebrafish protein. It is thus clear that more structural data are necessary in order to identify the determinants of cooperativity.
- the site-selectivity displayed by chicken I-BABP show differences compared to that reported for human I-BABP. Indeed, human I-BABP has been found to exhibit a high degree of ligand site selectivity in its interactions with GCA and GCDA, and a mutation of glutamine in position 51 results in the loss of site preference for both bile acids.

In conclusion, we present the characterization of bile acids interactions with chicken I-BABP, aimed at contributing to determine the chemical basis of intracellular bile acid transport, by taking advantage of comparative analyses

with homologous proteins. The chicken I-BABP can be considered as an allosteric system, which is able to establish long-range communication networks through a significant conformational rearrangement induced by a first binding event, but that does not displays positive binding cooperativity, as instead observed for the human protein and its paralogous chicken liver BABPs. The observed differences in binding features make chicken I-BABP a suitable model for the understanding of cooperativity among BABPs and may provide indications about the development of this efficient feature during species evolution.

My contribution in this work consisted in production of the protein samples, in the assignment of protein resonances and in the analysis of NMR titration experiments.

BJ www.biochemj.org
Biochem. J. (2010) 425, 413–424 (Printed in Great Britain) doi:10.1042/BJ20091209

Chicken ileal bile-acid-binding protein: a promising target of investigation to understand binding co-operativity across the protein family

Mara GUARIENTO*, Michael ASSFALG*, Serena ZANZONI*, Dimitrios FESSAS†, Renato LCNGHI‡ and Henriette MOLINARI*[1]

*Dipartimento di Biotecnologie, Università degli Studi di Verona, Strada Le Grazie 15, 37134 Verona, Italy, †Dipartimento di Tecnologie Alimentari e Microbiologiche, Università degli Studi di Milano, Via Celoria 2, 20133 Milano, Italy, and ‡Istituto di Chimica del Riconoscimento Molecolare, Consiglio Nazionale delle Ricerche, Via Mario Bianco 9, 20131 Milano, Italy

Protein–bile acid interactions are crucial microscopic events at the basis of both physiological and pathological biochemical pathways. BABPs (bile-acid-binding proteins) are intracellular transporters able to bind ligands with different stoichiometry, selectivity and co-operativity. The molecular determinants and energetics of interaction are the observables that connect the microscopic to the macroscopic frameworks. The present paper addresses the study and proposes a mechanism for the multi-site interaction of bile acids with chicken I-BABP (ileal BABP) with the aim of elucidating the determinants of ligand binding in comparison with homologous proteins from different species and tissues. A thermodynamic binding model describing two independent consecutive binding sites is derived from isothermal titration calorimetry experiments and validated on the basis of both protein-observed and ligand-observed NMR titration data. It emerges that a singly bound protein is relatively abundant at low ligand/protein molar ratios assessing the absence of strong co-operativity. Both the measured energetics of binding and the distributed protein chemical-shift perturbations are in agreement with a first binding event triggering a global structural rearrangement. The enthalpic and entropic contributions associated with binding of the first ligand indicate that the interaction increases stability and order of the bound protein. The results described in the present study point to the presence of a protein scaffold which is able to establish long-range communication networks, but does not manifest positive-binding co-operativity, as observed for the human protein. We consider chicken I-BABP a suitable model to address the molecular basis for a gain-of-function on going from non-mammalian to mammalian species.

Key words: bile-acid-binding protein, chicken ileum, co-operativity, lipid transport, multi-site interaction, NMR.

INTRODUCTION

BAs (bile acids) are amphipathic molecules with a steroid backbone that are synthesized from cholesterol in hepatocytes. Their well-known physiological function concerns hepatic bile formation and absorption of dietary lipids and fat-soluble vitamins from the small intestine [1]. The interaction of BAs with nuclear receptors was further shown to provide a negative-feedback pathway for BA synthesis which is critical for the maintenance of whole-body cholesterol homoeostasis [2]. The co-ordinate action of several transporter proteins is responsible for the conservation of the physiological pool of BA molecules, but also for the distribution of lipophilic exogenous drugs [3]. The intracellular BABPs (BA-binding proteins) involved in lipid trafficking belong to the FABP (fatty-acid-binding protein) family, the members of which are abundantly expressed small-molecular-mass (14–15 kDa) proteins [4]. Despite the considerable differences in their primary structure, the tertiary structure of all FABPs is highly conserved and consists of a ten-stranded β-barrel, made of two orthogonal antiparallel five-stranded sheets forming a 'clam'-shaped cavity which can host a diverse number of ligands. The opening of the binding site is framed by an N-terminal helix-turn-helix domain. FABPs bind lipid molecules with differences in ligand selectivity, binding affinity and binding mechanism as a result of small structural differences [5].

The binding capabilities of BABPs appear difficult to be captured: indeed, the first reported NMR structures, of pig I-BABP (ileal BABP) [6] and human I-BABP [7] refer to a singly ligated protein, in contrast with the later reports on human I-BABP [8] and chicken L-BABP (liver BABP) [9] showing a protein/ligand stoichiometry of 1:2. The chicken L-BABP/cholate is the only NMR structure of a ternary BABP complex so far reported. Recent studies addressing the binding stoichiometry of rabbit and zebrafish I-BABPs with different bile acids, using calorimetry combined with MS [10] and X-ray crystallography [11] respectively, suggested that these proteins may have the capability of binding two ligands inside their cavity and an additional one or two ligands on the molecular surface. As for the binding mechanism, the most comprehensive work has been reported on human I-BABP [12,13]. The latter protein binds two molecules of GCA (glycocholic acid) with low intrinsic affinity, but an extraordinarily high degree of positive co-operativity. This feature provides insights into the biological mechanism by which I-BABPs may act as buffering agents in ileocytes against high levels of free glycocholate that may trigger apoptosis or cytotoxicity. However, the structural basis for this positive co-operativity has not yet been elucidated [8].

A previous study has shown that chicken L-BABP, similarly to human I-BABP, displays very high binding co-operativity towards bile salts [14], and it was suggested that 'a population-shift model' [15] was responsible for the observed co-operativity. Subsequent studies on selected mutants of both human I-BABP [12] and chicken L-BABP [16] have started to identify protein residues that are functional to the binding mechanism. A further approach towards the description of the interaction mechanism

Abbreviations used: BA, bile acid; BABP, BA-binding protein; CSP, chemical-shift perturbation; CT-HSQC, constant time-heteronuclear single-quantum coherence; FABP, fatty-acid-binding protein; GCA, glycocholic acid; GCDA, glycochenodeoxycholic acid; HSQC, heteronuclear single-quantum coherence; I-BABP, ileal BABP; ITC, isothermal titration calorimetry; L-BABP, liver BABP, LB, Luria–Bertani; U-GCDA, unbound GCDA.

[1] To whom correspondence should be addressed (email henriette.molinari@univr.it).

may derive from the comparative analysis of similar proteins from different tissues and organisms. In the present study we address the characterization of the binding interactions of chicken I-BABP with bile salts using ITC (isothermal titration calorimetry) and NMR spectroscopy experiments, providing the binding stoichiometry and the thermodynamic analysis of binding. By taking advantage of comparative analyses based on an extended protein/ligand molecular space, the present results aim at contributing to understand the chemical basis of intracellular bile acid transport, which may open up new avenues for exploration of strategies for prevention and treatment of metabolic diseases and further biomedical applications [5,17,18].

EXPERIMENTAL

Cloning and mutagenesis

The pQE50 plasmid carrying the chicken I-BABP coding sequence [4] was mutated using the QuikChange® site-directed mutagenesis kit (Stratagene). Forward and reverse primers (MWG Biotech) were designed to insert a stop codon after the sequence coding for the first four residues of the thrombin cleavage site (at the C-terminal end of the chicken I-BABP coding sequence). The shorter construct coded for chicken I-BABP was fused with the same tag that would have been found after the proteolytic cut of the longer product. The short construct was used to transform *Escherichia coli* XL1Blue competent cells (Stratagene). After a miniprep extraction (Sigma–Aldrich), the mutated plasmid sequence was verified by DNA sequencing. The recombinant vector was then used to transform the *E. coli* SG expression strain (Qiagen) and transformed cells were selected on LB (Luria–Bertani) agar plates containing 100 μg/ml ampicillin (Sigma–Aldrich).

Protein expression and purification

Chicken I-BABP has been expressed as a soluble protein inoculating 1 litre of LB medium with 10 ml of starter culture and incubating at 37 °C until cells reached a D_{600} of 0.6. Protein expression was induced by the addition of 0.5 mM IPTG (isopropyl β-D-thiogalactoside) and incubation overnight at 20 °C. Owing to the presence of the C-terminal tag plus the four residue N-terminal tag, derived from the BamHI cleavage site used for the cloning step, the predicted isoelectric point of chicken I-BABP shifted from the original value of 6.8 to 8.8. Supplementary Figure S1 (at http://www.BiochemJ.org/bj/425/bj4250413add.htm) shows the constructs derived from different cloning strategies. Given the increased value of pI, the same purification protocol previously described [14] was used, based on ionic exchange instead of affinity chromatography, thus increasing the yield of soluble protein from 40 to 80 mg/l of culture in rich medium (LB). This result was possibly due to the elimination of the thrombin cleavage step, one of the causes of protein precipitation during the purification process. ^{15}N and ^{13}C isotope labelling was achieved using M9 minimal media containing 1 g/l NH_4Cl and 4 g/l ^{13}C-enriched glucose and incubating the cultures overnight at 28 °C. The final yield of purified proteins was 30 mg/l of culture for both ^{15}N and ^{15}N-^{13}C chicken I-BABP. The purity of the samples was verified by SDS/PAGE, and MS analysis was used to quantify the isotopic enrichment (close to 100 % in both ^{15}N and ^{15}N-^{13}C chicken I-BABP). Purified samples were concentrated, buffer-exchanged in 20 mM potassium phosphate, pH 6.5, 135 mM KCl, 10 mM NaCl, 1 mM DTT (dithiothreitol) and 0.05 % sodium azide, and stored as freeze-dried protein.

Determination of the protein aggregation state

The state of aggregation of chicken I-BABP was determined by dynamic light scattering. The protein was found to be monomeric at the concentration used for the NMR experiments (0.4–0.5 mM) in both apo and holo form [in complex with GCDA (glycochenodeoxycholic acid) at a protein/ligand ratio of 1:6]. Measurements of the centrifuged NMR samples were carried out at 298 K using a DYNA-PRO 801 dynamic light-scattering/molecular-sizing instrument (Protein Solutions). Data were processed with Dynamics software.

ITC

The protein and the ligand (GCDA) were dissolved in the following buffer: 20 mM potassium phosphate, 135 mM KCl and 10 mM NaCl (pH 6.5) with the addition of 0.02 % sodium azide. The GCDA concentration in the titrating solution was 12 mM and the initial protein concentration in the measurement cell (determined spectrophotometrically) was 0.2 mM. The measurements were performed at 25 °C using a CSC Nano Isothermal Titration Calorimeter III (model 5300). A total of 20 injections of 10 μl aliquots of titrating solution were added to the 973 μl of protein solution cell. The heat of the injections was corrected for the heat of dilution of the ligand into the buffer. Experiments were performed in triplicate.

Several binding models were tested to interpret the calorimetric data and the fitting functions have been described previously [19–22]. Briefly, the observable enthalpy is given by (eqn 1):

$$\Delta H(T, p, \mu_L) = -R\left[\frac{\partial \ln Q}{\partial(1/T)}\right]_{p,\mu_L} \quad (1)$$

and the degree of association, i.e the concentration ratio x = [bound ligand]/[total protein] is given by (eqn 2):

$$\bar{X} = RT\left[\frac{\partial \ln Q}{\partial \mu_L}\right]_{T,P} = \left[\frac{\partial \ln Q}{\partial \ln[L]}\right]_{T,p} \quad (2)$$

where R is the universal gas constant, μ_L is the chemical potential of the free ligand, [L] is the concentration of the free ligand and Q is the partition function of the system referred to as the free protein state. Since we can approximate these systems as diluted solutions, the thermodynamic activities of the solutes may be replaced with their molar concentrations. Under this assumption the partition function is the sum of the concentrations of all protein species, P_j, referred to the free protein, P_0 (eqn 3):

$$Q = \sum_{j=0}^{n} [P_j]/[P_0] \quad (3)$$

Q depends on the assumptions made on the association (binding) mechanism and is the key function used to simulate the enthalpy so as to check the model with the experimental data and to obtain the association (or binding) constant, K_b, and the binding enthalpy ΔH_b. The binding constant is a dimensionless quantity by definition. However, in order to stress the approximation of the thermodynamic activities with the molar concentrations, the use of M^{-1} units for this parameter is widely used and was adopted in the present paper.

Data fitting to the binding models was performed using the non-linear Levendberg–Marquardt method [23]. The errors of each fitting parameter were calculated with a 95.4 % confidence limit by the Monte Carlo simulation method. An error of 5 % in the protein concentration was also taken into account.

NMR spectroscopy

NMR samples of chicken I-BABP were prepared by resuspending the freeze-dried protein in 90 % H_2O/10 % ^2H_2O obtaining a final concentration of 0.4 mM. Unenriched GCDA and GCA were purchased as sodium salts from Sigma–Aldrich. ^{15}N-Enriched GCDA and GCA were prepared as previously described [4] and transformed into the corresponding bile salts by addition of 2.15 μl of 1.1 M KOH for each milligram of bile acid. Both unenriched and ^{15}N-enriched bile salts were then used to prepare stock solutions in a buffer of 20 mM potassium phosphate, pH 6.5, 135 mM KCl, 10 mM NaCl and 0.05 % sodium azide. Titration experiments using ^{15}N-labelled ligand consisted of fourteen titration points measured at ligand/protein ratios of $r = 0.3$, 0.6, 0.7, 0.9, 1.1, 1.5, 1.8, 2.2, 2.7, 3.6, 5.0, 6.5, 8.5 and 10.

NMR experiments were run on a Bruker Avance III 600 spectrometer, operating at 600.13 MHz proton Larmor frequency, equipped with a triple resonance TCI cryoprobe, incorporating gradients in the z-axis, or on a Varian INOVA 800 MHz, equipped with a HCN 5 mm z-PFG cryogenic probe. The experimental temperature was set to 298 K, unless otherwise specified. Standard pulse sequences were used, incorporating pulsed field gradients to achieve suppression of the solvent signal and spectral artefacts. Linear prediction was applied to extend the indirect ^{13}C-detected dimension. Direct and indirect dimensions were normally apodized using 90°-shifted squared sine-bell functions (for ^{13}C- and ^{15}N-edited dimensions) or Lorentzian-to-Gaussian functions (for the ^{1}H dimension), followed by zero filling and Fourier transformation.

Sequential assignment and assignment of Cα, Cβ, Hα, Hβ and carbonyl carbon resonances were obtained by performing HNCACB, CBCA(CO)NH, HNCO, HBHANH and HBHA(CO)NH experiments on ^{15}N-^{13}C chicken I-BABP. Each two-dimensional ^{1}H-^{15}N HSQC (heteronuclear single-quantum coherence) spectrum was acquired with a spectral width of 9615 Hz and 1024 complex points in the ^{1}H dimension and a spectral width of 2432 Hz and 256 complex points in the ^{15}N dimension. The number of scans was 120 for experiments performed on unenriched chicken I-BABP in complex with ^{15}N-bile salt or on ^{15}N-bile salt alone, and 16 for the experiments performed on ^{15}N chicken I-BABP.

Processing of all the spectra have been performed with TOPSPIN 2.1 (Bruker), whereas assignments have been performed using the software Analysis, developed for the Collaborative Computing Project for the NMR community [24] and NMRView [25].

Combined chemical-shift changes were calculated as

$$\Delta\delta_{HN,N} = [(\Delta\delta_{HN}^2 + \Delta\delta_N^2/25)/2]^{0.5}$$

Considerations on protein stability

The superimposition of two ^{1}H-^{15}N HSQC spectra of the same sample of chicken I-BABP–GCDA complex (1:5 ratio) recorded at different time points indicated that after a few months ten peaks (Ala^1, Phe^2, Thr^3, Gln^{42}, Asn^{43}, Gly^{44}, Ile^{65}, Gly^{88}, Lys^{89} and Ile^{90}) were shifted to a slightly different position, whereas three other peaks (Gly^{-2}, Ser^{-1} and Met^0) were missing. These residues are all localized near the N-terminal tag MRGSM added during the cloning: Gly^{-2}, Ser^{-1} and Met^0 belong to the tag; Ala^1, Phe^2 and Thr^3 were directly linked to the tag; Gln^{42}, Asn^{43}, Gly^{44}, Ile^{65}, Gly^{88}, Lys^{89} and Ile^{90} were all positioned in loops surrounding the N-terminus. These results suggest that chicken I-BABP has the tendency to lose the five N-terminal residues, an hypothesis confirmed by MS analysis. Possibly this tag has been removed by a contaminant protease or by an enzyme such as MAP1 (methionine aminopeptidase 1), which is responsible of the cleavage of the first methionine residue in *E. coli* proteins [26]. However, this modification did not affect the binding of bile acids to chicken I-BABP, as verified by ITC performed on a sample before and after separation of the two forms of protein.

Rosetta protein structure calculation

The protein structural model was generated from chemical shifts, as described in the Results session, using the software Rosetta [27], as implemented on the e-grid.

RESULTS

Protein NMR resonance assignment and backbone structure analysis

A series of heteronuclear three-dimensional NMR experiments were used for the assignment of ^{1}H, ^{13}C and ^{15}N backbone resonances of chicken I-BABP in the unbound form as well as in complex either with GCA or with GCDA. A complete assignment was obtained for the GCDA and GCA complexes, while an assignment level of 83 % was derived for the apo protein. For this protein form, spectral correlations were missing for the stretch Glu^{72}–Ala^{81}, as a result of fast exchange of amide protons with the solvent and/or conformational exchange. A similar behaviour was reported for apo chicken L-BABP [14]. The apparent flexibility of this specific region of the apo protein (E, F strand and EF loop) is consistent with the presence of an ensemble of native states. Conformational exchange of the apo form is quenched upon ligand binding, as observed for chicken L-BABP [9,14].

The CSI (chemical-shift index) [28] analysis is consistent with the expected secondary structure (Supplementary Figure S2 at http://www.BiochemJ.org/bj/425/bj4250413add.htm). The experimental chemical shifts of ^{13}Cα, ^{13}Cβ, ^{13}C′, ^{15}N, Hα and H_N of the holo I-BABP–GCDA complex were used as input for the CS-Rosetta software [26–28] to generate a preliminary protein structure. A decision on whether the CS-Rosetta structure generation process has converged was based, as reported, on how well the co-ordinates of the lowest-energy structures agree with one another [29]. A plot of Rosetta all-atom energy as a function of the Cα RMSD (root mean square deviation) relative to the model with the lowest energy shows a 'funnelling' distribution (Supplementary Figure S3 at http://www.BiochemJ.org/bj/425/bj4250413add.htm), identified by Shen et al. [29] as a convergence criterion. The ten lowest-energy structures are within 2 Å (1 Å = 0.1 nm) from the best model (Supplementary Figure S4 at http://www.BiochemJ.org/bj/425/bj4250413add.htm).

CSP (chemical-shift perturbation) mapping upon ligand binding

The ^{1}H-^{15}N correlation NMR spectrum of a protein and its specific pattern can be regarded as a fingerprint of its structure. Monitoring the perturbations of ^{15}N and H_N chemical shifts upon binding to a ligand is commonly used to map residues involved in binding sites and/or to identify conformational rearrangements. Figure 1 shows the superimposition of the ^{1}H-^{15}N HSQC spectrum of apo chicken I-BABP with the spectra of the protein in complex with GCDA and GCA (Figures 1a and 1c), and a superimposition of the spectra

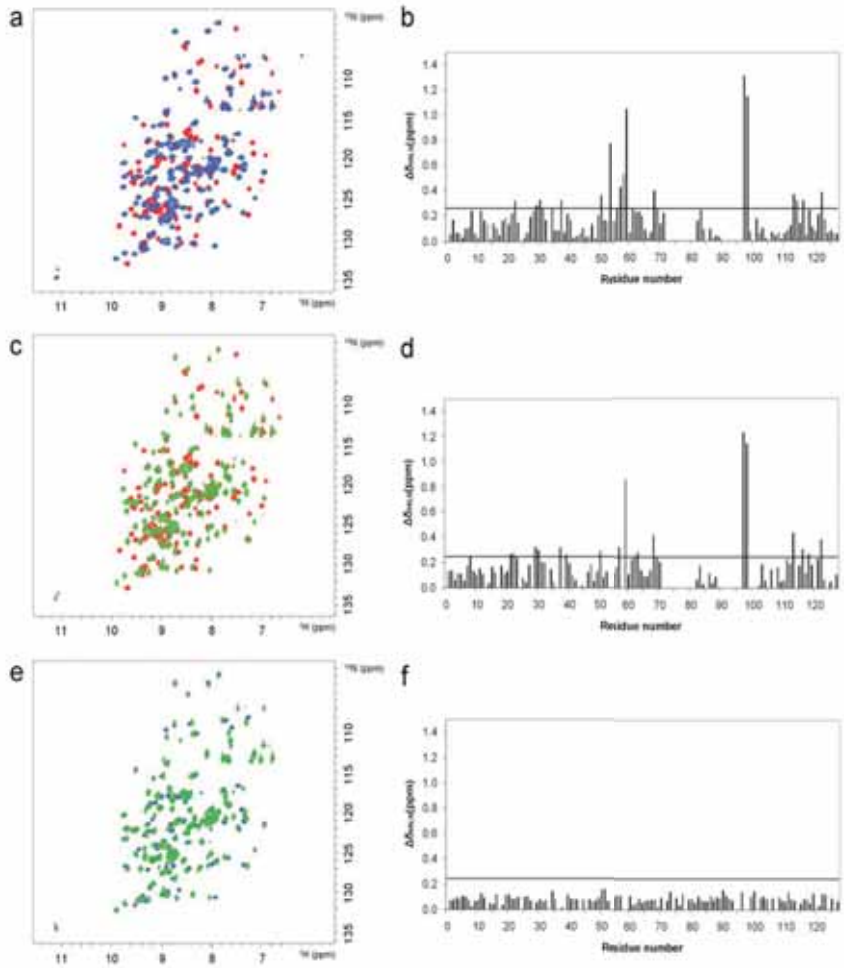


Figure 1 Chemical-shift changes of chicken I-BABP backbone ^{15}N-H resonances on ligand binding

Left-hand panels: superimpositions of ^{1}H-^{15}N HSQC spectra of (**a**) apo [^{15}N]I-BABP (in red) and [^{15}N]I-BABP–GCDA, 1:5 ratio (in blue; (**c**) apo [^{15}N]I-BABP (in red) and [^{15}N]I-BABP–GCA, 1:5 ratio (in green); and (**e**) [^{15}N]I-BABP–GCDA, 1:5 ratio (in blue) and [^{15}N]I-BABP–GCA, 1:5 ratio (in green). Right-hand panels: plots of the chemical-shift differences relative to (**b**) unbound protein and protein bound to GCDA; (**d**) unbound protein and protein bound to GCA; and (**f**) protein bound to GCDA and bound to GCA. Residues exhibiting a chemical shift perturbation above the displayed threshold line are Gly^{22}, Glu^{29}, Met^{30}, Val^{37}, Thr^{50}, Gly^{55}, Arg^{57}, Phe^{63}, Glu^{68}, Thr^{113}, Gly^{116}, Val^{118} and Thr^{122} for both complexes; Phe^{5}, Asp^{17}, Ile^{28}, Gly^{31}, Cys^{34}, Phe^{53}, Thr^{66}, Ala^{69}, Val^{83} and Ser^{114} for the complex with GCDA; and Ile^{21}, Glu^{39} and Gly^{56} for the complex with GCA. The missing bars refer to residues for which the assignment of the unbound form is missing, or to proline residues.

of the two holo forms (Figure 1e). The residue-specific CSPs derived from the comparison of apo and holo chicken I-BABPs, in complex with GCDA and GCA are shown in Figures 1(b) and 1(d) respectively, whereas Figure 1(f) shows the direct comparison between the two holo proteins. Both ligands induce chemical-shift changes that are distributed over the entire protein backbone and, in analogy with other proteins of the BABP family, it can be inferred that the perturbations are not only due to direct interaction with the ligands, but may originate from a global protein structural rearrangement. Thus chemical-shift effects induced by ligand binding are difficult to interpret directly in terms of binding sites. However, the limited differences observed between the spectra of the two bound forms suggest that the binding sites for the two ligands are structurally similar. Given this similarity, most of the NMR binding analysis reported in the present paper refers to the chicken I-BABP–GCDA complex.

ITC experiments

Figure 2 reports the experimental ITC data of GCDA binding to chicken I-BABP in a ΔH against r plot (ΔH = cumulative

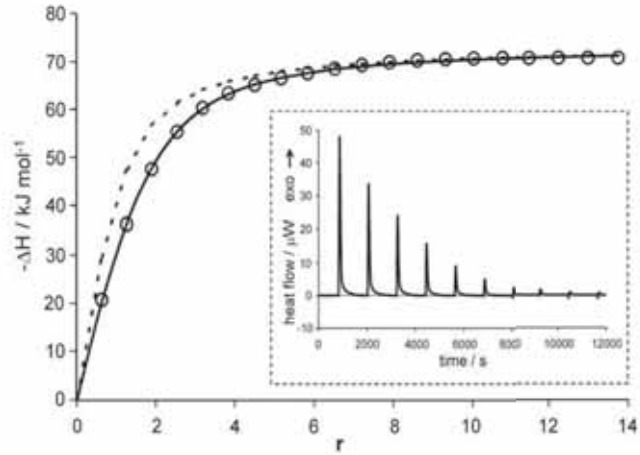


Figure 2 ITC experiment

GCDA binding to chicken I-BABP at 25 °C. The raw data for the first ten injections are shown in the insert. Circles represent the experimental data in the ΔH against r plot, where ΔH is the cumulative enthalpy (sum of the peak areas in the insert) expressed per mol of protein, and r is the total titrated ligand/total protein concentration ratio. The lines represent the best fit of the data using two thermodynamic binding models: single site (broken line) and two consecutive sites (continuous line).

enthalpy per mol of protein; r = concentration ratio = total titrated ligand/total protein). Thermodynamic binding models of increasing complexity, in terms of the number of fitting parameters and of protein states, were tested to interpret the calorimetry measurements, assuming that the interaction is best described by the simplest model in agreement with the experimental data. The best fit trials using the simplest model, i.e. one-step single-site binding, did not yield satisfactory results. An excellent fit (Figure 2) was instead achieved using the next thermodynamic model (according to the minimum complexity criterion) that implies the presence of two consecutive binding sites. The relevant partition function (see the Experimental section) in this case is: $Q = 1 + K_{b1}[L] + K_{b1}K_{b2}[L]^2$ and implies three protein states: free, one binding site occupied and two sites occupied. The values of the corresponding best fit thermodynamic parameters are reported in Table 1. We observed that in both binding events the interaction is enthalpically driven. Models with additional complexity (for example the two independent binding site model, described by $Q = 1 + K_{b1}[L] + K_{b2}[L] + K_{b1}K_{b2}[L]^2$, and implying four protein states) may also be considered appropriate. However, in the absence of additional experimental details their application is not justified, whereas the simplest model proposed represents well the average behaviour of all possible microstates.

NMR ^{15}N and ^{13}C titration results

The direct comparison of protein resonances for the unbound and bound forms provide information about the two extremes of the binding reaction. In principle, by monitoring protein signals in the presence of different amounts of ligand, it is possible to identify on-path intermediates along the binding co-ordinate. A titration experiment has been performed on ^{15}N-labelled chicken I-BABP using unlabelled GCDA. Forty-seven residues could be safely followed throughout the titration and all displayed a slow or slow-to-intermediate exchange on the NMR time scale. For some resonances the disappearance of apo peaks and appearance of holo peaks was accompanied by small chemical-shift changes. For an additional 26 residues only the peaks relative to the holo form could be analysed, due to lack of assignment or peak crowding for the corresponding apo resonances. The possibility of distinguishing differential behaviours in protein–ligand titrations may provide an indication about residues involved in different binding events or sites. Observed perturbations were, however, distributed all over the protein, and pattern recognition approaches, such as cluster analysis and principal component analysis, applied to the titration data did not allow precise grouping of residues, further indicating that protein residues could not be considered independent probes of ligand binding, but rather the reporters of a global change.

^{1}H-^{13}C correlation spectra of chicken I-BABP–GCDA at $r = 0$, 1, 2 and 4 were also analysed in the assumption that carbon chemical shifts may select more localized conformational changes and offer the possibility to localize the binding sites. Selected regions of ^{1}H-^{13}C CT (constant time)-HSQC spectra registered at varying molar ratios are shown in Figure 3. Two different behaviours could be distinguished. A group of residues exhibited a slow-exchange regime on the chemical-shift time scale, and at $r = 1$ both resonances due to the apo and holo forms were present, with the holo resonance already appearing at the chemical shift of the saturated holo protein (Figure 3). A second group of residues exhibited a slow-to-intermediate exchange regime, and a small but detectable chemical-shift change was observed both for the apo and the holo forms. The often encountered different direction of the shift changes of apo and holo forms suggests the presence of an intermediate, which is likely to be a singly ligated state, present at low r, which evolves, upon GCDA addition, towards the doubly ligated form. A thorough comparison of chemical shifts at various protein–ligand ratios was performed on 60 isolated CαHα resonances. The following observations were obtained: (i) a first group of residues did not exhibit any chemical-shift change on going from $r = 0$ to $r = 4$, thus indicating that they were not affected by binding. These residues are mainly located at the

Table 1 Thermodynamic parameters obtained from ITC measurements considering two or three stepwise binding models for the interaction of BAs and BABPs from different species

The reported data refer to the following ligands: GCDA (chicken and rabbit), GCA (human) and cholic acid (zebrafish), which were shown to display the highest affinities for the corresponding BABP. All measurements were performed at 25 °C.

	K_1 ($\times 10^6$ M^{-1})	ΔH_1 (kJ · mol^{-1})	$T\Delta S_1$ (kJ · mol^{-1})	K_2 ($\times 10^6$ M^{-1})	ΔH_2 (kJ · mol^{-1})	$T\Delta S_2$ (kJ · mol^{-1})	K_3 ($\times 10^6$ M^{-1})	ΔH_3 (kJ · mol^{-1})	$T\Delta S_3$ (kJ · mol^{-1})	$\Delta\Delta G$ (kJ · mol^{-1})§
Chicken	1.55 ± 0.15	−50.3 ± 5.0	−26.4 ± 2.7	0.55 ± 0.05	−23.3 ± 2.3	−1.95 ± 0.2	–	–	–	−0.87
Human*	0.13 ± 0.03	10.0 ± 6.2	6.7 ± 1.5	16.9 ± 3.2	−11.1 ± 0.1	−4.0 ± 0.1	–	–	–	−15.50
Rabbit†	0.47 ± 0.4	−0.2 ± 0.3	–	0.6 ± 0.1	−9.0 ± 2.1	–	2.5 ± 0.2	−18.7 ± 2.0	–	−6.96
Zebrafish‡	0.63 ± 0.1	−5.1 ± 0.5	–	1.02 ± 0.1	−1.6 ± 0.2	–	–	–	–	−4.55

*See [12].

†The value of $\Delta\Delta G$ for this system is calculated for the last two binding steps, under the assumption that they correspond to the binding of ligands inside the cavity, based on the reported large enthalpic contributions [10].

‡The ITC data were also consistent with a model of three independent binding sites [11].

§The reported values are the upper limits of free energy coupling between binding events corresponding to the situation of identical intrinsic binding constants and calculated as $\Delta\Delta G = -RT\ln(4K_2/K_1)$ [32].

bottom of the protein cavity, and in regions of strands I-J, adjacent to loop IJ; (ii) a second group of residues, evenly distributed on all the strands of the proteins, reached their final chemical shift at $r = 1$; (iii) a few residues, mainly located in helix I, still exhibited a minor chemical-shift change on going from $r = 1$ to $r = 4$.

Binding-site occupancies and site selectivity

An NMR titration was performed on the unlabelled chicken I-BABP with [^{15}N]GCDA to assess the stoichiometry and to monitor binding-site occupancies. Selected ^{1}H-^{15}N HSQC spectra collected after each ligand addition are shown in Figure 4. It appears that a single resonance (δH=7.84 p.p.m., δN = 119.29 p.p.m.) is observable up to $r = 1.8$, whereas a second resonance appears at $r = 2.2$ (δH = 7.21 p.p.m., δN = 112.06 p.p.m.), at higher field in both the proton and the nitrogen frequencies. Both peaks are assigned to a bound form of the ligand as their chemical shifts and linewidths are different from those of U-GCDA (unbound GCDA), determined separately in a control experiment. Interestingly, the peak resonating at lower field (δH = 7.72 p.p.m., δN = 119.17 p.p.m.) changes position and becomes sharper (Figure 4g) during the titration, reaching, at high r values, chemical shift (δH = 7.88 p.p.m., δN = 119.48 p.p.m.) and linewidth values similar to those of U-GCDA. This behaviour is consistent with an intermediate-to-fast chemical-exchange regime between U-GCDA and GCDA bound to a specific site (named site 1). By lowering the temperature to 283 K and repeating the experiment at higher field (900 MHz), it was still not possible to separate the resonances of the free and the bound ligand. The low-field peak was thus marked as 1/U indicating its composite nature. The high-field peak was marked as 2, referring to the binding of the ligand to a second site. This signal is also experiencing linewidth reduction and chemical-shift change when approaching saturation, possibly reflecting loss of chemical exchange with the ligand in the singly bound protein (Figure 4h).

In order to investigate whether BA binding was site-selective, as reported for the homologous human I-BABP [8], complementary competition experiments were performed with GCDA and GCA. In these experiments an unlabelled ligand (GCDA or GCA) was added to a complex formed by I-BABP and [^{15}N]GCA or [^{15}N]GCDA ($r = 3$) respectively. The corresponding spectra are shown in Figure 5. It can be observed that while peak 2 disappears almost immediately after addition of the competitor in both experiments, the signal named 1/U behaves differently. In this site, unlabelled GCDA is able to displace the bound [^{15}N]GCA, and indeed the resonance left at 7.92–121.51 p.p.m. is very close to that characteristic of unbound [^{15}N]GCA. On the other hand, unlabelled GCA is less efficient in the displacement of bound [^{15}N]GCDA, as deduced by the fact that the low-field peak is poorly affected by the addition of unlabelled competitor. These experiments revealed the absence of ligand-specificity of one site (site 2) together with some ligand-specificity displayed by the other site (site 1).

DISCUSSION

Description and assessment of the binding model

The binding model that best describes the ITC data relative to the chicken I-BABP–GCDA interaction involves two consecutive, enthalpically driven binding events. Both the existence of two binding sites and the succession of binding events is independently confirmed by NMR experiments based on the observation of ligand ^{15}N-H signals. The pertinence of the proposed

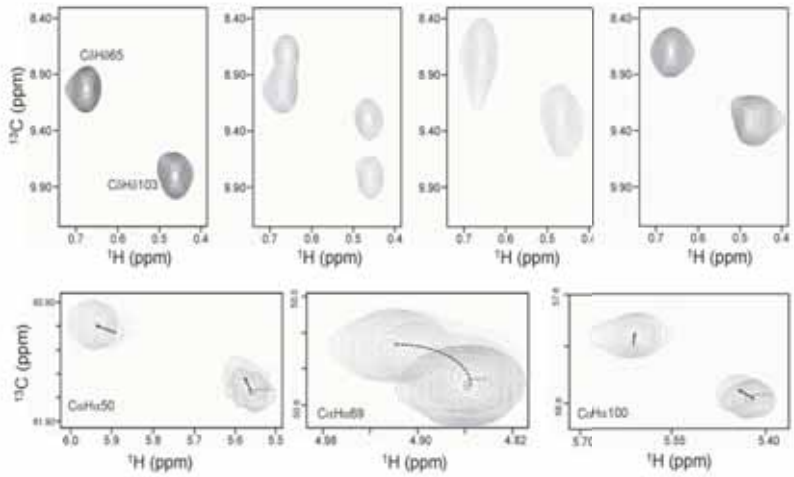


Figure 3 Protein ^{13}C-H CSPs on ligand binding

Selected regions of ^{1}H-^{13}C CT-HSQC spectra registered at 600 MHz at 25 °C, at increasing ligand/protein ratios, r. Upper panels: residues showing two distinct resonances in slow exchange, from left to right: $r = 0$, $r = 1$, $r = 2$ and $r = 4$. Lower panels: residues experiencing minor chemical-shift changes of the resonances corresponding to the unbound and fully bound forms, arrows follow the direction of peak shifts on increasing ligand concentration.

thermodynamic model was further evaluated by comparing the predicted population fractions of species in solution (Figure 6) with the experimental curves of protein signals in NMR titration experiments. The NMR experiments performed by observing either ^{15}N-H amide or ^{13}C-H protein resonances show the appearance of two sets of separate signals, changing in intensity and displaying only minor chemical-shift variations. Cross-peak intensities of resonances appearing in a slow chemical-exchange regime were considered reasonable reporters of relative protein concentrations. Because of the unambiguous assignment of one set of resonances to the unbound protein, the intensities of the corresponding signals were analysed, normalized with respect to the intensity measured at $r = 0$, and corrected to take into account the lower detection limit of the experimental method (corresponding to a protein concentration of approx. 100 μM). In Figure 7 the averaged normalized intensity of 22 analysed residues is plotted as a function of r, showing a striking agreement with the population fraction of the unbound protein, thus confirming the validity of the proposed model.

The ITC-derived model predicts a considerable concentration of singly ligated protein in a wide range of ligand–protein molar ratios, accounting for approx. 50% of all protein in the range of $1 < r < 2$. On the other hand, the complementary NMR experiments show the appearance of only two sets of separate signals. Thus the singly bound protein is either invisible in the present experimental conditions, or its signals are coincident with one of the two sets of resonances. By comparing the sum of the population fractions for both the unbound and singly bound forms with the above-described signal intensities, it is clear that the analysed set of resonances corresponds to only the unbound protein without mixing of the singly bound form. A similar analysis was also performed on the second set of signals and it appeared that the normalized intensities were not corresponding to the population of either the double-bound or the single-bound protein. It seems reasonable to conclude that the second set of resonances corresponds to a combination of single- and double-bound species. However, since the normalized intensity values also do not correspond to the sum of population fractions of the two species, their contribution to the total intensity should be weighted differently. This is physically plausible considering, for example, that a different dynamics or solvent accessibility of the two protein forms can differently affect their signal line-widths and intensities. Nonetheless, in order to avoid excessive data manipulations impairing a rigorous analysis, we chose not to further attempt a quantitative comparison of the data relative to this set of signals with the ITC-derived population fractions. An additional comparative analysis between ITC and NMR data, performed by considering the NMR titration based on the observation of ligand signals, is presented in Supplementary Figure S5 (at http://www.BiochemJ.org/bj/425/bj4250413add.htm).

The two identified binding events are not accompanied by perturbations localized in specific protein regions. On the contrary, from the comparison of the NMR spectrum of the ligand-saturated protein with that of the unbound species, it appears that the perturbations are distributed over the entire protein. This is particularly true when considering the ^{13}Cα shift changes, as they should be less sensitive to indirect binding-correlated events compared with ^{15}N-H signals, but rather report on a direct interaction or local change in structure. Interestingly, the appearance of the heteronuclear correlation spectra collected at $r = 2$ is already very close to that of spectra collected at saturation. Because at this molar ratio the population of the singly ligated species is maximal, these observations may indicate that the first binding event promotes a large structural rearrangement. The nature of the single-bound protein is, however, not straightforward to explore due to the superposition of observables. This is also confirmed by an even distribution in all of the protein strands of the residues showing a distinct second resonance at $r = 1$, but no further shift perturbation of this resonance in the later titration steps (Figure 8).

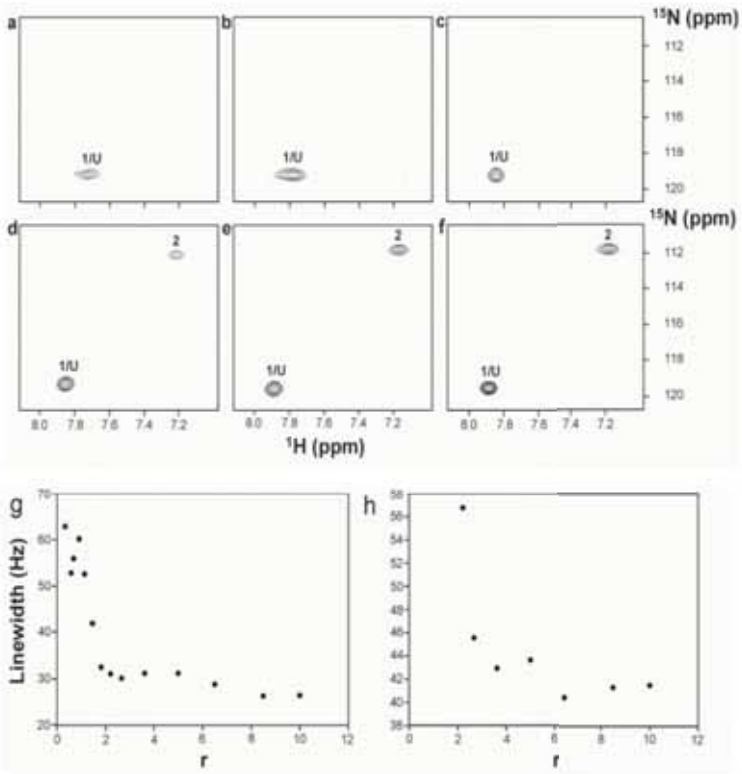


Figure 4 Ligand-observed NMR binding experiments

^{1}H-^{15}N-HSQC spectra registered during the titration of chicken I-BABP with [^{15}N]GCDA. Only a selection of spectra are reported, corresponding to a ligand/protein ratio *r* of (**a**) 0.3, (**b**) 0.9, (**c**) 1.8, (**d**) 2.2, (**e**) 3.6 and (**f**) 10. Line-width values measured during the titration are plotted as a function of *r* for peak 1/U (**g**) and peak 2 (**h**).

Comparative analysis of binding properties in I-BABPs

The binding mechanisms, stoichiometries and affinities of a few I-BABP/BA systems have been reported [8,10,11,13,30,31]. A common feature of BABP proteins from human, rabbit and zebrafish is the binding of two ligand molecules inside the internal cavity. In the case of chicken I-BABP, the strong favourable enthalpic contributions (see below) together with the observation of intermolecular NOEs (nuclear Overhauser effects) between the two ligands and protein residues pointing towards the internal cavity (Supplementary Figure S6 at http://www.BiochemJ.org/bj/425/bj4250413add.htm), again demonstrate the location of two ligands in the protein cavity. Additional secondary binding sites on the protein surface were described for the rabbit and zebrafish proteins. The available data allow a first comparative analysis of binding properties across different species to be made. Because only in the case of human I-BABP intrinsic binding parameters were determined, our analysis is focused on differences in macroscopic equilibrium constants derived from stepwise binding models, as summarized in Table 1. The binding co-operativity can be described by the free energy of interaction between binding sites, $\Delta\Delta G$, which provides a more convenient measure of co-operativity than the Hill coefficient [32]. When site-specific binding constants are not available, it is possible to calculate an upper limit for $\Delta\Delta G$ based on macroscopic binding constants. By comparison of the stepwise affinity constants and of the derived coupling energy, upper limit values of the four analysed protein–ligand adducts, it emerges that the human protein displays the largest co-operativity, the rabbit and zebrafish BABPs have intermediate co-operativity, and the chicken protein has poor energetic coupling between sites.

In order to investigate the determinants of interaction between binding sites, detailed structural data would be helpful. However, up to now, the only available high-resolution structure of an I-BABP with two bile salts bound inside the cavity is that of the protein from zebrafish [11]. We therefore focussed our attention on primary sequence data and structural models. A sequence alignment of the mentioned proteins is reported in Figure 9(B) together with the calculated identities among the proteins (Figure 9A). Non-conservative amino acid substitutions within the binding cavity, among the four investigated proteins, are highlighted on the ribbon structure of the chicken protein in Figure 9(C). These substitutions indicate subtle changes in both the cavity dimensions and in potential hydrophobic and

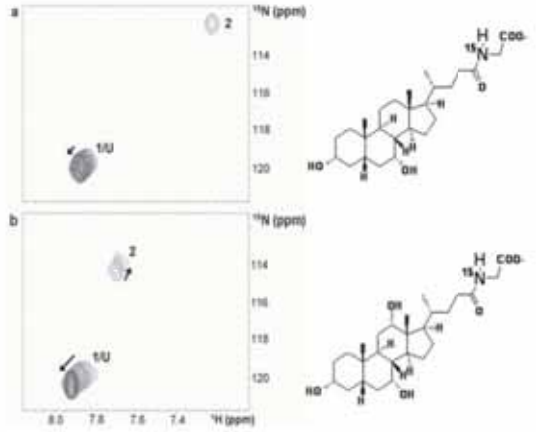


Figure 5 NMR ligand competition experiments

Superimposition of ^{1}H-^{15}N-HSQC spectra of samples containing: (**a**) unlabelled chicken I-BABP (P), [^{15}N]GCDA (L1) and unlabelled GCA (L2) or (**b**) unlabelled chicken I-BABP, unlabelled GCDA (L2) and [^{15}N]GCA (L1), in the following ratios: P/L1/L2 1:3:0, 1:3:1, 1:3:2, 1:3:3. Arrows follow the direction of peak shifts on increasing L2 concentration. The resonances are marked corresponding to ligand bound to site 2 (2) and site 1 or unbound (1/U). The ^{15}N-labelled bile salts are shown to the right of the corresponding spectra.

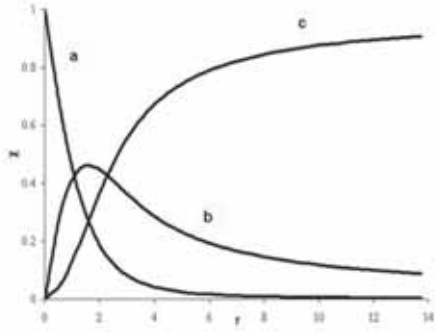


Figure 6 Population fractions of unbound and bound protein species

The population fractions χ are calculated according to the binding model parameters shown in Table 1. Fractions of the unbound, singly bound and double-bound protein forms correspond to curves a, b and c respectively.

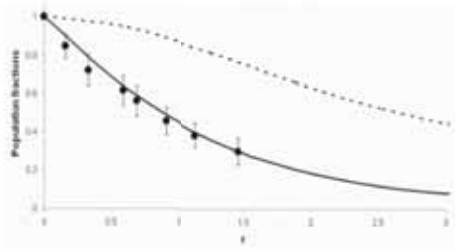


Figure 7 Evaluation of ITC-derived population fractions by NMR

Solid line: population fraction of unbound protein (P); broken line: sum of population fractions of unbound and singly bound protein (PL); ●, experimental average intensity decay of one set of ^{15}N-H protein signals (see the text for details). The error bars represents the variability of the 22 analysed resonances.

hydrophilic interactions with any ligand molecule bound inside the cavity. It is, however, not straightforward to correlate the occurring substitutions with the trends in binding co-operativity. In the sequence alignment reported in Figure 9(B), underlined amino acids mark residues establishing contacts with the ligands, as derived from the first generation NMR structure (never released) of the human protein and from the X-ray structure of the zebrafish protein (PDB code: 3EM0). It is worth noting that Thr^{38}, Trp^{49}, Gln^{51}, Asn^{61}, Gln^{99} and Glu^{110}, which make contacts with the ligands in the human protein, are conserved in all the sequences, with the only exceptions of positions 51 and 99 exhibiting histidine to glutamine mutations without any direct correlation with co-operativity. Indeed both human (highly co-operative) and chicken (poor co-operativity) display a glutamine at position 51, whereas human and zebrafish (intermediate co-operativity) display a glutamine at position 99. The residues making contacts with the ligands in the zebrafish structure are Tyr^{14}, Ile^{23}, Lys^{30}, Tyr^{33}, Val^{34}, Leu^{90}, Tyr^{97}, Gln^{99}, Thr^{101} and Arg^{125}. All of the amino acids occupying these positions are either conserved or exhibit conservative mutations in the analysed sub-family, with the only exception being Lys^{30} and Val^{34}, again without any clear correlation to the presence of co-operativity.

Mutations performed on the residues making contact with the ligands in the human protein indicated that W49A and N61A mutants result in loss of positive binding co-operativity [12]. However, both Trp^{49} and Asn^{61} are conserved in the chicken ileal protein. It was further suggested that these residues are part of a hydrogen-bonding network in the protein–bile salt complex that connects the two sites, giving rise to two co-operativity networks in the doubly ligated system: an upper network involving the steroid ring hydroxy group at position C-12, and a lower network involving the steroid ring hydroxy group at positions C-3 and C-7 [12]. The structural data do not allow us to confirm this hypothesis and indeed only a careful structural correlation of the hydrogen-bonding network in the four proteins

Figure 8 Evidence of ligand-induced global structural rearrangement

The protein structure, displayed in cartoon representation, corresponds to the co-ordinates of the best structure obtained from the chemical-shift-based structure determination method. Black spheres represent protein residues showing two distinct ^{13}C-H resonances in slow exchange along the protein/ligand NMR titration experiment and no further shift perturbation. These residues report on the global structural rearrangement induced by the first binding event (because the second signal is observed at ligand/protein ratios where the population of the double-bound protein is low), but are not sensitive to the presence of the second ligand molecule. Residue numbers are indicated.

of the family, accompanied by further thermodynamic analysis and molecular dynamic simulations can lead to the design of appropriate mutations capable of providing a mechanistic insight into the determinants of co-operativity.

The comparison of the enthalpic contributions with binding of the characterized I-BABPs from human, rabbit, zebrafish and chicken, revealed that the protein described in the present study exhibited a much larger enthalpic change of -73.6 kJ · mol^{-1} on binding (Table 1). Interestingly a large enthalpy change of -109.5 kJ · mol^{-1} was recently measured for a mutant of rabbit I-BABP ($\Delta\alpha$-ILBP) [30] where the helical capping motif, proposed to play a key role in modulating the cavity size and regulating ligand binding, was replaced by a relatively short flexible Gly-Gly-Ser-Gly linker. This mutation had dramatic effects on ligand uptake and protein stability. Indeed it was suggested that strong coupling between folding and binding was responsible for the significant enthalpic change observed, since ligand binding to I-BABPs generally produced a significantly smaller exothermicity. In view of this result we suggest that the enthalpy change measured here may be ascribed to the presence of local folding accompanying the binding event. This conclusion is even more strongly supported by the largely negative entropic contribution (Table 1), indicating that the system is shifted to a more highly ordered state by ligand binding. It is then likely that missing NMR connectivities of residues in the EF strands and EF loop of the apo protein (Glu^{75}–Ala^{81}) are to be attributed to conformational averaging among states that include locally unfolded protein forms. Finally, because most of the unfavourable entropic and favourable enthalpic changes are associated with the first binding event, it is reasonable to conclude that the first ligand molecule

A

	cl-BABP	hl-BABP	rl-BABP	zl-BABP
cl-BABP	100			
hl-BABP	62	100		
rl-BABP	59	82	100	
zl-BABP	55	54	50	100

B

C

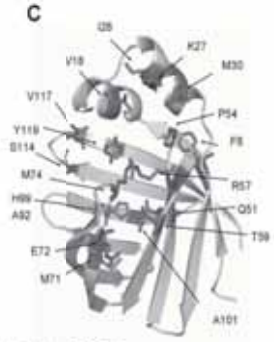


Figure 9 Sequence alignment of four ileal proteins of the BABP family

Sequence alignment of chicken I-BABP with human, rabbit and zebrafish I-BABPs. Conserved residues are highlighted in bold. Residues presenting non-conservative mutations in the four proteins pointing towards the ligand-binding cavity, as identified by the program CASTP (http://sts.bioengr.uic.edu/castp/) are indicated with ○. Residues underlined in the sequences are those reported to be involved in binding in human I-BABP on the basis of the first generation NMR structures, and zebrafish I-BABP, as obtained from the X-Ray structure (PDB code: 3EM0). cl-BABP, chicken I-BABP; hl-BABP, human I-BABP; rl-BABP, rabbit I-BABP; zl-BABP, zebrafish I-BABP.

promotes a major structural rearrangement, in agreement with the independent conclusion derived from the analysis of protein ^{15}N-H and ^{13}C-H CSPs.

A further remarkable difference displayed by chicken and human I-BABP concerns the site-selectivity. Indeed, human I-BABP has been found to exhibit a high degree of ligand site selectivity in its interactions with GCA and GCDA, a feature which appeared to be determined by localized enthalpic effects and could be removed by mutation of a glutamine residue in position 51 [12]. Competition experiments performed in the present study show GCDA selectivity for one binding site in chicken I-BABP, which is not affected by the mutation at position 51. The characterization of the binding properties of chicken I-BABP Q51A was performed to complete the binding comparison with human I-BABP.

It is interesting to note that the paralogous chicken L-BABP displays a binding mechanism which appears more similar to that of human I-BABP rather than chicken I-BABP. Indeed, a high degree of binding co-operativity was established by NMR investigations, although no site-selectivity was observed [16,33].

Conclusions

In summary, the analysis described in the present study points to a protein scaffold which is able to establish long-range communication networks through a significant conformational rearrangement induced by a first binding event. We conclude that chicken I-BABP can be considered as an allosteric system, which, however, does not manifest positive-binding co-operativity, as observed for the human protein. These considerations make chicken I-BABP a potentially invaluable model system for the understanding of co-operativity among BABPs and may provide clues about the development of this efficient feature during species evolution.

AUTHOR CONTRIBUTION

Henriette Molinari and Michael Assfalg designed the study and wrote the manuscript. Mara Guariento and Serena Zanzoni performed protein expression, purification and spectral assignment. Mara Guariento and Michael Assfalg designed and recorded NMR spectra. Dimitrios Fessas performed calorimetric experiments and subsequent analysis. Renato Longhi synthesized modified glycoconjugate ligands.

ACKNOWLEDGEMENTS

We thank the European Community (EU-NMR, Contract # RII3-026145) for access to the HWB (Henry Wellcome Building-NMR facility (Birmingham, U.K.), Università degli Studi di Verona and CIRMMP (Consorzio Interuniversitario di Risonanze Magnetiche di Metalloproteine Paramagnetiche).

FUNDING

This work was supported by the FIRB (Fondo per gli Investimenti della Ricerca di Base) 2003 [project number RBNE03PX83]; and the Fondazione Cariverona (Bando di ricerca scientifica 2007: Nuove applicazioni della Risonanza Magnetica Nucleare allo sviluppo di nano biotecnologie).

REFERENCES

1 Houten, S. M., Watanabe, M. and Auwerx, J. (2006) Endocrine functions of bile acids. EMBO J. **25**, 1419–1425
2 Nakahara, M., Furuya, N., Takagaki, K., Sugaya, T., Hirota, K., Fukamizu, A., Kanda, T., Fujii, H. and Sato, R. (2005) Ileal bile acid-binding protein, functionally associated with the farnesoid X receptor or the ileal bile acid transporter, regulates bile acid activity in the small intestine. J. Biol. Chem. **280**, 42283–42289
3 Dawson, P. A. and Oelkers, P. (1995) Bile acid transporters. Curr. Opin. Lipidol. **6**, 109–114
4 Guariento, M., Raimondo, D., Assfalg, M., Zanzoni, S., Pesente, P., Ragona, L., Tramontano, A. and Molinari, H. (2008) Identification and functional characterization of the bile acid transport proteins in non-mammalian ileum and mammalian liver. Proteins **70**, 462–472
5 Furuhashi, M. and Hotamisligil G. S. (2008) Fatty acid-binding proteins: role in metabolic diseases and potential as drug targets. Nat. Rev. Drug Discov. **7**, 489–503
6 Lücke, C., Zhang, F., Hamilton, J. A., Sacchettini, J. C. and Rüterjans, H. (2000) Solution structure of ileal lipid binding protein in complex with glycocholate. Eur. J. Biochem. **267**, 2929–2938
7 Kurz, M., Brachvogel, V., Mattei, H., Stengelin, S., Thuring, H. and Kramer, W. (2003) Insights into the bile acid transportation system: the human ileal lipid-binding protein-cholyltaurine complex and its comparison with homologous structures. Proteins **50**, 312–328
8 Tochtrop, G. P., Richter, K., Tang, C., Toner, J. J., Covey, D. F. and Cistola, D. P. (2002) Energetics by NMR: site-specific binding in a positively cooperative system. Proc. Natl. Acad. Sci. U.S.A. **99**, 1847–1852
9 Eliseo, T., Ragona, L., Catalano M., Assfalg, M., Paci, M., Zetta, L., Molinari, H. and Cicero, D. O. (2007) Structural and dynamic determinants of ligand binding in the ternary complex of chicken liver bile acid binding protein with two bile salts revealed by NMR. Biochemistry **46**, 12557–12567
10 Kouvatsos, N., Thurston, V., Bal, K., Oldham, N. J., Thomas, N. R. and Searle, M. S. (2007) Bile acid interactions with rabbit ileal lipid binding protein and an engineered helixless variant reveal novel ligand binding properties of a versatile β-clam shell protein scaffold. J. Mol. Biol. **371**, 1365–1377
11 Capaldi, S., Saccomani, G., Fessas, D., Signorelli, M., Perduca, M. and Monaco, H. L. (2009) The X-ray structure of zebrafish (*Danio rerio*) ileal bile acid-binding protein reveals the presence of binding sites on the surface of the protein molecule. J. Mol. Biol. **385**, 99–116
12 Toke, O., Monsey, J. D., DeKoster, G. T., Tochtrop, G. P., Tang, C. and Cistola, D. P. (2006) Determinants of cooperativity and site selectivity in human ileal bile acid binding protein. Biochemistry **45**, 727–737
13 Toke, O., Monsey, J. D. and Cistola, D. P. (2007) Kinetic mechanism of ligand binding in human ileal bile acid binding protein as determined by stopped-flow fluorescence analysis. Biochemistry **46**, 5427–5436
14 Ragona, L., Catalano, M., Luppi, M., Cicero, D., Eliseo, T., Foote, J., Fogolari, F., Zetta, L. and Molinari, H. (2006) NMR dynamic studies suggest that allosteric activation regulates ligand binding in chicken liver bile acid-binding protein. J. Biol. Chem. **281**, 9697–9709
15 Okazaki, K. and Takada, S. (2008) Dynamic energy landscape view of coupled binding and protein conformational change: induced-fit versus population-shift mechanisms. Proc. Natl. Acad. Sci. U.S.A. **105**, 11182–11187
16 Pedò, M., D'Onofrio, M., Ferranti, P., Molinari, H. and Assfalg, M. (2009) Towards the elucidation of molecular determinants of cooperativity in the liver bile acid binding protein. Proteins **77**, 718–31
17 Assfalg, M., Gianolio, E., Zanzoni, S., Tomaselli, S., Russo, V. L., Cabella, C., Ragona, L., Aime, S. and Molinari, H. (2007) NMR structural studies of the supramolecular adducts between a liver cytosolic bile acid binding protein and gadolinium(III)-chelates bearing bile acids residues: molecular determinants of the binding of a hepatospecific magnetic resonance imaging contrast agent. J. Med. Chem. **50**, 5257–5268
18 Tomaselli, S., Zanzoni, S., Ragona, L., Gianolio, E., Aime, S., Assfalg, M. and Molinari, H. (2008) Solution structure of the supramolecular adduct between a liver cytosolic bile acid binding protein and a bile acid-based gadolinium(III)-chelate, a potential hepatospecific magnetic resonance imaging contrast agent. J. Med. Chem. **51**, 6782–6792
19 Wyman, J. and Gill, S. J. (1990) Binding and Linkage, University Science Books, Mill Valley, CA
20 Gill, S. J., Robert, C. H. and Wyman, J. (1988) Biochemical Thermodynamics, 2nd edition (Jones, M.N., ed.), pp. 140–160, Elsevier Science Publishers B.V., Amsterdam
21 Castronuovo, G., Elia, V., Iannone, A., Niccoli, M. and Velleca, F. (2000) Factors determining the formation of complexes between α-cyclodextrin and alkylated substances in aqueous solutions: a calorimetric study at 25 degrees C. Carbohyd. Res. **325**, 278–286
22 Di Cera, E., Bassi, F. A. and Gill, S. J. (1989) Information theory and the analysis of ligand-binding data. Biophys. Chem. **34**, 19–28
23 Press, W. H., Flannery, B. P., Teukolsky, S. A. and Vetterling, W. T. (1989) Modeling of data. In Numerical Recipes, The Art of Scientific Computing, pp. 521–538, Cambridge University Press, Cambridge
24 Vranken, W. F., Boucher, W., Stevens, T. J., Fogh, R. H., Pajon, A., Llinas, M., Ulrich, E. L., Markley, J. L., Ionides, J. and Laue, E. D. (2005) The CCPN data model for NMR spectroscopy: development of a software pipeline. Proteins **59**, 687–696
25 Johnson, B. A. (2004) Using NMRView to visualize and analyze the NMR spectra of macromolecules. Methods Mol. Biol. **278**, 313–352
26 Frottin, F., Martinez, A., Peynot, P., Mitra, S., Holz, R. C., Giglione, C. and Meinnel, T. (2006) The proteomics of N-terminal methionine cleavage. Mol. Cell. Proteomics **5**, 2336–2349
27 Shen, Y., Vernon, R., Baker, D. and Bax, A. (2009) De novo protein structure generation from incomplete chemical shift assignments. J. Biomol. NMR **43**, 63–78

28 Wishart, D. S. and Sykes, B. D. (1994) The 13C chemical-shift index: a simple method for the identification of protein secondary structure using 13C chemical-shift data. J. Biomol. NMR **4**, 171–180
29 Shen, Y., Lange, O., Delaglio, F., Rossi, P., Aramini, J. M., Liu, G., Eletsky, A., Wu, Y., Singarapu, K. K., Lemak, A. et al. (2008) Consistent blind protein structure generation from NMR chemical shift data. Proc. Natl. Acad. Sci. U.S.A. **105**, 4685–4690
30 Rea, A. M., Thurston, V. and Searle, M. S. (2009) Mechanism of ligand-induced folding of a natively unfolded helixless variant of rabbit ILBP. Biochemistry **48**, 7556–7564
31 Kramer, W., Sauber, K., Baringhaus, K. H., Kurz, M., Stengelin, S., Lange, G., Corsiero, D., Girbig, F., Konig, W. and Weyland, C. (2001) Identification of the bile acid-binding site of the ileal lipid-binding protein by photoaffinity labeling, matrix-assisted laser desorption ionization-mass spectrometry, and NMR structure. J. Biol. Chem. **276**, 7291–7301
32 Svensson, B., Jonsson, B., Woodward, C. E. and Linse, S. (1991) Ion-binding properties of calbindin D9k: a Monte Carlo simulation study. Biochemistry **30**, 5209–5217
33 Tomaselli, S., Ragona, L., Zetta, L., Assfalg, M., Ferranti, P., Longhi, R., Bonvin, A. M. and Molinari, H. (2007) NMR-based modeling and binding studies of a ternary complex between chicken liver bile acid binding protein and bile acids. Proteins **69**, 177–191

Received 7 August 2009/9 October 2009; accepted 29 October 2009
Published as BJ Immediate Publication 29 October 2009, doi:10.1042/BJ20091209

Biochem. J. (2010) **425**, 413–424 (Printed in Great Britain) doi:10.1042/BJ20091209

SUPPLEMENTARY ONLINE DATA

Chicken ileal bile-acid-binding protein: a promising target of investigation to understand binding co-operativity across the protein family

Mara GUARIENTO*, Michael ASSFALG*, Serena ZANZONI*, Dimitrios FESSAS†, Renato LONGHI‡ and Henriette MOLINARI*[1]

*Dipartimento di Biotecnologie, Università degli Studi di Verona, Strada Le Grazie 15, 37134 Verona, Italy, †Dipartimento di Tecnologie Alimentari e Microbiologiche, Università degli Studi di Milano, Via Celoria 2, 20133 Milano, Italy, and ‡Istituto di Chimica del Riconoscimento Molecolare, Consiglio Nazionale delle Ricerche, Via Mario Bianco 9, 20131 Milano, Italy

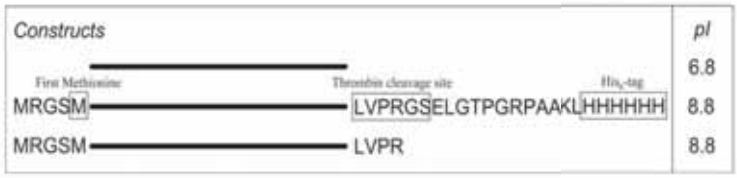


Figure S1 Scheme of the chicken I-BABP constructs and corresponding predicted isoelectric points

From the top: the original sequence of chicken I-BABP (from Ala^1 to Ala^{127}) is represented by a black bar; the long construct (as reported in [1]) with the short N-terminal tag preceding the first methionine residue and the long C-terminal tag containing the thrombin-cleavage site and an His_6-tag; the short construct (the present work) with the short N-terminal tag and the C-terminal tag cut after the first four residues.

[1] To whom correspondence should be addressed (email henriette.molinari@univr.it).

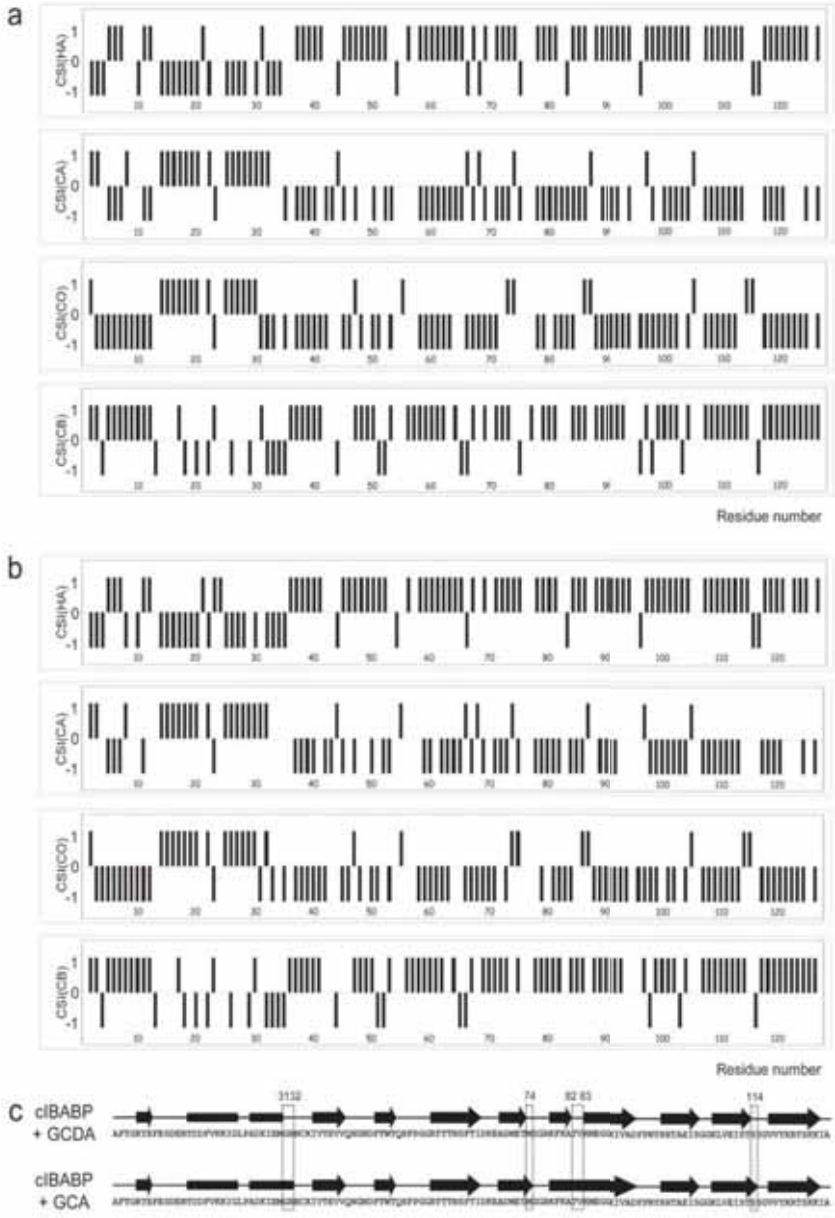


Figure S2 Histogram representation of the chemical-shift index (CSI)

CSI obtained for α-protons (HA), α-carbons (CA), β-carbons (CB) and carbonyl carbons (CO) of: (**a**) chicken I-BABP in complex with GCDA, 1:5 ratio, and (**b**) chicken I-BABP in complex with GCA, 1:5 ratio. (**c**) Consensus secondary structure assignments obtained combining the CSIs for the complexes of chicken I-BABP with GCDA and GCA. The residues exhibiting a difference between the two complexes are indicated.

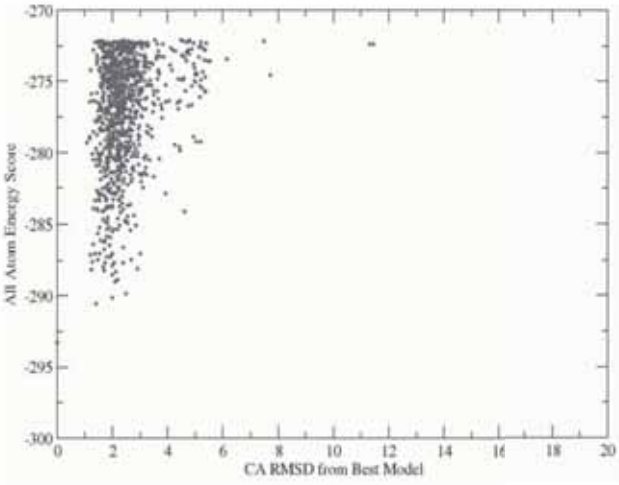


Figure S3 Plot of rescored Rosetta all-atom energy against Cα RMSD relative to the lowest-energy model (dot on vertical axis)

CA, cholic acid; RMSD, root mean square deviation.

Figure S4 Superposition of the ten lowest-energy CS-Rosetta models of holo chicken I-BABP

Notably, although the CS-Rosetta algorithm is not based on homology modelling but relies on chemical shifts data, the calculation predicts an I-BABP structure exhibiting 2.13 Å RMSD with respect to the homologous zebrafish I-BABP (PDB code: 3EM0).

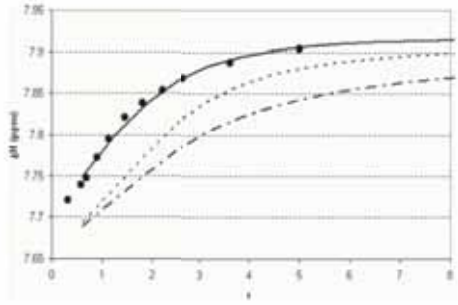


Figure S5 Evaluation of ITc-derived population fractions by ligand NMR chemical shifts

The lower field shifted ligand H_N signal labelled 1/U (Figure 4 in the main text) can be attributed to both unbound and bound ligand. The chemical-shift changes of the signal during the titration with chicken I-BABP can be measured with high precision and may be used to understand the nature of the observed signal. It is assumed that the ligand species are in a fast chemical-exchange regime and that he observed shift of the signal is a linear combination of the shifts of the unbound and some bound ligand. The latter could correspond to a molecule in a singly bound or a double-bound protein, or both. Ligand H_N chemical shifts for peak 1/U were simulated from linear combinations of the ITC-derived population fractions considering values of 7.62 and 7.92 p.p.m. for the bound and unbound forms respectively. It was also considered that the singly bound protein could exist in a variety of forms, not all of them contributing to the observed signal. In particular, after several simulation trials we considered that only 45% of the total single-bound protein could contribute to signal 1/U. The ligand in the remaining 55% singly bound protein may be bound less specifically or its signal be broadened by dynamic equilibria. The best agreement between ITC and NMR data was obtained considering that 1/U is produced by both unbound ligand and ligand present in approximately half of the singly ligated protein forms. The Figure shows the simulated shifts corresponding to a linear combination of the population fractions of unbound and: singly bound protein (PL); dashed line; sum of singly and double-bound protein (PL+PL$_2$; dashed/dotted line; and NMR observable singly bound protein (45% of PL); continuous line. ●, Experimental proton-shift changes of the H_N ligand signal 1/U.

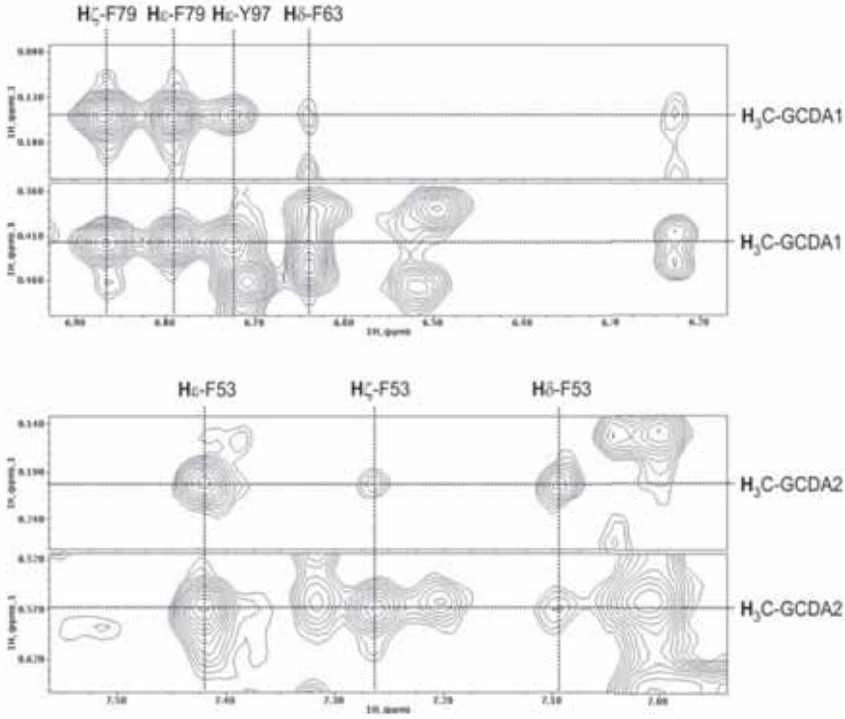


Figure S6 Selected regions of two-dimensional ^{1}H-^{1}H-NOESY (nuclear Overhauser enhancement spectroscopy) acquired at 600 MHz, 25°C, on a sample containing cl-BABP–GCDA at a 1:4 ratio in 30 mM P_i, pH 6.5

The spectra highlight the observed intermolecular NOE cross-peaks in the adduct involving aromatic residues. The chemical shifts of the methyl groups of the two bound ligands have been determined from the transformed first find of a F1-edited, F3-filtered three-dimensional HMQC-NOESY experiment where only resonances of protons xound to C-12 and exhibiting NOE with protein protons are selected. The assignment of methyl resonances to either ligand molecule was based on common NOE patterns.

REFERENCES

1 Guariento, M., Raimondo, D., Assfalg, M., Zanzoni, S., Pesente, P., Ragona, L., Tramontano, A. and Molinari, H. (2008) Identification and functional characterization of the bile acid transport proteins in non-mammalian ileum and mammalian liver. Proteins **70**, 462–472

Received 7 August 2009/9 October 2009; accepted 29 October 2009
Published as BJ Immediate Publication 29 October 2009, doi:10.1042/BJ20091209

Chapter 5: NMR investigation of the structural correlates of cooperativity in bile acid binding proteins

Short presentation of publication 3

In this study, we have determined the NMR solution structure of the chicken ileal BABP (cI-BABP) in complex with two molecules of glycochenodeoxycholic acid (GCDA) using NMR derived data in combination with the HADDOCK modeling procedure.

The extensive overlap of the resonances of the two bound identical ligands represents a real challenge for the structure determination of the complex. A wide variety of NMR experiments, including doubly filtered and edited-filtered schemes were performed to identify the intermolecular restraints for the structural determination of the complex. The employed HADDOCK procedure allowed to follow an iterative procedure for extracting the necessary intermolecular restraints to obtain the complex structure with high accuracy.

A comparison of the obtained ternary complex with the only available ileal BABP structure from zebrafish, shows that the internal ligand (GCDA-II) has a very similar orientation in the two proteins, with the major differences involving the external ligand GCDA-I.

A comparative structural analysis coupled to a bioinformatics screening extended to the other known members of the I-BABP family allowed us to identify a key interaction capable of affecting the binding mechanism, introducing a gain of cooperativity, as shown by the analysis of the produced mutant.

My contribution to this work consisted in the planning and analysis of the experiments, in the structure determination of ternary complex and in subsequent comparative analysis of different proteins of the I-BABP family.

NMR INVESTIGATION OF THE STRUCTURAL CORRELATES OF COOPERATIVITY IN BILE ACID BINDING PROTEINS*

Serena Zanzoni[1], Michael Assfalg[1], Alejandro Giorgetti[1], Mariapina D'Onofrio[1] and Henriette Molinari[1]

Dipartimento di Biotecnologie, Strada Le Grazie 15, Università di Verona, 37134 Verona, Italy.

Address correspondence to Henriette Molinari, Phone: +39 0458027901; Fax: +39 0458027929;

e-mail henriette.molinari@univr.it

Ileal bile acid binding proteins (I-BABP) belong to the family of intracellular lipid binding proteins and play an important role in transport and release, within the cytoplasm, of bile acids, thus controlling bile acid homeostasis in the enterohepatic circulation. We have previously shown that chicken I-BABP forms a ternary complex with two bile salt molecules and we have investigated the thermodynamics of binding in comparison with other proteins of the family. The protein under study, namely chicken I-BABP, represents a suitable model for the mechanistic investigation of cooperativy as it contains a protein scaffold which is able to establish a long-range communication network, as reflected by the significant conformational rearrangement induced by the first binding event. This behavior indicates that the protein behaves as an allosteric system, which, however, does not display binding cooperativity. The question addressed by the present work relates to whether we are capable of introducing mutations that lead to the gain of a function, namely cooperativity. Thus NMR derived data together with HADDOCK were employed to initially derive the structure of the ternary complex with the aim of correlating structural features with the level of cooperativity exhibited by different proteins of the family. This extensive analysis allowed us to single out a key interaction capable of affecting the binding mechanism.

Intracellular bile acid binding proteins (BABP) belong to the fatty acid binding protein (FABP) family that is composed of abundantly expressed small-molecular-mass (14-15 KDa) β-barrel proteins, presumably acting as lipid chaperones (1-3). Specifically, BABPs were shown to play a pivotal role in the transcellular trafficking and enterohepatic circulation of bile salts (4).

During the last decade a growing amount of evidence has been presented that the various intracellular lipid binding proteins (iLBPs) interacting with lipids not only facilitate their transport in

aqueous media but, by interacting with specific targets, also modulate their subsequent biological action or metabolism. Therefore, the notion arises that for better understanding the action of lipids as signalling compounds and the role of lipids in intermediary metabolism, it is essential to have detailed knowledge on the interactions between lipids and their cognate binding proteins (5).

The binding capability and mainly the binding mechanism of these proteins appear difficult to be captured, in spite of many attempts in this direction (6-9). It has been definitively ascertained that the proteins of the family can form at least ternary complexes with their putative binders, with two ligand molecules located inside the protein cavity (10). In few cases, namely for zebrafish (11) and rabbit ileal BABPs (12), a higher binding stoichiometry was detected by X-ray crystallography and mass spectrometry, respectively, with one or two additional ligands located on the protein surface. However, the most intriguing feature identified for some members of the family relates to macroscopic functional characteristics, such as cooperativity and site-selectivity (8). Human ileal BABP and chicken liver BABP, the two most widely studied members of the family, were shown to display an extraordinary degree of positive cooperativity, translated, for example, in a Hill coefficient close to 2 (13). Cooperativity and allostery are two closely related phenomena, but the exact biophysical characterization of allosteric communication remains in most cases a major challenge. This is particularly true for the protein family under study, also due to the lack of extensive structural characterization. Indeed within the ileal protein subfamily (I-BABP), which appears as a well-separated group from our phylogenetic analysis, only the structure of holo zebrafish with two cholate molecules located in the internal cavity is known. Data relative to the binding stoichiometry and thermodynamics were reported for the rabbit ileal protein (12). As far as the human form is concerned, hI-BABP, the reported holo structure (hI-BABP/GCA) (14) is bound to a single ligand, as it was obtained at temperatures where one of the two ligands could not be observed due to exchange broadening of the NMR signals. Later NMR and calorimetric studies performed on hI-BABP and on a series of mutants, for the first time addressed the issue of possible communication pathways between the two binding sites(9). However this analysis was hampered by the lack of the three-dimensional structure of the ternary complex. Along this line of investigation, we have reported an interaction study on another member of the I-BABP family, chicken I-BABP (cI-BABP), with the aim of enlarging the knowledge on this family and derive information on the binding mechanism (15,16). This protein revealed itself a very good model

for this kind of studies. Indeed ITC data, supported by NMR titration experiments, allowed establishing a thermodynamic binding model describing two independent consecutive binding sites. A singly bound protein was found to be relatively abundant at low protein/ligand ratio thus assessing the absence of cooperativity, differently from the human protein. The measured energetics of binding and the analysis of chemical shift perturbations suggested that the first binding event was capable of triggering a global structural rearrangement. Consistent with this finding, also the enthalpic and entropic contributions associated with the binding of the first ligand indicated an increased stability and order of the bound protein. In line with the classical definition of allostery, implying that the entire molecule can undergo a conformational transition even when it is only partially ligated (17), we concluded that I-BABP behaves as an allosteric system that does not display positive cooperativity. Thus it represents a suitable and attracting model to investigate the molecular basis for a gain-of-function on going from non-mammalian to mammalian species. To this aim we first address here the 3D NMR structural determination of cI-BABP in complex with two glycochenodeoxycholate (GCDA) molecules. This is a challenging NMR structural determination due to the extensive overlap of the resonances of the two identical ligands in the complex. In this contest we exploited the potentiality of the HADDOCK approach (18), allowing for an iterative procedure for extracting an increasing number of intermolecular restraints, to derive the 3D structure of the ternary complex. A comparative structural analysis coupled to a bioinformatics screening extended to the other known members of the I-BABP family allowed us to identify a key interaction capable of affecting the binding mechanism, introducing a gain of cooperativity, as shown by the analysis of the produced mutant.

Experimental procedures

Sample preparation. The expression plasmid for single (A101S) and double mutant (A101S/H99Q) was obtained from that of wild-type chicken I-BABP using the Quickchange mutagenesis kit. The presence of the desired mutations was confirmed by plasmid sequencing. Recombinant wild type and mutant I-BABPs were expressed as described previously using an *Escherichia coli*-based expression vector (16). Uniformly, $^{15}N^{13}C$-labeled protein was produced by growth in M9 minimal medium with ^{15}N labeled ammonium chloride and ^{13}C-labeled glucose as the only nitrogen and carbon sources, respectively. Protein concentrations for sample preparation were determined spectrophotometrically measuring the absorbance at 280 nm. Unenriched bile salts (GCA and GCDA) were purchased from

Sigma-Aldrich. ^{15}N-Glycine conjugates of CDA and CA were prepared as previously reported (15).

NMR Spectroscopy. NMR samples were prepared in 30 mM sodium phosphate buffer, pH 6.5, containing 0.05% NaN_3 and 90%/10% H_2O/D_2O or 99%D_2O. NMR spectra were acquired at 25 °C with an AvanceIII Bruker spectrometer, operating at 600.13 MHz and equipped with a 5 mm TCI cryoprobe and Z-field gradient. The NMR samples for the structure determination contained 0.8 mM $^{15}N^{13}C$ cI-BABP in complex with unlabeled GCDA at a protein:ligand ratio of 1:4. Backbone sequential resonance assignment has been previously reported (19). Side chain resonances assignment was achieved for almost all atoms based on (H)CCH-TOCSY, H(C)CH-TOCSY. Typically a SW of 8417.5, and 11312.2 was employed in the proton and carbon dimensions, respectively; 2048, 128 and 256 points were employed in F3, F2 and F1, respectively. A 3D NOESY-[^{1}H,^{15}N]-HSQC and two 3D NOESY-[^{1}H,^{13}C]-HSQC (one in H_2O and one in D_2O) were recorded to obtain intramolecular NOE-based distance restraints for the protein in the holo form.

An F_1/F_2-[^{15}N,^{13}C]-filtered NOESY experiment with mixing time of 70 ms was run to filter out the ^{1}H resonances of doubly labeled protein allowing the resonances assignment of ligands in the bound form. In order to detect intermolecolar NOEs between the labeled protein and unlabeled ligands, two 3D F_1-[^{13}C]-filtered, F_2-[^{13}C]-separated, F_3-[^{15}N,^{13}C]-edited NOESY spectra (one optimized for aliphatic and one for aromatic residues) were recorded in D_2O with mixing time 120 ms.

^{1}H-^{15}N HSQC spectra were employed in titration experiments where a 0.4 mM sample of ^{15}N labeled protein (either wild type or mutants) were treated with increasing amounts of ligand. The following protein:ligand ratios were analysed: 1:0.3, 1:0.6, 1:1, 1:2, 1:2.2, 1:4, 1:5. The spectra were acquired with a spectra width of 7214 Hz and of 973 Hz in the ^{1}H and ^{15}N dimensions, respectively. 2048 and 80 complex points were collected in F2 and F1 dimensions, respectively, employing standard pulse programs from the standard Bruker library. NMR data were processed with Topspin 2.1 (Bruker) and analyzed with NMRView software (20).

Protein Structure calculation. Intramolecular NOE cross-peaks identified and integrated in the 3D NOESY-HSQC spectra and backbone dihedral angle constraints derived from secondary chemical shifts, using TALOS+ program (21) were used as input in the protein structure calculations.

The structure of cI-BABP in the holo form was calculated using the torsion angle dynamics program CYANA 2.1 (22). The CANDID module of CYANA (23) was used for automated assignment of the NOE cross-peaks, followed by a manual check prior to the final

calculations. In the final round of CYANA refinement, a total of 2738 upper bound distance restraints and 222 dihedral angle restraints were used. The 20 best structures of the CYANA ensemble, in terms of target function, were chosen to represent the final family.

Structure determination of the protein-ligand complex. The structure of the ternary complex between cI-BABP and two molecules of GCDA was performed using the software HADDOCK 2.1 (24) in combination with CNS (25). HADDOCK can dock two or more molecules together, incorporating both unambiguous (NOE, hydrogen bond, RDC, dihedral angle) and ambiguous (chemical shift perturbation) restraints. The 10 best structures of the final bundle calculated by CYANA were used for the docking calculation as starting protein structures. GCDA coordinates were derived with the SMILE program (26) by adding a glycine residue to a CDA molecule. Topology and parameter files of the ligands were generated by the PRODRG server (27). The protein was kept fully flexible during docking, the ligands semi flexible. As such, both backbones and side chains of the protein are allowed to move in the calculation. The following restraints were used as input for the docking calculation: (i) ambiguous interaction restraints derived from chemical shift mapping data (16) and isotope-filtered NMR experiments; (ii) dihedral restraints (to maintain the secondary structure of the protein); (iii) unambiguous intra-residue NOEs (to ensure that the protein structure in the complex satisfies the experimentally derived intramolecular distances); (iv) intermolecular NOE-derived distances defined as upper distance limits of up to 6.0 Å between protein and ligand carbons.

The docking protocol consists of three distinct stages: randomization of the relative orientations and energy minimization, semi-rigid simulated annealing of the protein and ligands in torsion angle space, and finally refinement of the complexes in Cartesian space with explicit solvent. During the rigid body docking, 4000 structures were calculated. The best 400 solutions were selected for a fully-flexible simulated annealing, followed by water refinement. Electrostatic and van der Waals terms were calculated with a 8.5 Å distance cutoff using the OPLS non bonded parameters from the parallhdg5.3.pro parameter file (28).

The resulting solutions were clustered using the algorithm of Daura (29) with a 0.5 Å cut-off. The structures were divided in twelve clusters and the best twenty structures were selected for each cluster. According to the HADDOCK score we defined a final bundle of twenty structures. The protein-ligand contacts were analyzed using the software LIGPLOT (30) and the r.m.s.d. referred to ligand coordinates was calculated after all-atom fitting with PROFIT software (Martin, A. C. R.,

Profit http://www.bioninf.org.uk/software/profit.

RESULTS AND DISCUSSION

Solution structure determination of holo cI-BABP. Two-dimensional ^{1}H-^{15}N heteronuclear correlation spectra (HSQC) of uniformly ^{15}N labeled protein samples provide a fingerprint of the local environment of backbone amides and may give detailed insight into the structure and dynamics of a protein. Three-dimensional triple-resonance experiments are used to extend these amide resonances via their covalently linked nitrogen into a third (carbon) dimension which finally leads to sequence assignments for the residues responsible for the original HSQC spectrum. On the basis of the standard set of 3D experiments (see Experimentals), a complete backbone assignment was previously obtained and reported for cI-BABP in complex with glycochenodeoxycholic acid (GCDA) and glycocholic acid (GCA) (16). We have now focused on the structure determination of the complex cI-BABP/GCDA, in order to investigate, at a molecular level, its binding characteristics, and to extend this analysis to a series of homologous holo proteins.

The side-chains assignment of the complex was achieved with a set of 3D (H)CCH-TOCSY, and H(C)CH-TOCSY experiments while 3D NOESY-[^{1}H,^{15}N]-HSQC and 3D NOESY-[^{1}H,^{13}C]-HSQC were recorded to obtain intramolecular distance restraints. NOE cross-peak intensities were obtained by manual peak picking each strip of the 600 MHz 3D ^{15}N NOESY-HSQC spectra taken at the assigned ^{1}H,^{15}N amide resonances as well as in each strip of the 3D ^{13}C NOESY-HSQC spectra taken at the assigned ^{1}H,^{13}C frequencies, optimized for aliphatic and aromatic resonances, respectively.

The solution structure of ileal-BABP was then derived from 2738 distance constraints, 221 angle constraints from chemical shift analysis by TALOS+ software (21). The total of 2959 constraints was used in generating 100 refined structures with the CYANA 2.1 software, and the best 20 conformers, those with lowest energy that showed the fewest constraint violations, were chosen to represent the solution structure of ileal-BABP (Figure 1).

The r.m.s.d. of 0.49 ± 0.08Å for backbone heavy atoms and 1.07± 0.08 Å for all heavy atoms, 86% of the backbone angles in most favoured, 14% in additionally favoured regions of the Ramachandran plot, and the Procheck G-factor for backbone/all dihedral angles of -0.62/-0.72 are all parameters indicative of a high quality solution structure. All the statistics and the employed tools for structure validation are reported in Table 1.

Identification of intermolecular

restraints. We have previously shown, on the basis of NMR and calorimetry data that the protein forms a ternary complex with two ligand molecules. One of the most difficult problems to be solved for a ternary complex like that of cI-BABP/GCDA is the assignment of the two identical bound molecules of the unlabelled bile acids. The chemical shift dispersion of protons belonging to GCDA is very poor, but the presence of three methyl groups (C18, C19 and C21) which appear as two high intensity singlets at relatively high fields (reported chemical shifts for C18 and C19 are 0.67 and 0.93 ppm, respectively) and as a doublet (C21, 0.95 ppm) (Figure 2 a,b) constituted a good entry point for the assignment of the bound molecules. 2D F_1/F_2-[^{15}N,^{13}C]-filtered NOESY experiments, performed on a cI-BABP/GCDA complex at different protein:ligand ratio (from 1:1 to 1:4) were used to eliminate ^{1}H signals belonging to the protein, which is ^{15}N and ^{13}C labelled, thus giving the possibility of observing only the unlabelled bound and unbound ligands. Although it is suggested that a sample requirement for isotope-filtered experiments is that the partners should be present in near-stoichiometric ratio, experiments at different protein:ligand ratios were employed in the present study, to keep track of the different species present at each r ratio. At P:L 1:4 three new broad signals occurring at 0.15, 0.19 and 0.42 ppm were thus identified (see Figure 3a for the 1:4 P:L complex). They exhibited NOESY exchange peaks with intense signals at the same chemical shift of free GCDA (in excess in the complex), whose resonance assignment is known (31). Thus the assignment of methyl signals of one bound GCDA molecule (hereby called GCDA-II) was obtained (Figure 3a). Specifically the identification of C19 and C21, nearly superimposed in the free ligand, was possible on the basis of the double filtered TOCSY experiment of the complex, where the only methyl residue exhibiting a scalar correlation can be that corresponding to C21 (scalarly coupled to H20 and H17) (Figure 3b). However, if two molecules of GCDA are bound to cI-BABP, then a total of six different methyl signals are expected in this region, in addition to those of free GCDA. As shown in Figure 3b it is likely that the resonances belonging to the second bound ligand are superimposed to those of free GCDA. Further experiments confirmed that at P:L ratio 1:4 the resonances of the ligand bound at site I are coincident with those of the free ligand: i) ^{1}H-^{15}N HSQC spectra obtained from the complex of the unlabelled protein with ^{15}N labelled GCDA showed a unique resonance at 7.89 ppm for the amide group of the free ligand and of the ligand bound to site 1; ii) HACA and HNCA 3D spectra obtained from the complex of unlabelled protein with the ^{15}N,^{13}C glyco-conjugated bile salt,

exhibited a unique Cα chemical shift for the free and bound (to site 1) ligand. (Figure 1 of Supplementary Material). Titration experiments obtained upon addition of increasing amount of ^{15}N labelled GCDA to the unlabelled protein showed that the amide of ligand bound to site I exhibits fast exchange with free GCDA and move towards the chemical shift of free GCDA amide during the titration (Figure 2 of Supplementary Material). We reasoned that in the presence of an excess of ligand, exchange phenomena with the free ligand are enhanced, thus allowing for the assignment, while at lower protein/ligand ratio the sets of bound methyl groups in both singly and doubly ligated complexes should be the prominent species and therefore more easily distinguishable. Indeed, we have previously shown that the ITC-derived model, implying the presence of two consecutive binding sites, predicted a considerable concentration of singly ligated protein in a wide range of ligand–protein molar ratios, accounting for approximately 50 % of all protein in the range of 1:1 <P:L<1:2 (19). However, ^{1}H-^{13}CHSQC protein spectra obtained at increasing amounts of ligand, showed the appearance of only two sets of ^{13}C separate signals. Thus, in those spectra, the singly bound protein was either invisible in the employed experimental conditions, or its signals coincident with one of the two sets of resonances, as was established by comparing the sum of the population fractions for both the unbound and singly bound forms and for both the singly and doubly bound forms with the signal intensities of the two set of ^{13}C signals. The following analysis performed at different P:L values allowed for the possible identification of the singly ligated species. Indeed at P:L 1:2 (Figure 4) an intense cross-peak in the 2D F_1/F_2-[^{15}N,^{13}C]-filtered NOESY experiment connected the resonances at 0.42 and 0.57 ppm. We hypothesized that the methyl resonance of a singly ligated species (indicated in the Figure as 18sb) overlaps with the assigned C19 of GCDA-II. Thus this cross-peak represents an exchange cross-peak with the resonance occurring at 0.57 ppm, corresponding to C18 free ligand superimposed to C18 of GCDA-I. In Figure 4 the resonances of the other methyl groups corresponding to the ligand bound to site 1, to the singly bound species and to the free ligand are indicated. This assignment is confirmed by the fact that upon increasing the P:L ratio this cross-peak disappears. Indeed on going from P:L 1:1 to P:L 1:5 the population of the singly ligated species decreases and the resonances of the coincident free and bound-to-site-1 methyl groups move downfield towards the free ligand chemical shift (the ligand at site 1 is in fast exchange with the free ligand) (see Figure 3 of Supplementary Material).

Complementary NMR experiments, namely F1-edited, F3-filtered 3D HMQC-NOESY (32), with mixing times ranging from 0 to 200 ms, were then performed, in order to specifically detect signals of bound GCDA and eliminate those of the free form. In these experiments, only ^{1}H signals of GCDA experiencing NOE with the protein can be detected. All the parameters of the experiments were optimized to minimize the signals in the region 0-1 ppm at zero mixing time, where all the resonances should be filtered out. A superposition of the high field region of the spectra at P:L 1:2 obtained from the first FIDs at all mixing times is reported in Figure 5, clearly showing all the methyl groups (labelling is reported in the Figure) of the doubly and singly ligated species. All the spectra experience increasing NOE effects upon increasing the mixing time. All these data indicate that simplified spectra are obtained at higher P:L ratio due to the presence of a higher population of the doubly ligated species, and therefore the whole set of 3D experiments aimed at obtaining the assignment of the holo protein were performed at P:L ratio 1:4. Other resonances, beyond those of the bound methyl groups, could be unambiguously assigned at P:L 1:4 ratio, namely the resonances of H7 (H7 GCDA-I 3.85 ppm, H7 GCDA-II 3.89 ppm), as they occur outside of the bulk of the ligand resonances. H3 of GCDA-1 could be unambiguously identified at 3.46 ppm, as well as H25 of the bound species. These assignments allowed the identification of several intermolecular NOEs in the 3D F_1-[^{13}C]-filtered, F_2-[^{13}C]-separated, F_3-[^{13}C]-edited NOESY-HSQC spectra (Figure 6). The number of the identified intermolecular constraints was initially rather limited (35) and not devoid of ambiguities, as the chemical shifts of GCDA-I C19 and C21 are substantially superimposed.

As a record, it should be stressed that an analysis of the singly ligated species occurring at protein:ligand ratio 1:2 was hampered by the complexity of the spectra under examination. It is however worth mentioning that intermolecular NOEs were detected between the resonance occurring at 0.57 ppm (assigned to a singly ligated species) and residues located at the protein open end (30, 31, 34, 36, 53), thus suggesting that a singly bound species would involve the more superficial binding site, as previously shown for homologous proteins (33,34). At the same time the available data may not rule out that such interaction involves the protein surface giving rise to an adsorbed intermediate which later mature by migration of the ligand into a more specific binding site.

Analysis of the histidine tautomeric state. Additional experiments were run to investigate the histidine protonation state in the present experimental conditions. We have already pointed out the role of buried histidines in the binding

mechanism in other proteins of the family (13,35). To identify possible H-bonds between histidines and bound ligands it is relevant to know which tautomeric form is present, so to introduce the correct information on the histidine state in the computation procedure. Indeed the non-exchangeable Hδ2 and Hε1 protons, linked to nitrogen via weak two and three-bond couplings, can be detected in ^{1}H-^{15}N HMBC experiments (36-38). Specific assignment of ^{1}H and ^{15}N resonances of H52 and H98 were thus obtained, coupling the information derived from ^{1}H-^{15}N HMBC with ^{1}H-^{13}C HSQC-CT experiments, acquired in constant time (which removes the homonuclear coupling) and optimised for the aromatic ^{1}H-^{13}C coupling. In these experiments C4-Hδ2 correlations of histidine residues have negative sign (together with C2-H2 of tryptophan residues) and are thus easily identified. (Figure 4 of Supplementary Material). On the basis of the observed pattern and ^{15}N and ^{1}H shifts it appears that H52 and H98 are protonated and charged, as they exhibit a δN value of 173.2 ppm, expected for type α+ ^{15}N. Their Hε1 and Hδ2 protons are also downfield shifted, as expected for charged protonated histidines. Interestingly, only the resonances of H52 and H98 are detectable in ^{1}H-^{15}N HMBC spectra, while H99 resonances are not present, possibly due to exchange broadening between two tautomeric forms (N-Hε2 and N-Hδ2), occurring on a time scale comparable to the expected chemical shift difference between type α and type β nitrogens (82 ppm). C4-δ2 resonance of H99 is however present in the ^{1}H-^{13}C HSQC-CT spectrum, with Hδ2 occurring at 6.09 ppm, consistent with the presence of a neutral histidine. Additionally ^{1}H NMR spectra revealed the presence of an NH resonance at 12 ppm, which could be unambiguously assigned to the proton directly attached to the imidazole ring nitrogen (as deduced from ^{1}H NOESY spectra showing the correlation of N-Hε2 with Hε1 at 7.67ppm and Hδ2 at 6.09 ppm). NHε2 resonance is usually lost as it exchanges too quickly with water proton. Thus the unusual presence of N-Hε2 resonance confirms that H99 is deeply buried in the protein cavity, possibly involved in H-bonds with the ligand.

Structure calculation of the ternary complex between cI-BABP and two molecules of GCDA. The procedure selected for structure calculation employed Haddock (18), a robust data driven modelling program. The ability of Haddock in exploring the protein-ligand conformational space has been widely recognized, however only a limited number of studies are present in the literature where this approach is employed for the identification of a complex between a protein and small ligands (24,39-41). In our approach, to ensure success of the docking procedure, the starting structure for ligand docking was the calculated

structural ensemble of the holo protein. Therefore 2738 intramolecular NOEs were defined as non-ambiguous interaction restraints while 221 dihedral angles derived from secondary chemical shifts were employed to maintain the secondary structure of the protein. To identify the interaction surface, the average chemical shift changes between the free and the bound state, as derived from a series of $^{1}H,^{15}N$-HSQC spectra collected upon progressive addition of GCDA, were plotted versus the cI-BABP residue numbers (Figure7). Resonances with significant chemical shift perturbation (CSP), deviating more than two standard deviation from the mean CSP, included residues 53, 57-59, 97, 98 indicating that the interaction of cI-BABP with GCDA clearly affects residues in CD loop and in the H β-strand. These residues were therefore categorized as "active" and assigned as ambiguous interaction restraints (AIRs), which require their interaction with the substrate. Intermolecular NOEs, derived from edited-filtered experiments, were all used as ambiguous restraints (AIRs) in the first Haddock run. For these last constraints no calibration of the peak volumes was performed, as the filtered experiments do not provide signals corresponding to known distances, and an upper limit distance of 6 Å was employed. Histidines were defined as charged or non charged as mentioned in the previous paragraph. Subsequent rounds of Haddock procedure coupled with spectral analysis led to the identification of 48 non-ambiguous constraints between the ligands and the protein from which the structure of the complex was derived. A complete list of ambiguous and non-ambiguous restraints is given Table 2. A total of 400 structures of the complex were selected after rigid body docking and subjected to optimization by fully flexible simulated annealing, followed by refinement in explicit water. These structures were then clustered according to their pairwise root-mean- square deviation values, and the twenty lowest energy structures in the lowest energy cluster were selected. Statistics of the cluster with the best average HADDOCK score are presented in Table 3 while Figure 8 displays the HADDOCK model that represents best the structure of the complex. The twenty best structures in the selected cluster have an average backbone RMSD value from the mean of 0.044 ± 0.09 Å and 0.95 ± 0.10 Å for the backbone and all heavy atoms, respectively. The average r.m.s.d of the ligands, obtained with Profit, were 1.15 Å for GCDA-I and 0.72 Å for GCDA-II. The protein-ligands interface was analysed using Ligplot (30) (Figure 9).

Structural comparison with other proteins of the family. Only another ileal BABP structure from zebrafish forming a ternary complex with bile acids (pdb 3ELZ) is available for structural comparison. Few additional data

are available relative to the thermodynamics of binding for human (10), zebrafish (11) and rabbit ileal BABPs (12,42). The X-ray structure of the zebrafish protein shows that in two out of the three protein molecules present in the asymmetric unit, there are three cholate molecules in approximately identical position bound on the surface of the protein, in addition to the two internal ligands. We have previously shown, on the basis of NMR and calorimetry data that cL-BABP displays a 1:2 stoichiometry (13,43). We will therefore focus on the comparison of the internal binding pocket. Superposition of ileal zebrafish BABP (3ELZ) with the cI-BABP here reported, yielded r.m.s.d. values of 1.94 Å and indicated that helix-II, EF and IJ loops in zebrafish are substantially different from the chicken protein. These structural elements define the protein open end, which can determine the orientation of the more external ligand (GCDA-I) and eventually influence the binding cooperativity. Indeed, while the internal ligand (GCDA-II) has a very similar orientation in the two proteins, the external GCDA-I exhibits different orientations. To better analyse the surface binding cavities, Ligplot results obtained for zebrafish and ileal BABPs were compared (Figure 9). When the internal GCDA-II is considered, it appears that in the chicken protein, in at least 15 structures of the 20 analysed, favourable hydrophilic interactions (H-bonds) are formed between the carboxylate side chain of the ligand and residues T73, G75 and R77. Residues 73, 75 and 77 (including EF loop) are conserved in the entire protein family (residue 77 is R in the chicken ileal protein and lysine in all the remaining members of the family). In zebrafish holo protein the bound ligands are cholate molecules (CA), exhibiting shorter side chains terminating at C24 and an additional hydroxyl group in position 12. Internal cholate shows hydrophilic interactions between C24 carboxylate and V74 (EF loop), OH in position 12 and Y97 (strand H), OH in position 7 and E110 (strand I). A common anchoring point for the internal ligand in the two proteins is thus located at the level of EF loop. GCDA-I, in the chicken protein, has two H-bonds between the carbonyl group of C24 and the side chain of K27, the terminal carboxylate and the backbone amide of R57. In the zebrafish H-bonds involve OH7 and the side chain of Y53, OH12 and the side chain of Y14. When the hydrophobic interactions are compared, the internal cholate in zebrafish establishes hydrophobic interactions with Y97, F79, T73, M71, F63, N61, W49 and the external ligand with V74, Y53, G31, I23, and I21. The same analysis performed for the twenty Haddock structures of cI-BABP indicated that residues involved in hydrophobic interactions (present in at least 18 structures out of twenty) with GCDA-II are H99 (either H or Q in the family), Y97,

F94 (either F or S), A92, V83, F79, R77, G75, T73, M71, F63, N61, W49 and with GCDA-I are G75, M74, F53, M30, K27, L23 (either L or I), I21 (either I or L). The underlined residues are those conserved throughout the family. This analysis indicates that substantially the same residues are involved in the interaction with the internal ligand, though a more extended hydrophobic environment for GCDA-II is present in cI-BABP, consistent with the increased hydrophobicity of GCDA with respect to cholate. This analysis shows that the more pronounced differences in binding interface involve the environment of the external ligand.

Structural correlates of cooperativity: phylogenetic analysis of iLBP family. Sequence retrieval of the entire iLBP family was performed using both BLAST and Ssearch programs over the UNIPROT database. A multiple sequence alignment of the family was then obtained with the PROMALS program. The obtained alignment was then funnelled through the procedure of phylogenetic tree building using the MEGA program. Different approaches were used to the construction of the phylogenetic trees from the multiple sequence alignment, i.e. like Neighbour Joining, Minimum Evolution, UPGMA Maximum Parsimony and Maximun Likelihood, and an afterwards tree comparison was performed taking into account the work by Schaap (44). Figure 10 reports the obtained tree of the entire family. The classification has been done with a tissue specific criterion and each cluster contains proteins coming from the same tissue but from different species. A detailed analysis of the clusters indicates that the tissue-specialization of the proteins occurred before speciation. When the subfamilies of interest in the present work, i.e. liver FABPs, liver BABPs, and ileal BABPs are considered, it appears that the three groups are well separated. Interestingly the liver FABP group includes only mammalian species, the liver BABP group includes only non-mammalian species while the ileal BABP group includes both mammalian and non-mammalian species. Thus, for the analysis of the cI-BABP and its putative functional aminoacids, we concentrated on its closests relatives, that is the I-BABP family. In order to investigate the determinants of cooperativity between binding sites, detailed structural data, as those provided here, are of fundamental importance, as only the analysis of the sequences and functional conservation of certain features may allow the characterization of the key residues in the binding mechanism. This approach should further allow the design of mutations able to produce a gain of function, i.e. moving from a non-cooperative to a more cooperative protein. Figure 11 reports the sequence alignment of chicken, human, rabbit and zebrafish I-BABPs. As it is not straightforward to correlate the

occurring substitutions with the trends in binding cooperativity, we first performed a comparison of 3D-structures from chicken (non cooperative binding) and human (high cooperative binding) proteins. The human holo structure reported in the PDB (1O1V) is bound to a single ligand, as it was obtained at temperatures where the observation of one of the two ligand molecules was hampered by NMR exchange broadening (14,45). However we believe that the protein structure itself can be employed for structural comparison, while interaction data later reported for human ileal ILBP (8,9,46) were used when comparing protein-ligand interactions. We looked first at the non conserved aminoacids, in order to possibly select for those residues with lost or gained interactions. Among these, we discarded those residues that do not point into the internal cavity. The remaining residues were mapped onto the sequence alignment to extend the analysis to the other homologouos proteins with the purpose of identifying whether a correlation between the selected aminoacids and the decrease/increase of cooperativity exist. As it was not straightforward to single out aminoacids eventually involved in cooperativity, we went on analysing pairs of aminoacids. The 11 non conserved residues within the protein cavity may give rise to 55 possible pairs. Of these only eight pairs were made of residues exhibiting a reciprocal average distance shorter than 4 Å and, more specifically, six pairs, namely 53-54, 53-34, 54-34, 24-27, 27-30, 30-34 were found both in human and chicken proteins (Table 4). Although only the pairs 54-34 and 24-27 appear to change together with cooperativity effects, the kind of interaction formed in human and chicken proteins is conserved. Indeed, in both cases, the interactions are hydrophobic and do not involve their functional groups. In particular for the 54-34 pair P54, in chicken, interacts with Hα of C54 (not with SH). Similarly Hα of S54, in human species, interacts with Hβ atom of F34. For the 24-27 pair, chicken P24 interacts with the aliphatic chain of K27 and human S24 Hβ (not OH) interacts with a methyl group of V27. Furthermore, two pairs 54-27 and 99-101 were present only in the human protein, with only the 99-101 pair correlating with cooperativity effects (Table 4). Residues 99 and 101 are buried within the protein cavity, with H99 being the only residue making contacts with both ligands in chicken ileal protein. The pair 99-101 wit exception of ileal protein, is always made by polar residues, Q/T and H/S, in the intermediate cooperative zI-BABP and rI-BABP and Q/S in the human protein. As a likely way of energetic communication between the two binding sites shoud involve a hydrogen-bonding network, we reasoned that that disruption of such network may affect cooperativity. We thus looked for conserved H-bonds involving side chains in human and chiken ileal BABPs. The H-bond between Q99 and S101 (in

human) could not be formed in the ileal protein, where residue 101 is an alanine. Calorimetry data coupled to mutation data reported for the human protein in complex with GCDA indicated that the mutant Q99A exhibited a strong decrease in cooperativity for the binding to GCDA (8). It should be noted that the 99-101 H-bond is conserved also in the zebrafish species, where the pair T/Q is present. Altogether the ensamble of these data sugegsted that a key mutation could be located at the level of the buried pair 99/101.

Mutant analysis. Two mutants were then designed and enginereed, the single mutant cI-BABP A101S (to mimic the intermediate co-operativity as in rI-BABP) and the double mutant cI-BABP H99Q/A101S to mimick the optimal couple Q/S found in the human protein. The mutated proteins were then titrated with increasing amount of ^{15}N labelled GCDA, in the assumption that gaining of the cooperative function should lead to the observation of GCDA-II amide resonance at P:L ratio lower than 2. This at variance with the behaviour observed for the wild type protein where, as a result of a consecutive binding, resonance of ligand 2 only appeared in the HSQC spectrum at protein:ligand ratio higher that 2. Figure 12 reports a comparison of the titration experiments performed for the wild-type and the two mutants A101S and H99Q/A101S. A similar behaviour is substantially displayed by the wild-type and the A101S mutant, while a clear gain in cooperativity is obtained for the double mutant where the resonances corresponding to the two bound species are already present, at their final chemical shifts (i.e the chemical shifts of the saturated species), at a P:L ratio 1:1. This behaviour is a clear indication that the binding mechanism has changed and the two binding events are no more consecutive and independent.

Bile acid production is stimulated pri- marily by bile acids returning to the liver and is also influenced by the size of the bile acid pool and number of enterohepatic circulation cycles. It should be mentioned that in humans primary bile acids include both tri-hydroxy and di-hydroxy conjugates in approximatively 55 to 35% ratio, respectively (6), while, in avians, glycine and taurine CDA conjugates (47) represent 70–95% of the bile salt pool (48). In vitro studies have clearly established that conjugates of cholic acid, containing three hydroxyl groups on the steroid nucleus, are intrinsically less cytotoxic than conjugates of CDA, a dihydroxyl bile acid. It has been suggested that the addition of a third hydroxyl group has evolved, through six types of nuclear hydroxylation pathways, as a means of detoxifying CDA. The toxicity of high concentration of CDA, on one side, further suggests the important role of I-BABPs in the regulation of bile salts concentration in the cell, and, on the other side, may explain the adaptative response of evolving a single protein system capable of binding at the same time (site

seletivity) and with high efficiency (cooparativity) different components of the bile acid pool. To this aim an efficient network of hydrogen bonds must be established and the 99-101 pair is indeed part of an extended network which allows the establishment of communicationacross the protein (Figure 13). It has also been reported that the hydrophobic-hydrophilic balance of the pool that results from the characteristics and proportions of the individual bile acids present within the pool is critical as it may dictate most of the effects of bile acids on hepatic cholesterol metabolism where more hydrophobic bile acids seem to increase biliary lipid secretion and inhibit cholesterol and bile acid synthesis, whereas hydrophilic bile acids are less effective (4).

It is thus clear that the knowledge of the subtle structural and dynamic effects regulating the binding mechanism of bile salt derivatives may give an important contribution to the understanding, in molecular terms, of the link between metabolic diseases.

.

FOOTNOTES

*The authors are indebted to Simona Tomaselli for the help with HADDOCK calculations. The European Community (EU-NMR, Contract # RII3-026145) is acknowledged for access to the HWB- NMR facility at Birmingham, UK. CIRMMP (Consorzio Interuniversitario di Risonanze Magnetiche di Metalloproteine Paramagnetiche) is acknowleged for access to high field spectrometers. The eNMR project (European FP7 e-Infrastructure grant, contract no. 213010, www.enmr.eu), supported by the national GRID Initiatives of Italy, Germany and the Dutch BiG Grid project (Netherlands Organization for Scientific Research), is acknowledged for the use of web portals, computing and storage facilities. The University of Verona and Cariverona Foundation are acknowledged for financial support for the acquisition of NMR spectrometer and cryoprobe.

REFERENCES

1. Veerkamp, J. H., and Maatman, R. G. (1995) *Prog Lipid Res* **34**, 17-52
2. Houten, S. M. (2006) *Cell metabolism* **4**, 423-424
3. Houten, S. M., Watanabe, M., and Auwerx, J. (2006) *Embo J* **25**, 1419-1425
4. Argmann, C. A., Houten, S. M., Champy, M. F., and Auwerx, J. (2006) *Curr Protoc Mol Biol* **Chapter 29**, Unit 29B 22

5. Hamilton, J. A. (2004) *Progress in lipid research* **43**, 177-199
6. Tochtrop, G. P., Bruns, J. L., Tang, C., Covey, D. F., and Cistola, D. P. (2003) *Biochemistry* **42**, 11561-11567
7. Tochtrop, G. P., DeKoster, G. T., Covey, D. F., and Cistola, D. P. (2004) *J Am Chem Soc* **126**, 11024-11029
8. Toke, O., Monsey, J. D., DeKoster, G. T., Tochtrop, G. P., Tang, C., and Cistola, D. P. (2006) *Biochemistry* **45**, 727-737
9. Toke, O., Monsey, J. D., and Cistola, D. P. (2007) *Biochemistry* **46**, 5427-5436
10. Tochtrop, G. P., Richter, K., Tang, C., Toner, J. J., Covey, D. F., and Cistola, D. P. (2002) *Proc Natl Acad Sci U S A* **99**, 1847-1852
11. Capaldi, S., Saccomani, G., Fessas, D., Signorelli, M., Perduca, M., and Monaco, H. L. (2009) *Journal of molecular biology* **385**, 99-116
12. Kouvatsos, N., Thurston, V., Ball, K., Oldham, N. J., Thomas, N. R., and Searle, M. S. (2007) *Journal of molecular biology* **371**, 1365-1377
13. Pedo, M., D'Onofrio, M., Ferranti, P., Molinari, H., and Assfalg, M. (2009) *Proteins* **77**, 718-731
14. Kramer, W., Sauber, K., Baringhaus, K. H., Kurz, M., Stengelin, S., Lange, G., Corsiero, D., Girbig, F., Konig, W., and Weyland, C. (2001) *The Journal of biological chemistry* **276**, 7291-7301
15. Guariento, M., Raimondo, D., Assfalg, M., Zanzoni, S., Pesente, P., Ragona, L., Tramontano, A., and Molinari, H. (2008) *Proteins* **70**, 462-472
16. Guariento, M., Assfalg, M., Zanzoni, S., Fessas, D., Longhi, R., and Molinari, H. (2010) *The Biochemical journal* **425**, 413-424
17. Cui, Q., and Karplus, M. (2008) *Protein science : a publication of the Protein Society* **17**, 1295-1307
18. van Dijk, A. D., de Vries, S. J., Dominguez, C., Chen, H., Zhou, H. X., and Bonvin, A. M. (2005) *Proteins* **60**, 232-238
19. Guariento, M., Assfalg, M., Zanzoni, S., Fessas, D., Longhi, R., and Molinari, H. (2010) *Biochem J* **425**, 413-424
20. Johnson, B. A. (2004) *Methods Mol Biol* **278**, 313-352
21. Shen, Y., Delaglio, F., Cornilescu, G., and Bax, A. (2009) *Journal of biomolecular NMR* **44**, 213-223
22. Guntert, P. (2004) *Methods Mol Biol* **278**, 353-378
23. Herrmann, T., Guntert, P., and Wuthrich, K. (2002) *Journal of molecular biology* **319**, 209-227
24. Dominguez, C., Boelens, R., and Bonvin, A. M. (2003) *Journal of the American Chemical Society* **125**, 1731-1737
25. Brunger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J. S., Kuszewski, J., Nilges, M., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T., and Warren, G. L. (1998) *Acta Crystallogr D Biol Crystallogr* **54**, 905-921
26. Eufri, D., and Sironi, A. (1989) *J Mol Graph* **7**, 165-169, 158

27. Schuttelkopf, A. W., and van Aalten, D. M. (2004) *Acta Crystallogr D Biol Crystallogr* **60**, 1355-1363
28. Linge, J. P., Williams, M. A., Spronk, C. A., Bonvin, A. M., and Nilges, M. (2003) *Proteins* **50**, 496-506
29. Daura, X., Antes, I., van Gunsteren, W. F., Thiel, W., and Mark, A. E. (1999) *Proteins* **36**, 542-555
30. Wallace, A. C., Laskowski, R. A., and Thornton, J. M. (1995) *Protein Eng* **8**, 127-134
31. Ijare, O. B., Somashekar, B. S., Jadegoud, Y., and Nagana Gowda, G. A. (2005) *Lipids* **40**, 1031-1041
32. Eliseo, T., Ragona, L., Catalano, M., Assfalg, M., Paci, M., Zetta, L., Molinari, H., and Cicero, D. O. (2007) *Biochemistry* **46**, 12557-12567
33. Cogliati, C., Ragona, L., D'Onofrio, M., Gunther, U., Whittaker, S., Ludwig, C., Tomaselli, S., Assfalg, M., and Molinari, H. (2010) *Chemistry* **16**, 11300-11310
34. Capaldi, S., Guariento, M., Saccomani, G., Fessas, D., Perduca, M., and Monaco, H. L. (2007) *The Journal of biological chemistry* **282**, 31008-31018
35. Ragona, L., Catalano, M., Luppi, M., Cicero, D., Eliseo, T., Foote, J., Fogolari, F., Zetta, L., and Molinari, H. (2006) *The Journal of biological chemistry* **281**, 9697-9709
36. Pelton, J. G., Torchia, D. A., Meadow, N. D., and Roseman, S. (1993) *Protein science : a publication of the Protein Society* **2**, 543-558
37. Hass, M. A., Hansen, D. F., Christensen, H. E., Led, J. J., and Kay, L. E. (2008) *Journal of the American Chemical Society* **130**, 8460-8470
38. Wolff, N., Deniau, C., Letoffe, S., Simenel, C., Kumar, V., Stojiljkovic, I., Wandersman, C., Delepierre, M., and Lecroisey, A. (2002) *Protein science : a publication of the Protein Society* **11**, 757-765
39. Tomaselli, S., Ragona, L., Zetta, L., Assfalg, M., Ferranti, P., Longhi, R., Bonvin, A. M., and Molinari, H. (2007) *Proteins* **69**, 177-191
40. van Dijk, A. D., and Bonvin, A. M. (2006) *Bioinformatics* **22**, 2340-2347
41. Arnesano, F., Banci, L., Bertini, I., and Bonvin, A. M. (2004) *Structure* **12**, 669-676
42. Rea, A. M., Thurston, V., and Searle, M. S. (2009) *Biochemistry* **48**, 7556-7564
43. Cogliati, C., Tomaselli, S., Assfalg, M., Pedo, M., Ferranti, P., Zetta, L., Molinari, H., and Ragona, L. (2009) *The FEBS journal* **276**, 6011-6023
44. Schaap, F. G., van der Vusse, G. J., and Glatz, J. F. (2002) *Molecular and cellular biochemistry* **239**, 69-77
45. Kurz, M., Brachvogel, V., Matter, H., Stengelin, S., Thuring, H., and Kramer, W. (2003) *Proteins* **50**, 312-328
46. Lu, J., Cistola, D. P., and Li, E. (2006) *Biochemistry* **45**, 1629-1639
47. Elkin, R. G., Wood, K. V., and Hagey, L. R. (1990) *Comp Biochem Physiol B* **96**, 157-161

48. Hagey, L. R., Schteingart, C. D., Ton-Nu, H. T., and Hofmann, A. F. (1994) *Journal of lipid research* **35**, 2041-2048

FIGURE LEGENDS

Fig. 1. Superposition of the twenty lowest energy structures obtained for cI-BABP employing CYANA 2.1.

Fig. 2. 1D ^{1}H NMR spectrum recorded at 600.13 MHz of a 1mM sample of GCDA in 30 mM phosphate buffer at 298K (a); Structure of GCDA with the numbering of atoms.

Fig. 3. High field region of the 2D F_1/F_2-[^{15}N,^{13}C]-filtered NOESY spectrum of 0.4 mM sample of cI-BABP in complex with GCDA at a P:L ratio 1:4. The sample was dissolved in 30 mM phosphate buffer, pH 6.5, 298K. The employed mixing time was 70 ms. The assignments are reported on the spectrum (a); High-field region of the 2D F_1/F_2-[^{15}N,^{13}C]-filtered TOCSY recorded with an isotropic mixing of 70 ms. Assignments are reported on the spectrum (b).

Fig. 4. High field region of the 2D F_1/F_2-[^{15}N,^{13}C]-filtered NOESY spectrum of 0.4 mM sample of cI-BABP in complex with GCDA at a P:L ratio 1:2. The experimental conditions are the same reported in the caption of Fig.2. Assignments are reported on the spectrum.

Fig. 5. High field region observed in the transformed first fid of the F1-edited, F3-filtered 3D HMQC-NOESY. Only resonances of protons bound to ^{12}C and exhibiting NOEs with protein are present. Experiments were performed at increasing mixing times: 0 ms (black), 80 ms (red), 100 ms (green), 120 ms (magenta), 150 ms (yellow) and 200 ms (orange).

Fig. 6. F1-F3 slices of the 3D F_1-[^{13}C]-filtered, F_2-[^{13}C]-separated, F_3-[^{13}C]-edited NOESY-HSQC spectrum of cI-BABP/GCDA complex at a 1:4 molar ratio. The protein sample was 0.8 mM. The left upper panel corresponds to the chemical shift of I28-Cβ. The right upper panel corresponds to the chemical shift of G75-Cα. The left lower panel corresponds to the chemical shift of M73-Cγ The right lower panel corresponds to the chemical shift of M71-Cε. All NOEs with the ligands atoms are indicated.

Fig. 7. Plots of the chemical-shift differences relative to unbound protein and protein bound to GCDA; Residues exhibiting a chemical shift

perturbation above two standard deviations are T53, R57, T58, T59, Y97 and H98.

Fig. 8. Superposition of the twenty lowest energy structures belonging to the best Haddock- score cluster. The internal ligands are represented as sticks in red.

Fig. 9. Ligplot two-dimensional representation of interactions observed among GCDA-I, GCDA-II and cI-BABP. Hydrogen bonds are depicted with dashed green lines and hydrophobic interactions are shown as red arcs.

Fig. 10. The phylogenetic tree obtained from the multiple sequence alignment of the entire FABP family.

Fig. 11. Sequence alignment of chicken I-BABP with human, rabbit and zebrafish I-BABPs. Conserved residues are highlighted in bold. Residues presenting non-conservative mutations in the four proteins pointing towards the ligand-binding cavity, as identified by the program CASTP (http://sts.bioengr.uic.edu/castp/) are indicated with a circle.

Fig. 12. ^{1}H-^{15}N-HSQC spectra registered during the titration of cI-BABP with ^{15}N-GCDA at different molar ratios, indicated in each panel (A); titration of A101S mutant (B); titration of A101SH99Q mutant (c).

Fig. 13. Ribbon drawing of human ileal bile acid binding protein (pdb 1O1V) showing the network of hydrogen bonds crossing the entire protein. The name of the involved residues is reported.

Fig. S1. ^{1}H-^{15}N-HSQC spectrum recorded on 0.4 mM sample of unlabelled cI-BABP in complex with $^{15}N^{13}C$-GCDA at a P:L ratio 1:4 (a); HNCA spectrum recorded on the same sample (b). The GCDA molecule bound to the site 1 (GCDA-I) and the free GCDA exhibit a unique Cα chemical shift.

Fig. S2. Superimposition of 1H-15N-HSQC spectra of a sample containing unlabelled chicken I-BABP and $^{15}N^{13}C$-GCDA in the following protein:ligand ratios: 1:0.3, 1:1, 1:1.5, 1:2.2, 1:5. Arrow follows the direction of peak shifts corresponding to ligand bound to site 1, which moves towards the chemical shift of free GCDA.

Fig. S3. Superimposition of high field regions of the 2D F_1/F_2-[^{15}N,^{13}C]-filtered NOESY spectra of a sample of doubly labelled cI-BABP and unlabelled GCDA in the following protein:ligand ratios: 1:2 (black), 1:4

(blue), 1:5 (red). Arrows follow the direction of peak shifts corresponding to the coincident free and bound-to-site-1 (GCDA-I) methyl groups moving downfield towards the free ligand chemical shift.

Fig. S4. The spectra highlight the observed correlations of histidine atoms. the adduct involving aromatic residues. $^{1}H^{13}C$ HSQC-CT experiment recorded on 0.4 mM sample of cI-BABP in complex with GCDA at a P:L ratio 1:4. The assignments are reported on the spectrum (a); ^{1}H-^{15}N HMBC experiment recorded on the same sample. The assignments are reported on the spectrum (b); 2D from $^{1}H^{1}H$ NOESY spectrum showing the only observable correlations of H99. The assignments are reported on the spectrum (c).

Figure 1

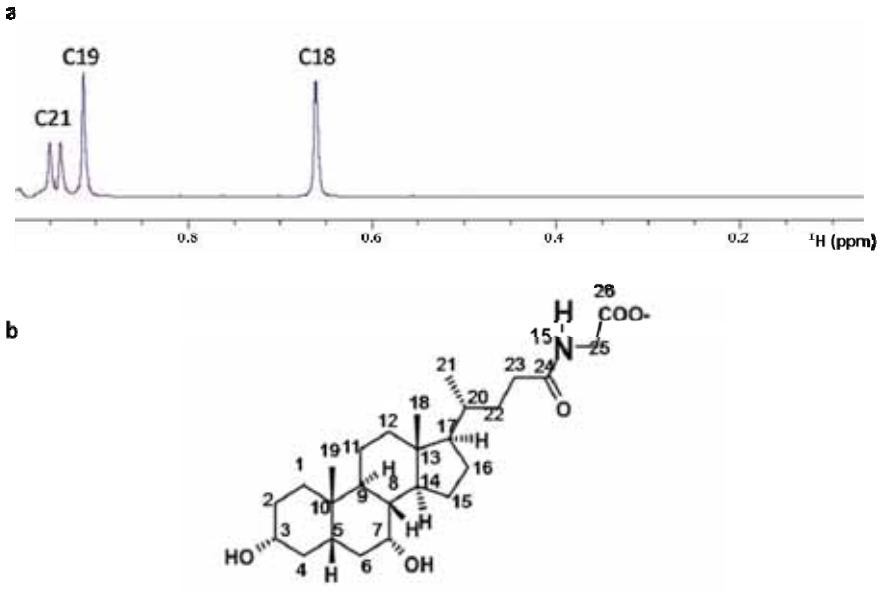


Figure 2

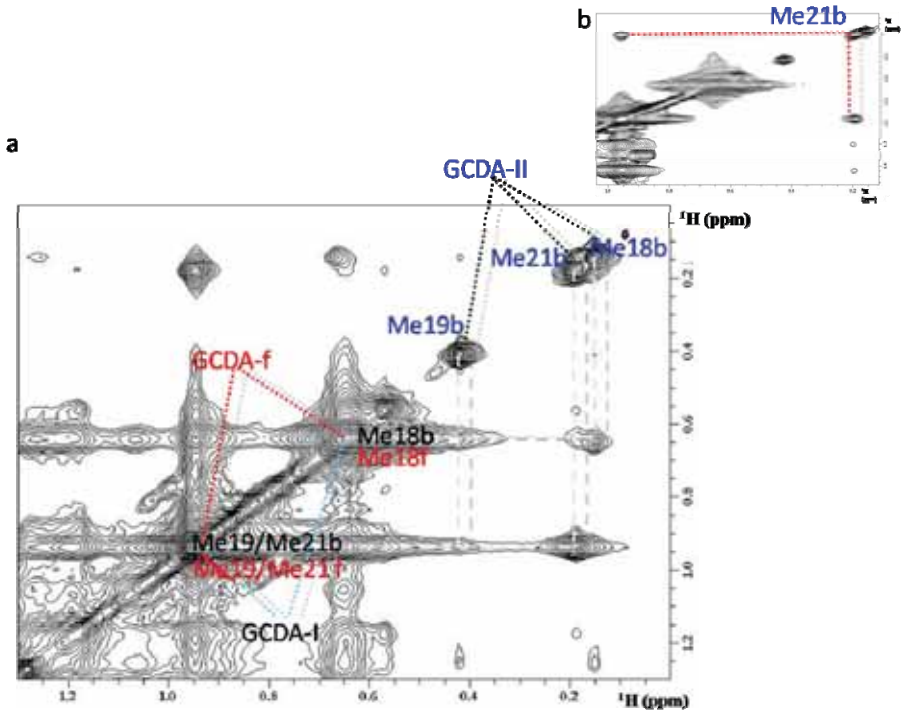


Figure 3

Figure 3

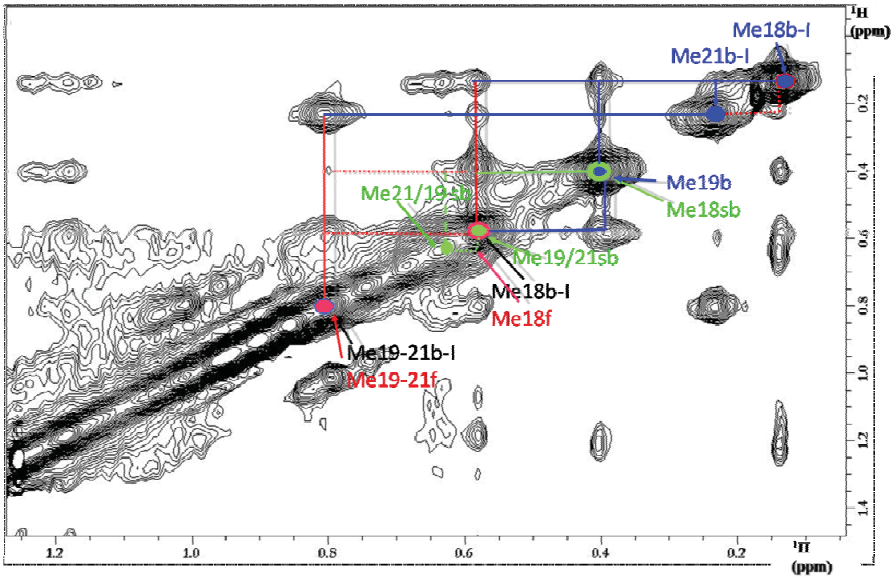

Me18b-l
Me21b-l
Me19b
Me18sb
Me21/19sb
Me19/21sb
Me18b-l
Me18f
Me19-21b-l
Me19-21f
^{1}H (ppm)
^{1}H (ppm)

Figure 4

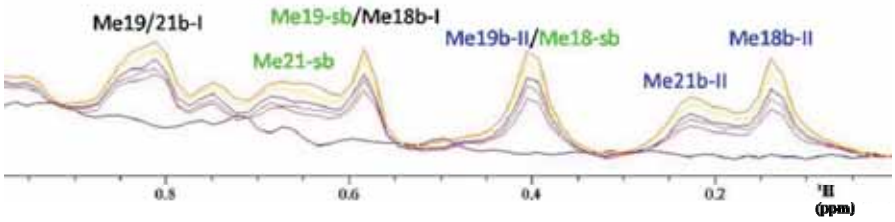

Me19/21b-I
Me19-sb/Me18b-I
Me21-sb
Me19b-II/Me18-sb
Me21b-II
Me18b-II
0.8
0.6
0.4
0.2
¹H (ppm)

Figure 5

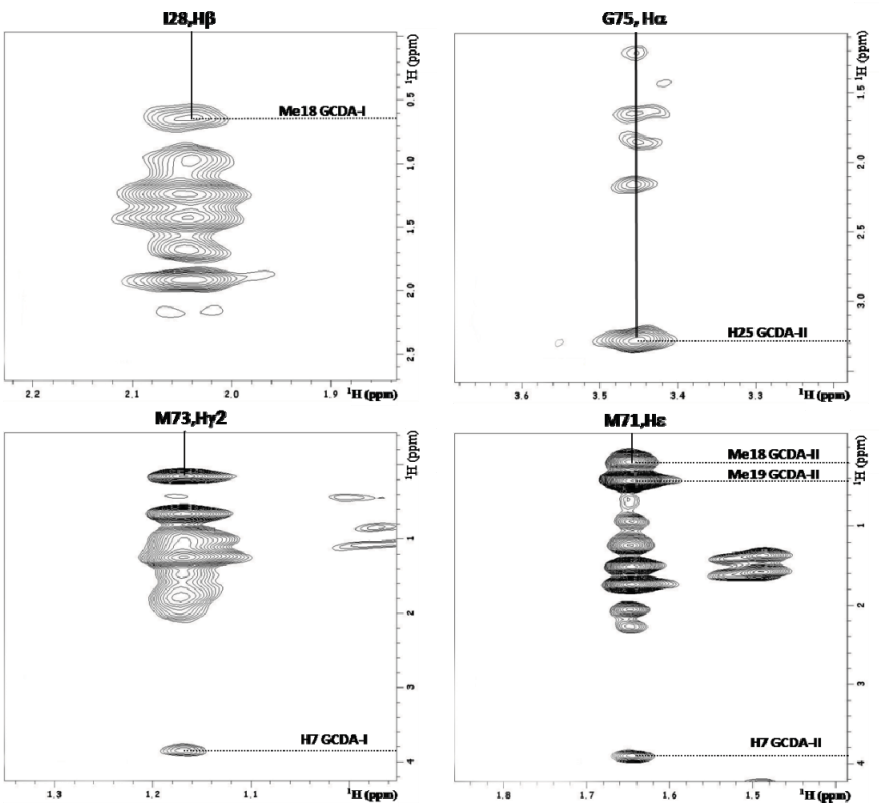

I28,Hβ
Me18 GCDA-I
G75, Hα
H25 GCDA-II
M73,Hγ2
H7 GCDA-I
M71,Hε
Me18 GCDA-II
Me19 GCDA-II
H7 GCDA-II
^{1}H (ppm)

Figure 6

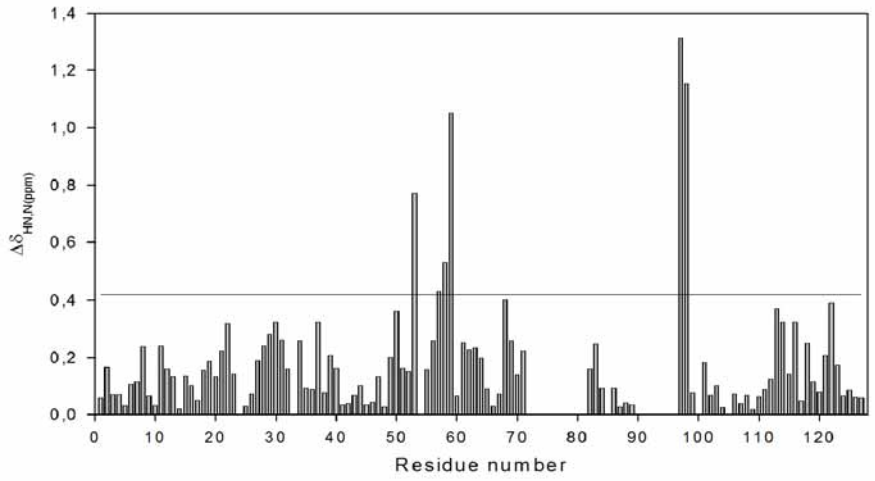

1,4
1,2
1,0
0,8
0,6
0,4
0,2
0,0
$\Delta\delta_{HN/N(ppm)}$
0
10
20
30
40
50
60
70
80
90
100
110
120
Residue number

Figure 7

Figure 8

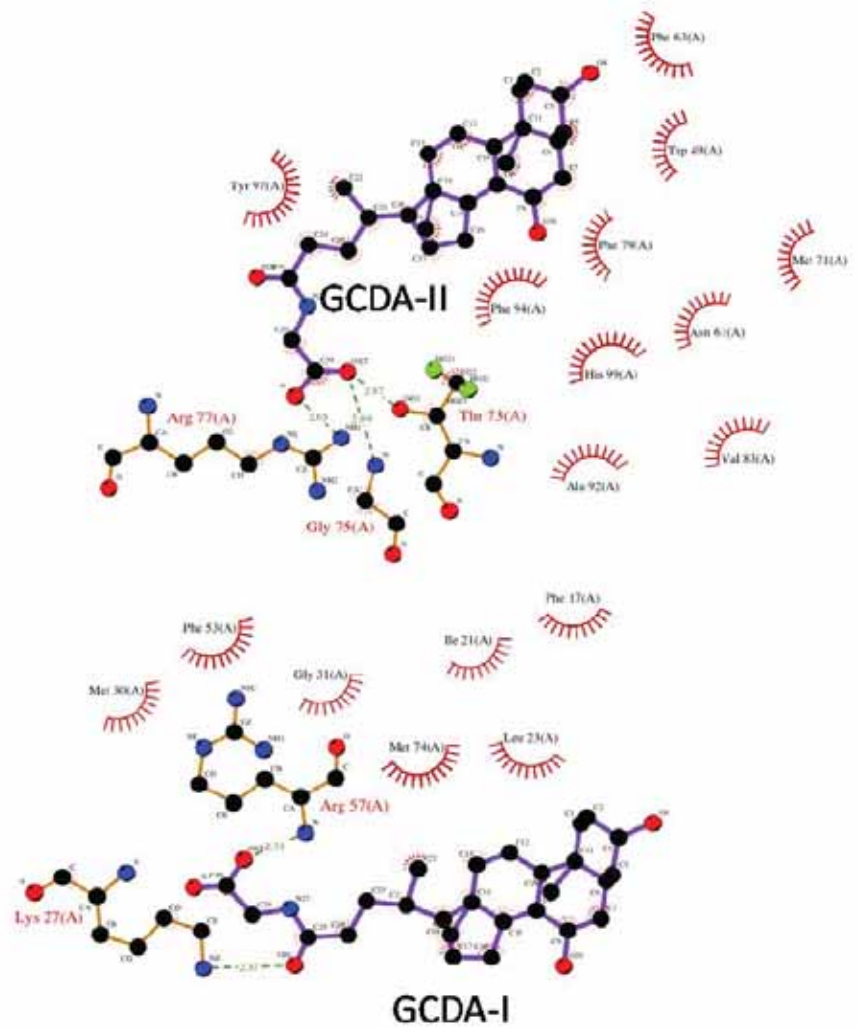

GCDA-II
Phe 63(A)
Trp 49(A)
Tyr 97(A)
Phe 79(A)
Met 71(A)
Phe 94(A)
Asn 67(A)
His 99(A)
Arg 77(A)
Thr 73(A)
Val 83(A)
Ala 92(A)
Gly 75(A)
Phe 17(A)
Phe 53(A)
Ile 21(A)
Met 30(A)
Gly 31(A)
Met 74(A)
Leu 23(A)
Arg 57(A)
Lys 27(A)
GCDA-I

Figure 9

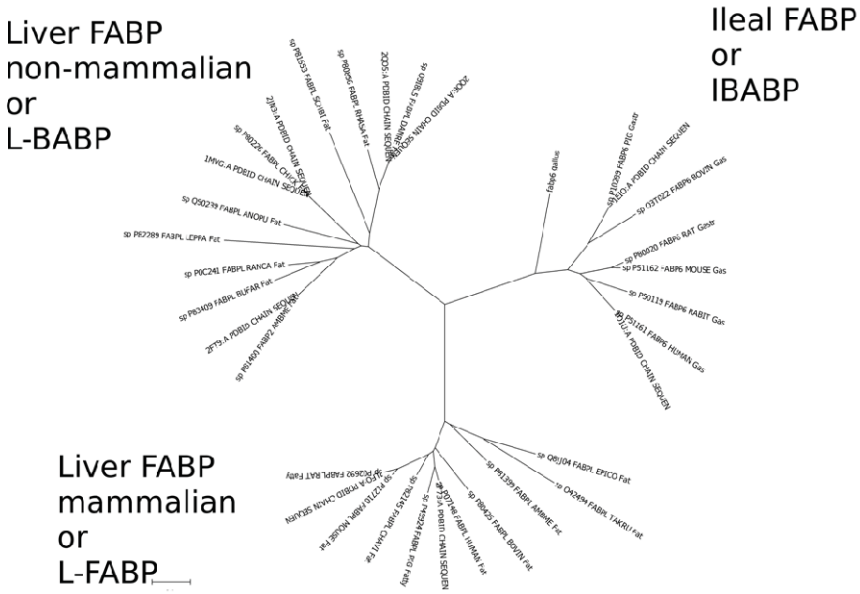

Liver FABP
non-mammalian
or
L-BABP
Ileal FABP
or
IBABP
Liver FABP
mammalian
or
L-FABP

Figure 10

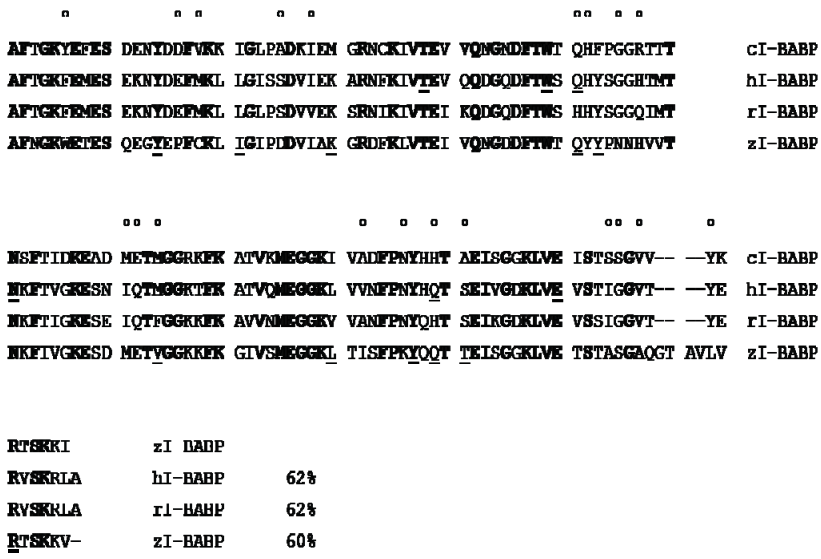

AFTGKYEFES DENYDDFVKK IGLPADKIEM GRNCKIVTEV VQNGNDFTWT QHFPGGRTIT cI-BABP
AFTGKFEMES EKNYDEFMKL LGISSDVIEK ARNFKIVTEV QQDGQDFTWS QHYSGGHTMT hI-BABP
AFTGKFEMES EKNYDEFMKL LGLPSDVVEK SRNIKIVTEI KQDGQDFTWS HHYSGGQIMT rI-BABP
AFNGKWETES QEGYEPFCKL IGIPDDVIAK GRDFKLVTEI VQNGDDFTWT QYYPNNHVVT zI-BABP
MSFTIDKEAD METMQGRKFK ATVKMEGGKI VADFPNYHHT AEISGGKLVE ISTSSGVV-- --YK cI-BABP
MKFTVGKESN IQTMQGKTFK ATVQMEGGKL VVNFPNYHQT SEIVGDKLVE VSTIGGVT-- --YE hI-BABP
MKFTIGKESE IQTFGGKKFK AVVNMEGGKV VANFPNYQHT SEIKGDKLVE VSSIGGVT-- --YE rI-BABP
MKFIVGKESD METVGGKKFK GIVSMEGGKL TISFPKYQQT TEISGGKLVE TSTASGAQGT AVLV zI-BABP
RTSKKI zI BADP
RVSKRLA hI-BABP 62%
RVSKRLA rI-BABP 62%
RTSKKV- zI-BABP 60%

Figure 11

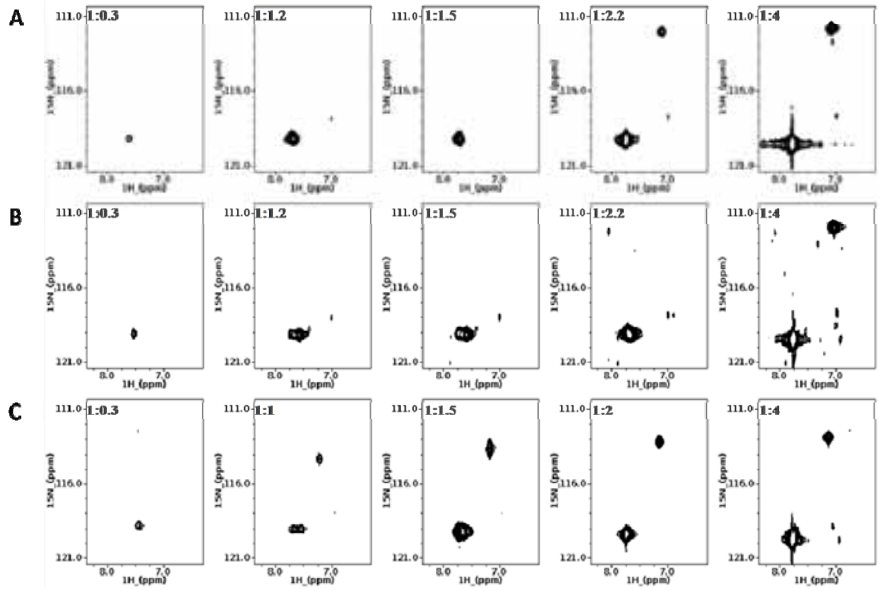


Figure 12


LYS-35
GLU-9
LYS-52
GLN-51
LYS-10
GLU-120
GLU-110
GLN-99
TRP-49
SER-100
GLU-102
LYS-89

Figure 13

Table 1. Statistics for the Final Structural Ensemble Obtained

average rmsd (Å)	
⟨rmsd⟩ backbone (Cα, NH, and C')[a]	0.49 + 0.08 Å
⟨rmsd⟩ heavy atoms	1.07 ± 0.08 Å
experimental restraints	
short-range NOEs ($\|i-j\| \leq 1$)	1342
medium-range NOEs ($1 < \|i-j\| < 5$)	361
long-range NOEs ($\|i-j\| > 5$)	1035
total	2738
dihedral angle restraints	221
experimental restraints violations[b]	
Distance violations / structure	
0.1 - 0.2 Å	5.2
0.2 - 0.5 Å	0.6
> 0.5 Å	0
RMS of distance violation / constraint	0.01 Å
Maximum distance violation[c]	0.31 Å
Dihedral angle violations / structure	
1 - 10 °	2.65
> 10 °	0
RMS of dihedral angle violation / constraint	0.20 °
Maximum dihedral angle violation[c]	3.40 °
Ramachandran analysis	
Residues in the favoured region (%)	86.1
Residues in allowed regions (%)	13.9
Residues in generous regions (%)	0
Residues in disallowed regions (%)	0
Close Contacts and Deviations from Ideal Geometry (from PDB validation software)	
Number of close contacts[d]	0
RMS deviation for bond angles	0.2 °
RMS deviation for bond lengths	0.001 Å

Global quality score

Program	Verify3D	ProsaII(-ve)	Procheck(phi-psi)	Procheck(all)	MolProbity Clashscore
Raw score	0.47	0.52	-0.62	-0.72	16.68
Z-score[e]	0.16	-0.54	-2.12	-4.20	-1.34

[a]Root mean square deviations of atomic coordinates were calculated over residues 1-127 using MOLMOL. [b]Calculated for all constraints for the given residues, using sum over r^-6. [c]Largest constraint violation among all the reported structures. [d]Within 1.6 Å for H atoms, 2.2 Å for heavy atoms. [e] Z-Score computed by PSVS.

Table 2. Ambiguous and non-ambigous intermolecolar restraints used to calculate the structure of the complex		
Ambiguous interaction restraints	*Non-ambiguous interaction restraints*	
	Protein atom	*Ligand atom*
F17 (NOE)[a]	F17, Cε	C3 GCDA-I
V18 (NOE)	F17, Cε	C7 GCDA-I
I21 (NOE)	V18, Cγ	Me18 GCDA-I
L23 (NOE)	V18, Cγ	Me19 GCDA-I
I28 (NOE)	I21, Cγ and Cδ	Me19 GCDA-I
M30 (NOE)	L23, Cδ	Me18 GCDA-I
I36 (NOE)	L23, Cδ	Me19 GCDA-I
F53 (CSP,NOE)	L23, Cδ	C25 GCDA-II
R57 (CSP,NOE)	I28, Cα and Cβ	Me18 GCDA-I
T58 (CSP)	M30, Cβ, Cγ and Cε	Me21 GCDA-I
T59 (CSP)	M30, Cε	C25 GCDA-I
N61 (NOE)	I36, Cδ	Me21 GCDA-I
A69 (NOE)	F53, Cβ, Cδ, Cε and Cζ	Me21 GCDA-I
M71(NOE)	R57, Cδ	C25 GCDA-I
T73 (NOE)	N61, Cβ and Nδ2	C7 GCDA-II
M74 (NOE)	N61, Cβ	Me19 GCDA-II
G75 (NOE)	A69, Cβ	Me19 GCDA-II
F79 (NOE)	M71, Cβ	Me18 GCDA-II
A81 (NOE)	M71, Cε	Me19 GCDA-II
V83 (NOE)	M71, Cε	C7 GCDA-II
A92 (NOE)	T73, Cγ	C7 GCDA-I
F94 (NOE)	T73, Cγ	C25 GCDA-II
Y97 (CSP)	M74, Cε	Me21 GCDA-I
H98 (CSP)	G75, Cα	C7 GCDA-I
Y119 (NOE)	G75, Cα	C25 GCDA-II
	F79, Cε and Cζ	Me18 GCDA-II
	F79, Cε and Cζ	Me19 GCDA-II
	F79, Cε and Cζ	C7 GCDA-II
	A81, Cβ	Me18 GCDA-II
	A81, Cβ	Me19 GCDA-II
	V83, Cγ	Me19 GCDA-II
	A92, Cβ	Me18 GCDA-II
	A92, Cβ	Me19 GCDA-II
	F94, Cε and Cζ	Me18 GCDA-II
	F94, Cε and Cζ	Me19 GCDA-II
	Y119, Cε	C3 GCDA-I

[a]The experimental method used to define the residue as active is indicated in parenthesis.

Table 3. Structural Statistics of twenty lowest energy structures in the lowest energy cluster	
average rmsd (Å)	
〈rmsd〉 backbone (Cα, NH, and C')[a]	0.44 ± 0.08 Å
〈rmsd〉 heavy atoms	0.95 ± 0.10 Å
Backbone RMSD (Å) with respect to mean	
Flexible interface backbone	0.29 ± 0.09 Å
All backbone	0.32 ± 0.13 Å
Experimental restraints	
Number of ambigous interaction restraints	49
Number of intermolecular interaction restraints	48
Number of intramolecular interaction restraints	2738
Number of dihedral angle restraints	221
Restraints violations	
Total ambigous violations	1
Total unambigous violations	0
Intramolecular distance violations / structure[b]	
0.1 - 0.2 Å	0.15
0.2 - 0.5 Å	0
> 0.5 Å	0
Dihedral angle violations / structure	
1 - 10 °	10
> 10 °	0
Intermolecular energies after water refinement	
E_{vdw} (kcal mol^{-1})	-61.66 ± 3.35
E_{elec} (kcal mol^{-1})	-282.4 ± 36.22
Buried surface area ($Å^2$)	1663.37 ±27.24
Ramachandran analysis	
Residues in the favoured region (%)	91.4
Residues in allowed regions (%)	8.6
Residues in generous regions (%)	0
Residues in disallowed regions (%)	0
Close Contacts and Deviations from Ideal Geometry (from PDB validation software)	
Number of close contacts[c]	37
RMS deviation for bond angles	0.6 °
RMS deviation for bond lengths	0.005 Å

[a]Root mean square deviations of atomic coordinates were calculated over residues 1-127 using MOLMOL. [b]Calculated for all constraints for the given residues, using sum over r^-6. [c]Within 1.6 Å for H atoms, 2.2 Å for heavy atoms.

Table 4. Pairs of non conserved residues that show recoprocal interactions				
Residue pair	cI-BABP residues	zI-BABP residues	rI-BABP residues	hI-BABP residues
53-54	F-P	Y-P	Y-S	Y-S
53-34	F-C	Y-K	Y-I	Y-F
54-34	P-C	P-F	S-I	S-F
24-27	P-K	P-V	P-V	S-V
27-30	K-M	V-K	V-K	V-K
30-34	M-C	K-F	K-I	K-F
54-27[a]	P-K	P-V	S-V	S-V
99-101[a]	H-A	Q-T	H-S	Q-S
ΔΔG	-0.87	-4.55	-6.96	-15.50

[a] Residue pair displays interactions only in human I-BABP.

SUPPLEMENTARY MATERIAL

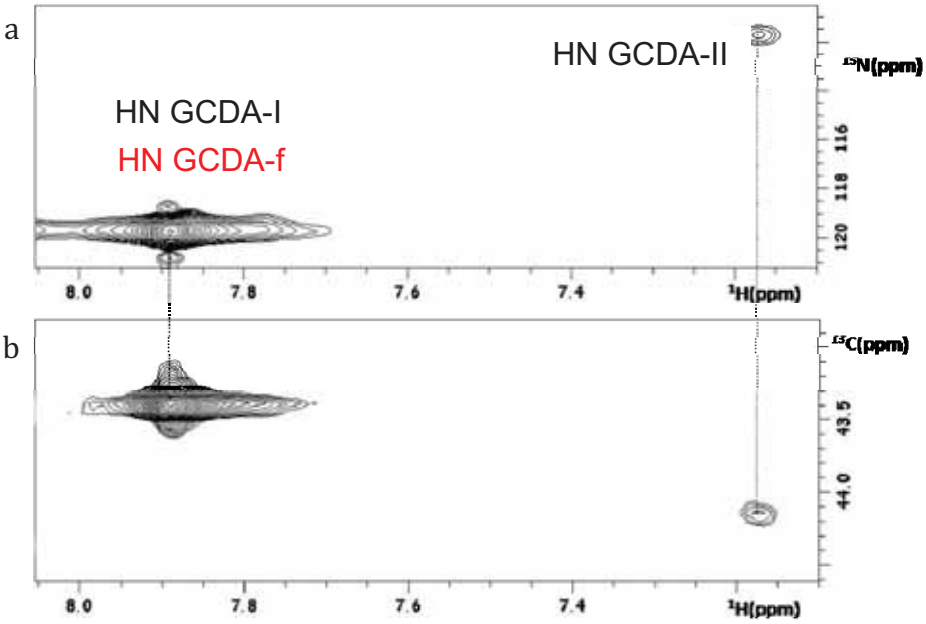


Figure S1

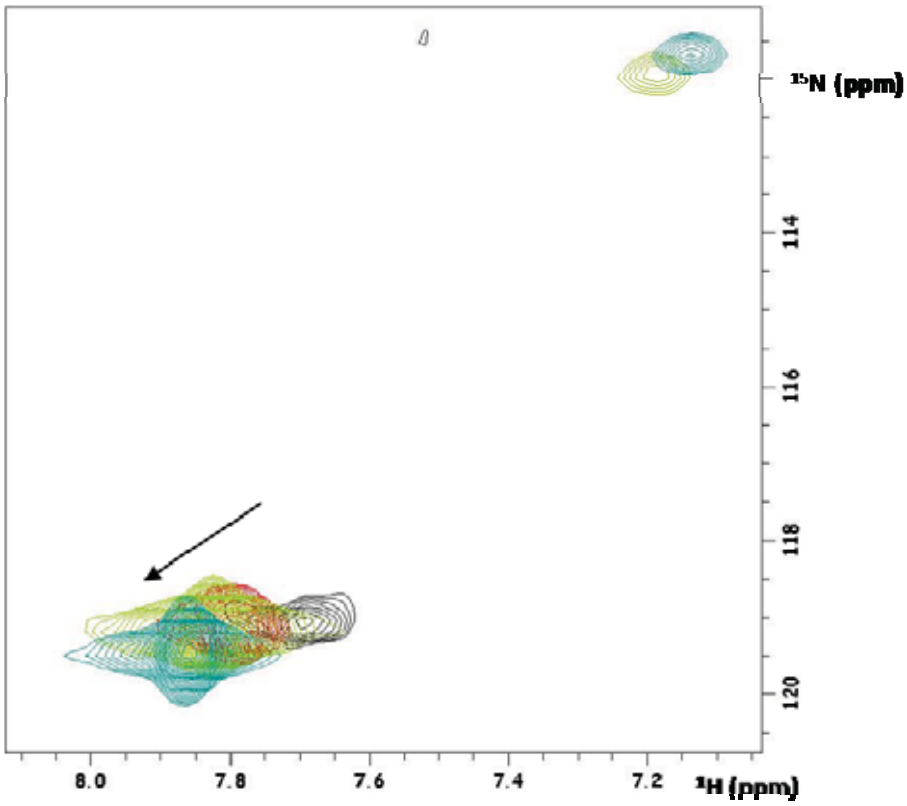


Figure S2

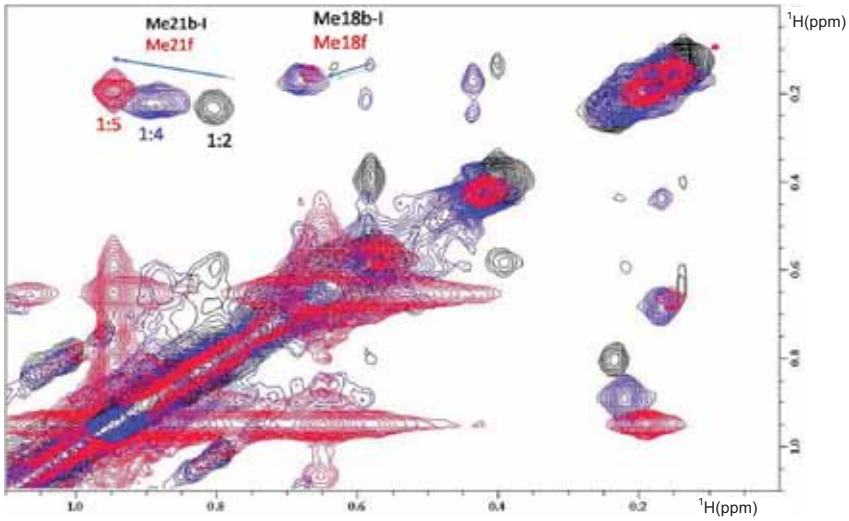


Figure S3

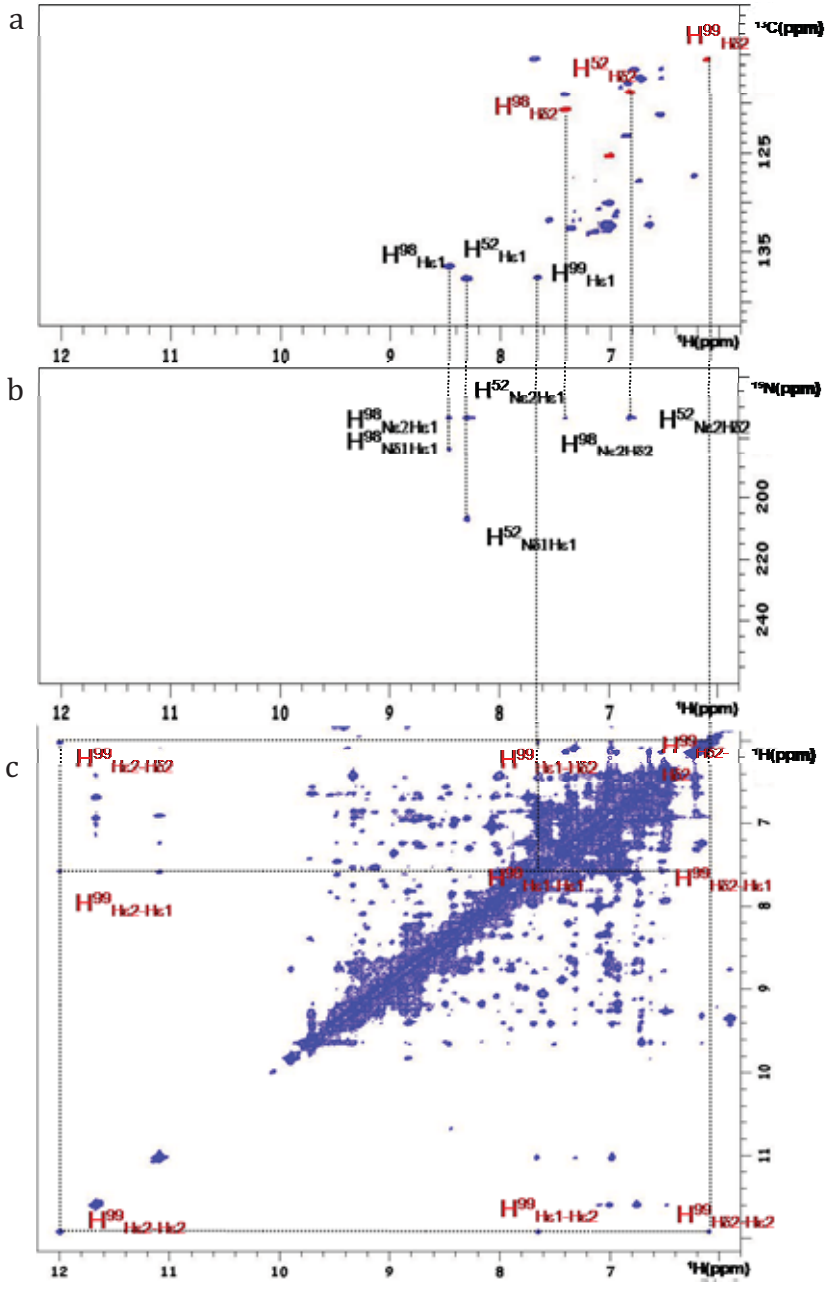

a
b
c
^{13}C(ppm)
^{15}N(ppm)
^{1}H(ppm)
$H^{99}_{Hδ2}$
$H^{52}_{Hδ2}$
$H^{98}_{Hδ2}$
$H^{98}_{Hε1}$
$H^{52}_{Hε1}$
$H^{99}_{Hε1}$
$H^{52}_{Nε2Hε1}$
$H^{98}_{Nε2Hε1}$
$H^{98}_{Nδ1Hε1}$
$H^{98}_{Nε2Hδ2}$
$H^{52}_{Nε2Hδ2}$
$H^{52}_{Nδ1Hε1}$
$H^{99}_{Hε2-Hδ2}$
$H^{99}_{Hε1-Hδ2}$
$H^{99}_{Hε2-Hε1}$
$H^{99}_{Hε1-Hε1}$
$H^{99}_{Hδ2-Hε1}$
$H^{99}_{Hε2-Hε2}$
$H^{99}_{Hε1-Hε2}$
$H^{99}_{Hδ2-Hε2}$

Figure S4

Conclusions

Bile acids play an important role in digestion and absorption of dietary fats and in the regulation of cholesterol homoeostasis. They are synthesized from cholesterol in the liver and secreted with bile into the intestine where they help in dietary fat digestion. They also play an important role as signaling molecules, capable of interacting with target nuclear receptors regulating their own transcription.
The importance of bile acid enterohepatic circulation (recycling system) is related to the maintenance of bile acid and cholesterol balance. In the bile acids enterohepatic circulation, three key steps are present, mediated by:

- a receptor system, that binds bile salts on one surface and translocates them into the cell;
- a cellular bile salt binding protein, that transports bile acids across the cell;
- an exit system, which moves bile salts out of the other side of the cell.

Much has been learned about the mechanism by which bile acids enter and exit liver and intestinal cells, but very little is known about how bile acids are shuttled in the cytosol, although the BABPs were proposed to act as the putative bile acid carriers in this compartment.

The study presented here is aimed at the NMR structural characterization of the intracellular transporters present both in liver and ileum of chicken. In this regard, I have addressed the study of the interactions between: i) liver-BABP and a potential hepatospecific contrast agent, Gd-DTPA-conjugated to 5β-cholanoic acid (Gd-I); ii) ileal-BABP and its physiological ligands, namely glycochenodeoxycholic and glycocholic acids (GCDA and GCA respectively).

The study of the mechanism involved in the intracellular transport of the bile acid-derived contrast agent in the hepatocytes is crucial for the definition of the drugs pharmacokinetics. An efficient contrast agent

should provide in fact a high contrast at low dose and have a rapid excretion after administration. The solution structure of liver-BABP in complex with the synthetic Gd-I has been obtained and fully described in paper N°1 (Tomaselli *et al*. J. Med. Chem.**,** 2008, 51, 6782–6792). The relevance of the presented results is related to the identification of the molecular determinants of the interactions in the described system. The obtained structural data represent a necessary step to guide the rational design of hepatocyte-selective gadolinium chelates for liver targeting.

In the second part of my thesis I have investigated the binding features of ileal-BABP with bile acid, employing both calorimetric and NMR approaches. The analysis described in paper N° 2 (Guariento *et al*. Biochem. J., 2010, 425, 413–424) derived from ITC measurements in combination with NMR titration experiments indicated the absence of cooperative binding, at variance with the homologous human protein. The most relevant result of the paper is related to the identification of a system, the chicken ileal protein bound to bile salts, containing a protein scaffold which is able to establish a long-range communication network in spite of lack of cooperativity. These findings set the basis for paper N° 3, ready for submission, (Zanzoni *et al* .) which addresses the molecular determinants for a "gain-of-function" on going from non-mammalian to mammalian species. A high variety of NMR experiments has been combined to derive the structure of the ternary complex. The most relevant result of the paper is the identification, on the basis of extensive structural and bioinformatics analysis, of the key interaction capable of affecting the binding mechanism, introducing a gain of cooperativity. Thus it is likely that a well-tuned energetic coupling, necessary in a cooperativity interaction, has evolved due to an adaptation process for efficient bile salt homeostasis.

Bibliography

1. Chen B., Dong J.Q., Chen Y.J., Wang J.M., Tian J., Wang C.B., Zou S.Q. (2007) Two-dimensional electrophoresis for comparative proteomic analysis of human bile. *Hepatobiliary Pancreat. Dis. Int.* 6:402-6.
2. Houten S. M., Auwerx J. (2004) The enterohepatic nuclear receptors are major regulators of the enterohepatic circulation of bile salts. *Ann. Med.* 36:482-91.
3. Hofmann A.F. (2007) Biliary secretion and excretion in health and disease: Current concepts. *Annals of Hepatology* 6:15-27.
4. Chiang J. Y. (2004) Regulation of bile acid synthesis: pathways, nuclear receptors, and mechanisms. *J. Hepatol.* 40:539-51.
5. Hylemon P.B, Harder J. (1998) Biotransformation of monoterpenes, bile acids, and other isoprenoids in anaerobic ecosystems. *FEMS Microbiol. Rev*. 22:475-88.
6. Cowen A. E., Korman M. G., Hofmann A. F., Cass O. W., Coffin S. B. (1975) Metabolism of lithocholate in healthy man. II. Enterohepatic circulation. *Gastroenterology* 69:67-76.
7. Allan R. N., Thistle J. L., Hofmann A. F. (1976) Lithocholate metabolism during chenotherapy for gallstone dissolution. II. Absorption and sulfation. *Gut.* 17:413-19.
8. Trauner M., Boyer J. L. (2003) Bile salt transporters: molecular characterization, function, and regulation. *Physiol. Rev.* 83:633-71.
9. Wiener I., Inoue K., Fagan C.J., Lilja P., Watson L.C., Thompson J.C. (1981) Release of cholecystokinin in man:Correlation of blood levels with gallbladder contraction. *Ann. Surg*. 194:321-27.
10. Bajor A., Gillberg P.G., Abrahamsson H. (2010) Bile acids: short and long term effects in the intestine. *Scand. J. Gastroenterol*. 45:645-64.
11. Wilson F. (1981) Intestinal transport of bile acids. *Am. J. Physiol*. 241:G83-92.
12. Weinman S.A., Carruth, M.W., Dawson P.A. (1998) Bile acid uptake via the human apical sodium-bile acid cotransporter is electrogenic. *J. Biol.Chem.* 273*:*34691-5.

13. Bahar R.J, Stolz A. (1999) Bile acid transport. *Gastroenterol. Clin. North. Am.* 28:27-58.
14. Lin M. C., Kramer W., Wilson F. A. (1990) Identification of cytosolic and microsomal bile acid-binding proteins in rat ileal enterocytes. *J. Biol. Chem.* 265:14986-95.
15. Tochtrop G. P., Richter K., Tang C., Toner J. J., Covey D. F., Cistola D. P. (2002) Energetics by NMR: site-specific binding in a positively cooperative system. *Proc. Natl. Acad. Sci. U.S.A*. 99:1847-52.
16. Lücke C., Zhang F., Hamilton J. A., Sacchettini J. C., Rüterjans H. (2000) Solution structure of ileal lipid binding protein in complex with glycocholate. *Eur. J. Biochem.* 267:2929-38.
17. Kurz M., Brachvogel V., Matter H., Stengelin S., Thuering H., Kramer W. (2003) Insights into the bile acid transportation system: the human ileal lipid-binding protein-cholyltaurine complex and its comparison with homologous structures. *Proteins* 50:312-28.
18. Dawson P. A., Hubbert M., Haywood J., Craddock A. L., Zerangue N., Christian W. V., Ballatori N. (2005) The heteromeric organic solute transporter alpha-beta, Ostalpha-Ostbeta, is an ileal basolateral bile acid transporter. *J. Biol. Chem.* 280:6960-8.
19. Meier P.J. (1995) Molecular mechanisms of hepatic bile salt transport from sinusoidal blood into bile. *Am. J. Physiol. Gastrointest. Liver Physiol.* 269: G801-12.
20. Kullak-Ublick G. A., Stieger B., Meier P. J. (2004) Enterohepatic bile salt transporters in normal physiology and liver disease. *Gastroenterology* 126:322-42.
21. Meier P. J., Stieger B. (2002) Bile salt transporters. *Annu. Rev. Physiol.* 64:635-61.
22. Von Dippe P., Amoui M., Alves C., Levy D. (1993) Na^+-dependent bile acid transport by hepatocytes is mediated by a protein similar to microsomal epoxide hydrolase. *Am. J. Physiol.* 264:G528-34.
23. Zhu Q.S., Xing W., Qian B., von Dippe P., Shneider B.L., Fox V.L., Levy D. (2003) Inhibition of human m-epoxide hydrolase gene expression in a case of hypercholanemia. *Biochim. Biophys. Acta.*

1638:208-16.
24. Kullak-Ublick G.A., Stieger B., Hagenbuch B., Meier P.J. (2000) Hepatic Transport of bile salts. *Semin. Liver. Dis.* 20: 273-92.
25. Kullak-Ublick G.A., Ismair M.G., Stieger B., Landmann L., Huber R., Pizzagalli F., Fattinger K., Meier P.J., Hagenbuch B. (2001) Organic anion-transporting polypeptide B (OATP-B) and its functional comparison with three other OATPs of human liver. *Gastroenterology* 120: 525-33.
26. Agellon L.B., Torchia E.C. (2000) Intracellular transport of bile acids. *Biochim. Biophys. Acta.* 1486:198-209.
27. Hofmann A.F. (1999) The Continuing Importance of Bile Acids in Liver and Intestinal Disease. *Arch. Intern. Med.* 159:2647-58.
28. Vlahcevic Z.R., Heuman D.M., Hylemon P.B. (1991) Regulation of bile acid synthesis. *Hepatology* 13:590-600.
29. Parks D.J., Blanchard S.G., Bledsoe R.K., Chandra G., Consler T.G., Kliewer S.A., Stimmel J.B., Willson T.M., Zavacki A.M., Moore D.D., Lehmann J.M. (1999) Bile acids: natural ligands for an orphan nuclear receptor. *Science* 284:1365-8.
30. Holt J. A., Luo G., Billin A. N, Bisi J., McNeill Y. Y., Kozarsky K. F., Donahee M., Wang D. Y., Mansfield T. A., Kliewer S. A., Goodwin B., Jones S.A.(2003) Definition of a novel growth factor-dependent signal cascade for the suppression of bile acid biosynthesis. *Genes. Dev*. 17:1581-91 .
31. Feingold K. R., Spady D. K., Pollock A. S., Moser A. H., Grunfeld C. (1996) Endotoxin, TNF, and IL-1 decrease cholesterol 7alpha-hydroxylase mRNA levels and activity. *J. Lipid Res.* 37:223-28 .
32. Datta S., Wang L., Moore D.D., Osborne T.F. (2006) Regulation of 3-Hydroxy-3 methylglutaryl Coenzyme A Reductase Promoter by Nuclear Receptors Liver Receptor Homologue-1 and Small Heterodimer Partner. *J. Biol. Chem*. 281:807-12.
33. Fickert P., Zollner G., Fuchsbichler A., Stumptner C., Pojer C., Zenz R., Lammert F., Stieger B., Meier P. J., Zatloukal K., Denk H., Trauner M. (2001) Effects of ursodeoxycholic and cholic acid feeding on

hepatocellular transporter expression in mouse liver. *Gastroenterology* 121:170-83.

34. Kast H.R., Goodwin B., Tarr P.T., Jones S.A., Anisfeld A.M., Stoltz C.M., Tontonoz P., Kliewer S., Willson T.M., Edwards P.A. (2002) Regulation of multidrug resistance-associated protein 2 (ABCC2) by the nuclear receptors pregnane X receptor, farnesoid X-activated receptor, and constitutive androstane receptor. *J. Biol. Chem.* 277:2908-15.

35. Denson L. A., Sturm E., Echevarria W., Zimmerman T. L., Makishima M., Mangelsdorf D. J., Karpen S. J. (2001) The orphan nuclear receptor, shp, mediates bile acid-induced inhibition of the rat bile acid transporter, ntcp. *Gastroenterology* 121:140-47.

36. Landrier J. F., Eloranta J. J., Vavricka S. R., Kullak- Ublick G. A. (2006) The nuclear receptor for bile acids, FXR, transactivates human organic solute transporter-alpha and -beta genes. *Am. J. Physiol. Gastrointest. Liver. Physiol*. 290:G476-85.

37. Kanda T., Foucand L., Nakamura Y., Niot I., Besnard P., Fujita M., Sakai Y., Hatakeyama K., Ono T., Fujii H. (1998) Regulation of expression of human intestinal bile acidbinding protein in Caco-2 cells. *Biochem. J.* 330:261-65.

38. Chen F., Ma L., Dawson P. A., Sinal C. J., Sehayek E., Gonzalez F. J., Breslow J., Ananthanarayanan M., Shneider B. L. (2003) Liver receptor homologue-1 mediates species- and cell line-specific bile acid-dependent negative feedback regulation of the apical sodium-dependent bile acid transporter. *J. Biol. Chem.* 278:19909-16.

39. Cariou B., van Harmelen K., Duran-Sandoval D., van Dijk T. H., Grefhorst A., Abdelkarim M., Caron S., Torpier G., Fruchart J. C., Gonzalez F. J., Kuipers F., Staels B. (2006) The farnesoid X receptor modulates adiposity and peripheral insulin sensitivity in mice. *J. Biol. Chem.* 281:11039-49.

40. Kalaany N. Y., Mangelsdorf D. J. (2006) LXRS and FXR: the yin and yang of cholesterol and fat metabolism. *Annu. Rev. Physiol.* 68:159-91.

41. Watanabe M., Houten S. M., Wang L., Moschetta A., Mangelsdorf D. J., Heyman R. A., Moore D. D., Auwerx J. (2004) Bile acids lower

triglyceride levels via a pathway involving FXR, SHP, and SREBP-1c. *J. Clin. Invest.* 113:1408-18.

42. Kawamata Y., Fujii R., Hosoya M., Harada M., Yoshida H., Miwa M., Fukusumi S., Habata Y., Itoh T., Shintani Y., Hinuma S., Fujisawa Y., Fujino M. A. (2003) G protein-coupled receptor responsive to bile acids. *J. Biol. Chem.* 278:9435-40.

43. Maruyama T., Miyamoto Y., Nakamura T., Tamai Y., Okada H., Sugiyama E., Nakamura T., Itadani H., Tanaka K. (2002) Identification of membrane-type receptor for bile acids (M-BAR). *Biochem. Biophys. Res. Commun.* 298:714-9.

44. Watanabe M., Houten S. M., Mataki C., Christoffolete M. A., Kim B. W., Sato H., Messaddeq N., Harney J. W., Ezaki O., Kodama T., Schoonjans K., Bianco A. C., Auwerx J. (2006) Bile acids induce energy expenditure by promoting intracellular thyroid hormone activation. *Nature* 439:484-9.

45. Zimmermann A. W., Veerkamp J. H. (2002) New insight into the structure and function of fatty acid-binding proteins. *Cell. Mol. Life Sci.* 59:1096-116.

46. Ockner R. K., Manning J. A., Poppenhausen R. B., Ho W. K. (1972) A binding protein for fatty acids in cytosol of intestinal mucosa, liver, myocardium, and other tissues. *Science* 177:56-8.

47. Banaszak L., Winter N., Xu Z. H., Bernlohr D. A., Cowan S., Jones T. A. (1994) Lipid-binding proteins: a family of fatty acid and retinoid transport proteins. *Adv. Prot. Chem*. 45:89-151.

48. Veerkamp J.H., van Kuppevelt T.H., Maatman R.G., Prinsen C.F. (1993) Structural and functional aspects of cytosolic fatty acid-binding proteins. *Prostaglandins Leukot. Essent. Fatty Acids* 49:887-906.

49. Veerkamp J.H. Maatman R.G.H.J. (1995) Cytoplasmic fatty acidbinding proteins: their structure and genes. *Prog. Lipid Res.* 34:17-52.

50. Glatz J.F.C., Vork, M.M., Cistola D.P., van der Vusse G.J. (1993) Cytoplasmatic fatty acid binding protein: significance for intracellular transport of fatty acids and putative role on signal transduction pathways. *Prostaglandins Leukotrienes Essent. Fatty Acids* 48:33-41.

51. Guariento M., Raimondo D., Assfalg M., Zanzoni S., Pesente P., Ragona L., Tramontano A., Molinari H. (2008) Identification and functional characterization of the bile acid transport proteins in nonmammalian ileum and mammalian liver. *Proteins* 70:462-72.
52. Thompson J., Winter N., Terwey D., Bratt J., Banaszak L. (1997) The Crystal Structure of the Liver Fatty Acid-binding Protein. A complex with two bound oleates. *J. Biol. Chem.* 272:7140-50.
53. Haunerland, N., Jagschies, G., Schulenberg, H., Spener, F., (1984) Fatty-acid-binding proteins. *Hoppe-Seyler's, Z. Physiol. Chem.* 365:365-76.
54. Keuper H.J.K., Klein R.A., Spener F. (1985) Spectroscopic investigations on the binding site of bovine hepatic fatty acid binding protein. Evidence for the existence of a single binding site for two fatty acid molecules. *Chem. Phys. Lipids* 38:159-73.
55. Ceciliani F., Monaco H. L., Ronchi S., Faotto L., Spadon P. (1994) The primary structure of a basic (pI 9.0) fatty acid-binding protein from liver of Gallus domesticus. *Comp. Biochem. Physiol*.109:261-71.
56. Ko Y.H., Cheng C.H., Shen T.F., Ding S.T.. (2004) Cloning and expression of Tsaiya duck liver fatty acid binding protein. *Poultry Sci*. 83:1832-38.
57. Vasile F., Ragona L., Catalano M., Zetta L., Perduca M., Monaco H., Molinari H. (2003) Letter to the Editor: Solution structure of chicken liver basic fatty acid binding protein. *J. Biomol. NMR*. 25:157-60.
58. Di Pietro S. M., Veerkamp J. H., Santome´ J. A. (1999) Isolation, amino acid sequence determination and binding properties of two fatty-acid-binding proteins from axolotl (Ambistoma mexicanum) liver. Evolutionary relationship. *Eur. J. Biochem.* 259:127-34.
59. Baba K., Abe T. K., Tsunasawa S., Odani, S. (1999) Characterization and primary structure of a fatty acid-binding protein and its isoforms from the liver of the Amphibia, Rana catesbeiana. *J. Biochem*. 125:115-22.
60. Di Pietro S. M., Corsico B., Perduca M., Monaco H. L., Santome J. A. (2003) Structural and biochemical characterization of toad liver fatty acid-binding protein. *Biochemistry* 42:8192-203.

61. Morales M. H., Ortiz, N. A., Cordero-Lopez N. (1995) Databases GenBank accession number U28756; (Submitted).
62. Di Pietro S. M., Santome´ J. A. (1996) Presence of two new fatty acid binding proteins in catfish liver. *Biochem. Cell. Biol*. 74:675-80.
63. Cordoba O. L., Sanchez E. I., Santome´ J. A. (1999) The main fatty acid-binding protein in the liver of the shark (Halaetunus bivius) belongs to the liver basic type. Isolation, amino acid sequence determination and characterization. *Eur. J. Biochem.* 265:832-38.
64. Denovan-Wright E. M., Pierce M., Sharma M. K., Wright J.M. (2000) Nucleotide sequence of a cDNA clone coding for an intestinal-type fatty acid binding protein and its tissue-specific expression in zebrafish (Danio rerio). *Biochim. Biophys. Acta.* 1492:227-32.
65. Di Pietro S. M., Santome´ J. A. (2001) Structural and biochemical characterization of the lungfish (Lepidosiren paradoxa) liver basic fatty acid binding protein. *Arch. Biochem. Biophys*. 388:81-90.
66. Odani S., Baba K., Tsuchida Y., Aoyagi Y., Wakui S., Takahashi Y. (2001) Hepatic fatty acid-binding proteins of a teleost, Lateolabrax japonicus. The primary structures and location of a disulfide bond. *J. Biochem*. 129:69-76.
67. Jordal A. E., Hordvik I., Pelsers M., Bernlohr D. A., Torstensen B. E. (2006) Comp FABP3 and FABP10 in Atlantic salmon (Salmo salar L.)--general effects of dietary fatty acid composition and life cycle variations. *Biochem. Physiol. B. Biochem. Mol. Biol.* 145:147-58.
68. Nichesola D., Perduca M., Capaldi S., Carrizo M.E., Righetti P.G., Monaco H.L. (2004) Crystal structure of chicken liver basic Fatty Acid-binding protein complexed with cholic acid. *Biochemistry* 43: 14072-79.
69. Tomaselli S., Ragona L., Zetta L., Assfalg M., Ferranti P., Longhi R., Bonvin A.M., Molinari H. (2007) NMR-based modeling and binding studies of a ternary complex between chicken liver bile acid binding protein and bile acids. *Proteins* 69:177-91.
70. Eliseo T., Ragona L., Catalano M., Assfalg M., Paci M., Zetta L., Molinari H., Cicero D.O. (2007) Structural and dynamic determinants of ligand binding in the ternary complex of chicken liver bile acid binding

protein with two bile salts revealed by NMR. *Biochemistry* 46:12557-67.
71. Ragona L., Catalano M., Luppi M., Cicero D., Eliseo T., Foote J., Fogolari F., Zetta L., Molinari H. NMR dynamic studies suggest that allosteric activation regulates ligand binding in chicken liver bile acid-binding protein. *J. Biol. Chem.* 281:9697-709.
72. Pedò M., D'Onofrio M., Ferranti P., Molinari H., Assfalg M. (2009) Towards the elucidation of molecular determinants of cooperativity in the liver bile acid binding protein. *Proteins* 77:718-31.
73. Capaldi S., Saccomani G., Fessas D., Signorelli M., Perduca M. Monaco H. L. (2009) The X-ray structure of zebrafish (Danio rerio) ileal bile acid-binding protein reveals the presence of binding sites on the surface of the protein molecule. *J. Mol. Biol.* 385:99-116.
74. Tochtrop G. P., Bruns J. M., Tang C., Covey D. F. Cistola D. P. (2003) Steroid ring hydroxylation patterns govern cooperativity in human bile acid binding protein. *Biochemistry* 42:11561-7.
75. Tochtrop G. P., DeKoster G. T., Covey D. F. Cistola D. P. (2004) A single hydroxyl group governs ligand site selectivity in human ileal bile acid binding protein. *J. Am. Chem. Soc*. 126:11024-9.
76. Toke O., Monsey J. D., DeKoster G. T., Tochtrop G. P., Tang C. Cistola D. P. (2006) Determinants of cooperativity and site-selectivity in human ileal bile acid-binding protein. *Biochemistry* 45:727-37.
77. Kouvatsos N., Thurston V., Ball K., Oldham N. J., Thomas N. R. Searle M. S. (2007) Bile acid interactions with rabbit ileal lipid binding protein and an engineered helixless variant reveal novel ligand binding properties of a versatile beta-clam shell protein scaffold. *J. Mol. Biol.* 371:1365-77.
78. C. Burtea, Laurent S., Vanijder Elst L., Muller R. N. (2008) Contrast Agents: Magnetic Resonance; In Handbook of Experimental Pharmacology. Vol. 185/1.
79. Aime S., Botta M., Fasano M., Terreno E. (2001) Protein-bound metal chelates; In The chemistry of contrast agents in medical magnetic resonance imaging; John Wiley & Sons, New York.
80. Andre J.P., Toth E., Fischer H., Seelig A., Macke H.R., Merbach A.E.

(1999) High relaxivity for monomeric Gd(DOTA)-based MRI contrast agents, thanks to micellar self-organization. *Chem. Eur. J.* 5:2977-83.
81. Aime S., Botta M., Garino E., Crich S.G., Giovenzana G., Pagliarin R., Palmisano G., Sisti M. (2000) Non-covalent conjugates between cationic polyamino acids and Gd-III chelates: a route for seeking accumulation of MRI-contrast agents at tumor targeting sites. *Chem. Eur. J.* 6:2609-17.
82. Bligh S.W., Harding C.T., Sadler P.J., Bulman R.A., Bydder G.M., Pennock J.M., Kelly J.D., Latham I.A., Marriott J.A. (1991) Use of paramagnetic chelated metal derivatives of polysaccharides and spin-labeled polysaccharides as contrast agents in magnetic resonance imaging. *Magn. Reson. Med.* 17:516-32.
83. Wiener E.C., Brechbiel M.W., Brothers H., Magin R.L., Gansow O.A., Tomalia D.A., Lauterbur P.C. (1994) Dendrimer based metal chelates—a new class of magnetic resonance imaging contrast agents. *Magn. Res. Med.* 31:1-8.
84. Bianchi A., Calabi L., Corana F., Fontana F., Losi P., Maiocchi A., Paleari, L., Valtancoli B. (2000) *Coordination Chemistry Reviews* 204:309-93.
85. Aime S., Fasano M. Terreno E. (1998) Lanthanide(III) chelates for NMR biomedical applications. *Chem. Soc. Rev*. 27:19-29.
86. Zhang Z., Nair S.A., McMurry T.J. (2005) Gadolinium Meets Medical Chemistry: MRI Contrast Agent Development. *Curr. Med. Chem*. 12:751-78.
87. Clement, O., Muhler, A., Vexler V., Berthezene Y., Brasch R. C. (1992) Gadolinium-ethoxybenzyl-DTPA, a new liver-specific magnetic resonance contrast agent. Kinetic and enhancement patterns in normal and cholestatic rats. *Invest. Radiol*. 27:612-19.
88. Vittadini G., Felder E., Musu C., Tirone P. (1990) Preclinical profile of Gd-BOPTA. A liver-specific MRI contrast agent. *Invest. Radiol*. 25:S59-60.
89. Libra A., Fernetti C., Lorusso V., Visigalli M., Anelli P.L., Staud F., Tiribelli C., Pascolo L. (2006) Molecular Determinants in the Transport of a Bile Acid-Derived Diagnostic Agent in Tumoral and Nontumoral Cell

Lines of Human Liver. *J. Pharmacol. Exp. Ther.* 319:809-17.
90. Pascolo L., Petrovic S., Cupelli F., Bruschi C.V., Anelli P.L., Lorusso V., Visigalli M., Uggeri F., Tiribelli C. (2001) ABC Protein Transport of MRI Contrast Agents in Canalicular Rat Liver Plasma Vesicles and Yeast Vacuoles. *Biochem. Biophys. Res. Commun*. 282:60-6.
91. Von Dippe P., Levy D. (1990) Expression of the Bile Acid Transport Protein during Liver Development and in Hepatoma Cells. *J. Biol. Chem.* 265:5942-45.
92. Anelli P. L., Lattuada L., Lorusso V., Lux G., Morisetti A., Morosini P., Serleti M., Uggeri F. (2004) Conjugates of gadolinium complexes to bile acids as hepatocyte-directed contrast agents for magnetic resonance imaging. *J. Med. Chem*. 47:3629-41.
93. De Haen C., Beltrami A., Cappelletti E., Lattuada L., Virtuani M. (2003) Bile acid conjugates with metal ion chelates and the use thereof. U.S. Patent US2003113265.
94. Assfalg M., Gianolio E., Zanzoni S., Tomaselli S., Russo V.L., Cabella C., Ragona L., Aime S., Molinari H. (2007) NMR structural studies of the supramolecular adducts between a liver cytosolic bile acid binding protein and gadolinium(III)-chelates bearing bile acids residues: molecular determinants of the binding of a hepatospecific magnetic resonance imaging contrast agent. *J. Med. Chem*. 50:5257-68.
95. Brunger A.T., Adams P.D., Clore G.M., Delano W.L., Gros P., Grosse-Kunstleve R.W., Jiang J.-S.,Kuszewski J., Nilges N., Pannu N.S., Read R.J., Rice L.M., Simonson T., Warren G.L. (1998) Crystallography and NMR System (CNS): A new software system for macromolecular structure determination. *Acta Cryst*. D54:905-21.
96. Dominguez C., Boelens R. Bonvin A.M. (2003) HADDOCK: a protein-protein docking approach based on biochemical and/or biophysical information. *J. Am. Chem. Soc.* 125:1731-37.
97. De Vries S.J., van Dijk A.D.J., Krzeminski M., van Dijk M., Thureau A., Hsu V., Wassenaar T., Bonvin A.M. (2007) HADDOCK versus HADDOCK: New features and performance of HADDOCK2.0 on the CAPRI targets. *Proteins* 69:726-33.

98. McGovern S.L., Shoichet B.K. (2003) Information decay in molecular docking screens against holo, apo, and modeled conformations of enzymes. *J. Med.Chem.* 46:2895-907.
99. Gordon, S.R. (2006) Fundamentals of Protein NMR Spectroscopy; Springer, Dordrecht.
100. Wüthrich K. (1986) NMR of Proteins and Nucleic Acids; John Wiley & Sons, New York.
101. Kay L. E., Ikura M., Tschudin R., Bax A. (1990) Three-dimensional triple-resonance NMR spectroscopy of isotopically enriched proteins. *J. Magn. Reson*. 89: 496-514.
102. Grzesiek S., Bax A. (1992) Improved 3D Triple-Resonance NMR Techniques Applied to a 31-kDa Protein. *J. Magn. Reson.* 96:432-40.
103. Farmer B.T., Venters L. D., Spicer L. D., Wittekind M.G. Muller L. (1992) A refocused and optimized HNCA: increased sensitivity and resolution in large macromolecules. *J. Biomol. NMR* 2:195-202.
104. Bax A., Ikura M. (1991) An efficient 3D NMR technique for correlating the proton and 15N backbone amide resonances with the alpha-carbon of the preceding residue in uniformly 15N/13C enriched proteins. *J. Biomol. NMR* 1:99-104.
105. Muhandiram D.R., Kay L.E. (1994) Gradient-enhanced triple-resonance three-dimensional proved sensitivity, *J. Magn. Reson.*, *Ser. B* 103:203-16.
106. Clubb R.T., Thanabal V., Wagner G. (1992) A Constant time three-dimensional triple resonance pulse scheme to correlate intraresidue HN, 15N and 13C' chemical shifts in 15N-13C labeled protein *J. Magn. Reson.* 97:213-17.
107. Wittekind M., Mueller L. (1993) HNCACB: A high sensitivity 3D NMR experiment to correlate amide proton and nitrogen resonances with the alpha and beta carbon resonances in proteins. *J. Magn. Reson.* 101:201-5.
108. Grzesiek S., Bax A. (1992) Correlating backbone amide and sidechain resonances in larger proteins by multiple relayed triple resonance NMR. *J. Am. chem. Soc.* 114:6291-93.

109. Grzesiek S., Bax A. (1993) Amino acid type determination in the sequential assignment procedure of uniformly 13C/15N-enriched proteins *J. Biomol. NMR* 3:185-204.

110. Bax A., Clore G.M., Gronenborn A.M. (1990) ^{1}H-^{1}H correlation via isotropic mixing of ^{13}C magnetization, a new three-dimensional approach for assigning ^{1}H and ^{13}C spectra of ^{13}C-enriched proteins. *J. magn. Reson.* 88:425-31.

111. Olejniczak E.T., Xu R.X., Fesik S.W. (1992) A 4D HCCH-TOCSY experiment for assigning the side chain 1H and 13C resonances of proteins. *J. Biomol. NMR* 2:655-59.

112. Wishart D.S., Sykes B.D., Richards F.M. (1992) The chemical shift index: a fast and simple method for the assignment of protein secondary structure through NMR spectroscopy. *Biochemistry* 31:1647-51.

113. Shen Y., Delaglio F., Cornilescu G., Bax A. (2009) TALOS+: A hybrid method for predicting protein backbone torsion angles from NMR chemical shifts. *J. Biomol. NMR* 44:213-23.

114. Shen Y., Vernon R., Baker D., Bax, A. (2009) De novo protein structure generation from incomplete chemical shift assignments. *J. Biomol. NMR* 43:63-78.

115. Marion D., Driscoll P.C., Kay L.E., Wingfield P.T., Bax A., Gronenborn A.M., Clore G.M. (1989) Overcoming the overlap problem in the assignment of 1H NMR spectra of larger proteins by use of three-dimensional heteronuclear 1H-15N Hartmann-Hahn-multiple quantum coherence and nuclear Overhauser-multiple quantum coherence spectroscopy: Application to interleukin 1β. *Biochemistry* 28:6150-56.

116. Marion D., Kay L.E., Sparks S.W., Torchia D.A., Bax A. (1989) Three-dimensional heteronuclear NMR of 15N-labeled proteins. *J. Am. Chem. Soc.* 111:1515-17.

117. Zuiderweg E.R.P., Fesik S.W. (1989) Heteronuclear three-dimensional NMR spectroscopy of the inflammatory protein C5a. *Biochemistry* 28:2387−91.

118. Bertini I., Luchinat C. (1996) NMR of paramagnetic substances. *Coord. Chem. Rev.* 150:1-296.

119. Kogler H., Sorensen O. W., Bodenhausen G., Ernst R. R. (1983) Low-pass J-filters suppression of neighbor peaks in heteronuclear relayed correlation spectra. *J. Magn. Reson*. 55:157-63.
120. Otting G., Senn H., Wagner G., Wuthrich K. (1986) Editing of 2D proton NMR spectra using X half-filters Combined use with residue-selective nitrogen-15 labeling of proteins J. *Magn. Reson*. 70:500-5.
121. Otting G., Wuthrich K. (1988) Efficient purging scheme for proton-detected heteronuclear two-dimensional NMR. *J. Magn. Reson*. 76:569-74.
122. A. L. Breeze (2000) Isotope-filtered NMR methods for the study of biomolecular structure and interactions. *Prog. NMR Spectrosc.* 36:323-72.
123. Ikura M., Bax A. (1992) Isotope-Filtered 2D NMR of a Protein-Peptide Complex: Study of a Skeletal Muscle Myosin Light Chain Kinase Fragment Bound to Calmodulin. *J. Am. Chem. Soc.* 114:2433-40.
124. Zwahlen C., Legault P., Vincent S.J.F., Greenblatt J., Konrat R., Kay, L.E. (1997) Methods for measurement of intermolecular NOEs by multinuclear NMR spectroscopy: Application to a bacteriophage λ N-peptide/*boxB* RNA complex. *J. Am. Chem. Soc.* 119: 6711-21.
125. Lee W., Revington M. J., Arrowsmith C., Kay L. E. (1994) A pulsed field gradient isotope-filtered 3D ^{13}C HMQC-NOESY experiment for extracting intermolecular NOE contacts in molecular complexes. *FEBS Lett.* 350:87−90.
126. Guntert P. (2004) Automated NMR structure calculation with CYANA. *Methods Mol. Biol*. 278:353-78.
127. Herrmann T., Guntert P., Wuthrich K. (2002) Protein NMR structure determination with automated NOE assignment using the new software CANDID and the torsion angle dynamics algorithm DYANA. *J. Mol. Biol.* 319:209-27.
128. Vuister G.W., Doreleijers J. F., da Silva A.W.S. CING: Validation Tools for Biomolecular NMR Structures.
129. Bhattacharya A., Tejero R., Montelione G.T. (2007) Evaluating protein structures determined by structural genomics consortia *Proteins*. 66:778-95.

130. Linge J. P., O'Donoghue S. I., Nilges M. (2001) Automated assignment of ambiguous nuclear overhauser effects with ARIA. *Methods Enzymol*. 339:71-90.
131. Eufri D., Sironi A. (1989) SMILEs shaded molecular imaging on low-cost equipment. *J. Mol. Graph.* 7:165–69.
132. Schuettelkopf A. W., van Aalten D. M. F. (2004) PRODRG - a tool for high-throughput crystallography of protein-ligand complexes. *Acta Cryst.* D60:1355-63.