

# NMR-based modeling and binding studies of a ternary complex between chicken liver bile acid binding protein and bile acids

Simona Tomaselli,<sup>1</sup> Laura Ragona,<sup>1\*</sup> Lucia Zetta,<sup>1</sup> Michael Assfalg,<sup>2</sup> Pasquale Ferranti,<sup>3</sup> Renato Longhi,<sup>4</sup> Alexandre M.J.J. Bonvin,<sup>5</sup> and Henriette Molinari<sup>2\*</sup>

<sup>1</sup> Laboratorio NMR, ISMAC, CNR, via Bassini 15, 20133, Milano, Italy

<sup>2</sup> Dipartimento Scientifico e Tecnologico, Università di Verona, Strada Le Grazie 15, 37134, Verona, Italy

<sup>3</sup> Dipartimento di Scienza degli Alimenti, Università di Napoli Federico II, Parco Gussone, Portici 80055, Italy

<sup>4</sup> Istituto di Chimica del Riconoscimento Molecolare, CNR, via M. Bianco, Milano, Italy

<sup>5</sup> NMR Laboratory Faculty of Sciences, Bijvoet Center for Biomolecular Research, Utrecht University, 3584CH Utrecht, The Netherlands

## ABSTRACT

Chicken liver bile acid binding protein (cL-BABP) is involved in bile acid transport in the liver cytosol. A detailed study of the mechanism of binding and selectivity of bile acids binding proteins towards the physiological pool of bile salts is a key issue for the complete understanding of the role of these proteins and their involvement in cholesterol homeostasis. In the present study, we modeled the ternary complex of cL-BABP with two molecules of bile salts using the data driven docking program HADDOCK on the basis of NMR and mass spectrometry data. Docking resulted in good 3D models, satisfying the majority of experimental restraints. The docking procedure represents a necessary step to help in the structure determination and in functional analysis of such systems, in view of the high complexity of the 3D structure determination of a ternary complex with two identical ligands. HADDOCK models show that residues involved in binding are mainly located in the C-terminal end of the protein, with two loops, CD and EF, playing a major role in ligand binding. A spine, comprising polar residues pointing toward the protein interior and involved in motion communication, has a prominent role in ligand interaction. The modeling approach has been complemented with NMR interaction and competition studies of cL-BABP with chenodeoxycholic and cholic acids. A higher affinity for chenodeoxycholic acid was observed and a  $K_d$  upper limit estimate was obtained. The binding is highly cooperative and no site selectivity was detected for the different bile salts, thus indicating that site selectivity and cooperativity are not correlated. Differences in physiological pathways and bile salt pools in different species is discussed in light of the binding results thus enlarging the body of knowledge of BABPs biological functions.

Proteins 2007; 69:177–191.  
© 2007 Wiley-Liss, Inc.

**Key words:** bile acid binding protein; ternary complex; binding selectivity; HADDOCK; NMR; dissociation constant.

## INTRODUCTION

Bile salts have long been known to facilitate digestion and absorption of lipids in the small intestine. Recently it has been demonstrated that bile acids, undergoing enterohepatic circulation from liver to intestine, also function as hormones by activating ligand-inducible transcription factors that are members of the superfamily of nuclear hormone receptors, such as farnesoid X receptor and liver X alpha receptor.<sup>1</sup> This activation provides a strong link between bile acids production and cholesterol homeostasis. Indeed in humans, inherited, genetic defects associated with the impairment of bile acid homeostasis lead to severe cholestatic disease and vitamin absorption deficiency.<sup>2</sup>

An important question is related to the mechanism by which bile acids, poorly soluble in aqueous environment, traverse the cytoplasm, enter the nucleus, and reach their cognate receptors. It is known that in cell bile acids bind to intracellular lipid binding proteins (iLBPs), a family of small proteins (ca. 15 kDa) that, in spite of the remarkably similar  $\beta$ -barrel structures, diverge in their primary structure and in the selectivity exhibited towards different lipophilic ligands.<sup>3</sup> It has been recently suggested that the role of iLBPs may not only be limited to the transport of bile acids in cytoplasm, but could be related to

The Supplementary Material referred to in this article can be found at <http://www.interscience.wiley.com/jpages/0887-3585/suppmat/>

Grant sponsor: Italian Ministry for Education, University and Research; Grant numbers: FIRB 2003 RBNE03B8KK, MIUR 2004; Grant sponsor: Large Scale Facility SONNMR of Utrecht University; Grant number: FP6 EU-NMR 026145; Grant sponsor: FP6 NDDP STREP; Grant number: 512077.

\*Correspondence to: Henriette Molinari, Dipartimento Scientifico e Tecnologico, Università di Verona, Strada Le Grazie 15, 37134 Verona, Italy. E-mail: molinari@sci.univr.it or Laura Ragona, ISMAC, CNR, via Bassini 15, 20133, Milano, Italy. E-mail: laura.ragona.imac.cnr.it

Received 11 January 2007; Revised 6 March 2007; Accepted 13 March 2007  
Published online 2 July 2007 in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/prot.21517

subcellular targeting.<sup>4</sup> Studies of cellular localization of a few iLBPs, such as CRAB-II, K-FABP, and A-FABP, revealed that these proteins are indeed cytosolic in the absence of ligands; they however translocate to the nucleus upon treatment of cells with appropriate ligands, that have the capability of switching a rearrangement that results in the presence of a nuclear localization signal (NLS) with a topology strikingly similar to the classic NLS.<sup>4</sup> It is however still under debate whether this translocation mechanism is common to all proteins of the family or not. In this line we have recently reported that the mechanism of ligand binding in chicken liver bile acid binding protein (cL-BABP), and possibly in the other proteins of the family, is consistent with a model of allosteric activation, where the protonation state equilibrium of a buried histidine modulates a functionally important motion, that is, the opening/closure of loops at the open end, and ligand binding stabilizes one of the preexisting conformations.<sup>5</sup>

cL-BABP consists of 10 antiparallel  $\beta$ -strands (A-J) organized in two nearly orthogonal  $\beta$ -sheets that form a  $\beta$ -clam type structure. Helices I and II, inserted between the A and B strands, together with the CD, EF, GH, and IJ loops define the protein “open end”, that is, the entrance to the binding cavity. The formation of a ternary complex of cL-BABP with 2 molecules of bile salts has been previously demonstrated.<sup>5,6</sup>

A detailed study of the mechanism of binding and selectivity of bile acids binding proteins (BABP) towards the pool of bile salts has become a key issue for the complete understanding of the role of these proteins and their involvement in cholesterol homeostasis, as this pathway is a target for pharmaceutical intervention in patients with elevated cholesterol levels or metabolic syndrome.<sup>7,8</sup> Pioneering work in the field has been reported for human ileal bile acid binding protein (hI-BABP), which was shown to exhibit low intrinsic affinity but a very high degree of positive cooperativity and site selectivity in its interactions with glycolate (GCA) and glycochenodeoxycholate (GCDA), the two major bile salts in humans.<sup>9,10–12</sup>

In the present article we address structural aspects, mechanism and site selectivity of bile acid binding to cL-BABP, a member of the BABP family. This protein has been suggested to perform bile acid transport in non-mammals liver cytosol, thus playing the same role that ileal BABP performs in enterocytes.<sup>5</sup> The guiding principle of this research is that only the correlation between binding data reported for different proteins of the same family would add to the establishment of firm guidelines for the determinants of binding.

In the first part of the article we present a model of the ternary complex of cL-BABP with two molecules of chenodeoxycholic acid (CDA), the most abundant bile salt in chicken, obtained using the data-driven docking program HADDOCK.<sup>13</sup> For this, two CDA molecules were docked simultaneously (three body docking) into

the binding pocket of cL-BABP based on multiple sets of NMR and mass spectrometry data. Flexible proteins exhibiting more than one binding site place severe demands on the docking procedures, and we show here the efficiency of the used multibody docking algorithm in dealing with a multisite interaction problem. Our results provide a basis for a thorough analysis of the binding mode of other proteins of the family currently under investigation in our laboratory. The models obtained represent a valuable and necessary contribution to the NMR 3D structural determination, which is not a trivial problem due to the presence of two identical ligand molecules and of resonance overlap in the aliphatic region. In the second part of the article an NMR binding competition study is presented for cL-BABP at different temperatures to investigate site selectivity and cooperativity towards chenodeoxycholic and cholic (CA) acids. The obtained results are discussed in the light of the binding properties reported for hI-BABP. An analysis of physiological pathways and bile salt pools in different species has been performed to trace the biological implications of the present findings.

## MATERIALS AND METHODS

### Protein expression and purification and ligand synthesis

Recombinant cL-BABP was expressed from *E. coli* and purified to homogeneity as previously described.<sup>5</sup> Delipidated cL-BABP was obtained in a yield of 90 mg/L of rich medium. <sup>13</sup>C and <sup>15</sup>N isotope labeling was achieved using M9 minimal media containing 1 g/L <sup>15</sup>NH<sub>4</sub>Cl, and 4 g/L <sup>13</sup>C enriched glucose following protocols reported in the literature.<sup>14</sup> The extent of <sup>15</sup>N and <sup>13</sup>C labeling was verified by MALDI mass analysis and the isotope incorporation was found to be more than 92%. <sup>13</sup>C-<sup>15</sup>N cL-BABP was obtained in a yield of 25 mg/L of minimal media. Protein concentrations, for samples preparation, were determined spectrophotometrically. Unenriched bile acids and [24-<sup>13</sup>C]-glycocholate (GCA) were purchased from Sigma. <sup>15</sup>N glycine conjugates of bile acids were prepared as follows: <sup>15</sup>N Fmoc-Glycine was anchored to 2-chlorotrityl chloride resin in the presence of *N,N*-diisopropylethylamine according to the procedures of Barlos *et al.*<sup>15</sup> After removal of the Fmoc group by 20% piperidine in dimethylformamide and further washings, a pre-activated pentafluorophenyl ester of CDA or CA in dimethylformamide was added in the presence of 1-hydroxybenzotriazole to <sup>15</sup>N-Glycine anchored to the resin and vortexed until the “Kaiser test”<sup>16</sup> was negative. After extensive washings to remove the reagents, GCDA and GCA <sup>15</sup>N enriched in the glycine moiety were removed from the resin by repeated treatment with 25% 1,1,1,3,3,3-hexafluoro-2-propanol in dichloromethane. Solvents were removed in vacuum, the adduct was dis-

solved in aqueous acetonitrile and concentrated, with concomitant precipitation of the products, recovered by filtration. The homogeneity of the derivatives was assessed by Thin Layer Chromatography.

For the preparation of the holo-protein samples, a procedure previously described was followed.<sup>9</sup> Bile salts were dissolved in tetrahydrofuran, and the concentration of the stock solutions was determined by measuring dry weights. Appropriate amounts of the stock solution of bile acid were aliquoted, and the solvent was evaporated under a stream of nitrogen. The bile acid was solubilized using 1:1 equivalents of 1M NaOH. Then 17.5  $\mu\text{L}$  of  $\text{D}_2\text{O}$  were added to 332.5  $\mu\text{L}$  of protein solution to bring the total volume of the NMR sample to 350  $\mu\text{L}$ . The final 0.5 mM protein samples were dissolved in 30 mM potassium phosphate buffer, in 95% $\text{H}_2\text{O}$ /5% $\text{D}_2\text{O}$ . The pH of the solutions was 7.2. To minimize errors each ligand:protein mole ratio sample was prepared twice.

### NMR experiments

All NMR spectra were recorded on Bruker DMX and DRX spectrometers operating at 500 MHz equipped with a triple resonance probehead, incorporating gradients in the  $z$ -axis. Sensitivity-enhanced  $^1\text{H}$ - $^{15}\text{N}$  heteronuclear correlation spectra were collected using standard pulse sequences. All HSQC spectra were acquired with a  $^1\text{H}$  spectral width of 6510 Hz and 1024 points, zero-filled to a total of 2048 points. Relaxation delays of 1.7 s were employed. In the  $^{15}\text{N}$  dimension, 256 increments were collected, with a sweep width of 2032 Hz, zero-filled to a total of 1024 points.

HMBC<sup>17</sup> and 3D  $^1\text{H}$ - $^{15}\text{N}$  HSQC-NOESY<sup>18</sup> spectra were acquired using standard pulse sequences. The delay in the HMBC sequence was optimized for a long range coupling constant of 5.3 Hz. In the 3D  $^1\text{H}$ - $^{15}\text{N}$  HSQC-NOESY spectra 1024 and 128 points were used for  $^1\text{H}$  first and third dimension, respectively with a spectral width of 6510 Hz; in the  $^{15}\text{N}$  dimension 20 increments were collected with a sweep width of 1013 Hz. A mixing time of 250 ms was employed.

For the acquisition of saturation transfer difference (STD) NMR spectra<sup>19</sup> a sample of cL-BABP (70  $\mu\text{M}$ ) in the presence of CDA (1.6 mM) in 99.9%  $\text{D}_2\text{O}$  30 mM buffer phosphate was prepared and a 1D pulse sequence incorporating a  $T_{1\rho}$  filter was used. Spectra were recorded with a spectral width of 6510 Hz and 32 K data points. On-resonance and off resonance irradiations were performed at 3380 and  $-15000$  Hz, respectively, using 60 Gaussian pulses with a 1% truncation and 50 ms duration to give a total saturation time of 3 s. The duration of the  $T_{1\rho}$  filter was 30 ms. STD NMR spectra were acquired with a total of 128 transients in addition to 16 scans to allow the sample to reach equilibrium. The subtraction of the two FIDs was achieved by phase cycling.

Reference spectra on free ligand were acquired using the same conditions.

An  $\text{F}_2$ - $^{15}\text{N}$ , $^{13}\text{C}$ -filtered NOESY experiment was run using a standard Bruker pulse sequence that employs simultaneous  $^{15}\text{N}$  and  $^{13}\text{C}$  low-pass  $J$ -filters.<sup>20</sup> The two, two-step  $^{13}\text{C}$  purge filters were tuned to aromatic  $^1J_{\text{CH}}$  couplings (150 and 170 Hz). A mixing time of 70 ms was used. The spectrum was registered with a  $^1\text{H}$  spectral width of 6009 Hz, 512 scans, and  $2048 \times 256$  data points. For this experiment, a  $^{15}\text{N}$ , $^{13}\text{C}$ -labeled protein: unlabeled ligand (1:2.5) sample was dissolved in 90% $\text{D}_2\text{O}$ /10% $\text{H}_2\text{O}$  and the residual water signal was suppressed using the Watergate scheme.

Data were processed with XWINNMR and NMRPipe<sup>21</sup> and peak volume integrations were carried out with NMRView 5.0.3 software package.<sup>22</sup> To decrease the systematic errors measurements were repeated twice under identical conditions. The derived standard deviation is reported on the graph as an error-bar.

Spectra calibration was performed with TSP as internal standard for  $^1\text{H}$ , whereas for  $^{15}\text{N}$  and  $^{13}\text{C}$  dimension an indirect referencing was used as previously described.<sup>23</sup>

### Mass spectrometry

Limited Proteolysis was performed for apo- and holo-cL-BABP in complex with CDA. The samples were digested with trypsin (enzyme/substrate 1:1000 w/w) in 0.1% ammonium bicarbonate buffer, pH 7.8. Aliquots of the incubation mixtures, corresponding to 5 nmol of the original protein, were withdrawn at times ranging from 15 s to 30 min and freeze-dried. The digests were then dissolved in 0.1 mL of aqueous 0.1% trifluoroacetic acid and analyzed by MALDI TOF MS or by Liquid Chromatography Electrospray Mass Spectrometry (LC-ESI-MS). MALDI TOF MS analysis was performed on a DE-Pro (Applied Biosystems) instrument operating in reflector mode. For LC-ESI-MS analysis, a 2.1 mm i.d.  $\times$  250 mm, C18, 5 mm (Vydac) reverse-phase column with a flow rate of 0.2 mL  $\text{min}^{-1}$  was used on a Agilent 1100 MSD system. For peptide analysis, the separation was carried out with a linear gradient from 8 to 40% acetonitrile containing 0.1% (v/v) TFA over 60 min and mass spectra were acquired in the range 1600–400 m/z at scan cycle of 5 s/scan. The source temperature was 393 K. Mass scale calibration was carried out using myoglobin as reference compound.

### Docking calculations

Docking of bile salts to cL-BABP was performed using the software HADDOCK (version 2.0\_dev)<sup>13</sup> in combination with CNS.<sup>24</sup> An ensemble of 10 NMR apo protein structures<sup>5,25</sup> was used as starting structures for docking. CDA coordinates were derived with the SMILE<sup>26</sup> program by changing the hydroxylation pattern of a CA molecule (<http://xray.bmc.uu.se/hicup/>).

Protein residues directly affected by ligand interaction are defined as “active residues”, while neighbors of active residues are defined as “passive residues”. Active and passive residues form the protein/ligand interface. Since the two CDA ligand molecules can not be distinguished, the Ambiguous Interaction Restraints (AIRs) were defined ambiguously to both ligands letting HADDOCK choose the best conformation. In that way no bias is introduced in the modeling. Additional restraints were introduced in the flexible stages of HADDOCK to maintain the secondary structure of the protein by defining dihedral angle restraints (measured value  $\pm 20^\circ$ ) for  $\beta$ -sheet and  $\alpha$ -helical regions, derived from the secondary structural elements of the apo structure (PDB id: 1MVG). Dihedral angle prediction from secondary chemical shift analysis performed by TALOS,<sup>27</sup> supports the use of such restraints.

The HADDOCK docking protocol consisted of the following three consecutive stages<sup>13</sup>:

- i. randomization of orientations followed by rigid body energy minimization (EM);
- ii. semiflexible simulated annealing in torsion angle space, which consists of (a) a rigid body Molecular Dynamics search and a first simulated annealing, (b) a semiflexible simulated annealing during which only side chains at the interface are free to move, and (c) a semiflexible simulated annealing during which only side chains and backbone at the interface are free to move;
- iii. final refinement in Cartesian space with explicit solvent during which side chains and backbone at the interface are free to move.

The rigid body docking stage is performed a number of times (in our case: 5 times), and the best resulting structure is saved. 4000 structures were generated at the rigid body docking stage, the best 400 of which were selected for further semiflexible refinement. The best 200 structures were then selected for the final refinement in explicit water. Nonbonded energies (sum of van der Waals and electrostatic terms) were calculated with an 8.5 Å distance cut-off using the OPLS nonbonded parameters<sup>28</sup> from the parallhdg5.3.pro parameter file.<sup>29</sup> Parameters for the ligands were obtained from the PRODGR server.<sup>30</sup> The dielectric constant  $\epsilon$  was set to 10.0 to damp the electrostatic contribution in vacuum and to 1.0 during explicit solvent refinement. The overall HADDOCK score used at the various stages to rank and select solutions was calculated as a weighted sum of different terms (rigid body stage:  $0.01 \cdot E_{vdW} + 1.0 \cdot E_{elec} + 0.01 \cdot E_{AIR} - 0.01 \cdot BSA \pm 1.0 \cdot E_{desolv}$ ; semiflexible refinement:  $1.0 \cdot E_{vdW} + 1.0 \cdot E_{elec} + 0.1 \cdot E_{AIR} - 0.01 \cdot BSA + 1.0 \cdot E_{desolv}$ ; water-refinement:  $1.0 \cdot E_{vdW} + 0.2 \cdot E_{elec} + 0.1 \cdot E_{AIR} + 1.0 \cdot E_{desolv}$ ). Here the subscripts are: vdW the van der Waals energy, Elec the electrostatic energy, AIR the AIR energy, BSA the buried surface area and Desolv the desolvation energy calculated using the atomic desolvation parameters of Fernandez-Recio *et al.*<sup>31</sup>

The resulting solutions were clustered using the algorithm of Daura *et al.*<sup>32</sup> with a 3 Å cut off according to RMSDs of the two ligands calculated after fitting on the backbone of the flexible interface for the protein. Docking solution could be divided into 21 clusters, which were ranked based on the average HADDOCK score (see earlier) calculated on the top two solutions (see supplementary Table). All the protein ligand interactions of the models were analyzed using the LIGPLOT program.<sup>33</sup> The RMSD at ligands coordinates (carbons) were calculated after global superposition of protein backbone atoms using PROFIT program (Martin, A. C. R., ProFit. <http://www.bioinf.org.uk/software/profit>).

## RESULTS

The ternary complex of cL-BABP with 2 molecules of CDA has been modeled using the HADDOCK program, which can incorporate experimental information obtained at various levels, such as NMR chemical shift perturbation analysis, <sup>15</sup>N relaxation measurements, NOE, mutagenesis data, and any other known attribute of the complex under study. In the following paragraphs a description of all the experimental data used to drive the docking simulation is reported. Protein residues directly involved in binding were annotated as “active residues”, while all residues sequentially and structurally close to an active residue were defined “passive residues”. Active and passive residues are summarized in Table I.

### Mass spectrometry analysis of apo- and holo-cL-BABP limited proteolysis

The accessibility of apo- and holo-cL-BABP was investigated with limited proteolysis experiments using trypsin. This protease was selected since cL-BABP presents a large number of potential hydrolysis sites (thirteen lysine and five arginine residues). Fragments 1–52 and 1–55 are present in the chromatogram of the apo protein after 5 min and both are quickly further digested at the level of R32. The peptide fragment 53–76 is produced after only 5 min of incubation, and after 10 min fragment 56–76 becomes detectable, while it proves quite resistant to further hydrolysis, possibly because of the presence of the stable K66-E67 bond. In summary, limited proteolysis experiments on apo cL-BABP showed that R32 (C-terminal end of Helix II), K52, R55 (CD loop), and K76 (EF loop) are the primary cleavage sites.

In proteolysis experiments on cL-BABP complexed with CDA, fragments 53–76 and, subsequently, 56–76 were observed only after 20 min, suggesting a higher protection in the presence of the ligands. Fragment 1–29 is observed after 30 min. Residues K52, K76, and K29 appear therefore to be the primary cleavage sites. No cleavage was observed at sites R32 and R55, and both residues were considered as “active”.

**Table 1**

Active and Passive Residues of cL-BABP

Active	Passive
Y14 (NOE <sup>a</sup> )	N13
F17 (NOE)	E15
L21 (NOE)	E16
R32 (CSP, Proteolysis)	L18
T53 (CSP)	A20
R55 (Proteolysis)	A22
T57 (CSP)	A31
E67 (CSP, Dynamics)	D33
A68 (CSP, Dynamics)	K52
T72 (CSP, Dynamics)	P54
M73 (Dynamics)	Q56
D74 (CSP)	V58
G75 (CSP)	K66
K76 (CSP)	D69
K79 (CSP)	T71
C80 (CSP)	K77
T81 (CSP, Dynamics)	L78
L89 (CSP, Dynamics)	V82
V90 (Dynamics)	K88
T91 (CSP)	K92
S93 (Dynamics)	E94
K95 (Dynamics)	E101
F96 (CSP, NOE)	T110
S97 (CSP, Dynamics)	T112
H98 (CSP, NOE)	
E99 (Dynamics)	
Q100 (Dynamics)	
I111 (CSP, Dynamics)	

<sup>a</sup>The experimental method used to define the residue as active is indicated in parenthesis. CSP, chemical shift perturbation data; Dynamics, dynamics data from <sup>15</sup>N relaxation measurements; NOE, NOESY data; Proteolysis, mass spectrometry analysis of limited proteolysis data.

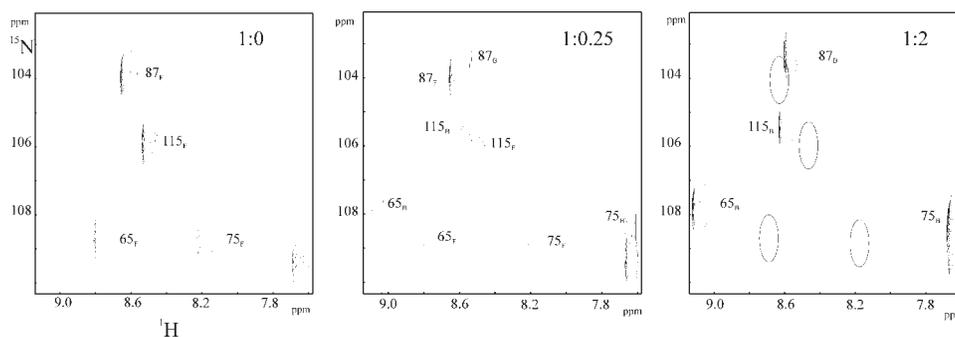
### Protein chemical shift perturbation experiments

The complete assignment of <sup>1</sup>H and <sup>15</sup>N apo- and holo-cL-BABP, in complex with CDA, has been previ-

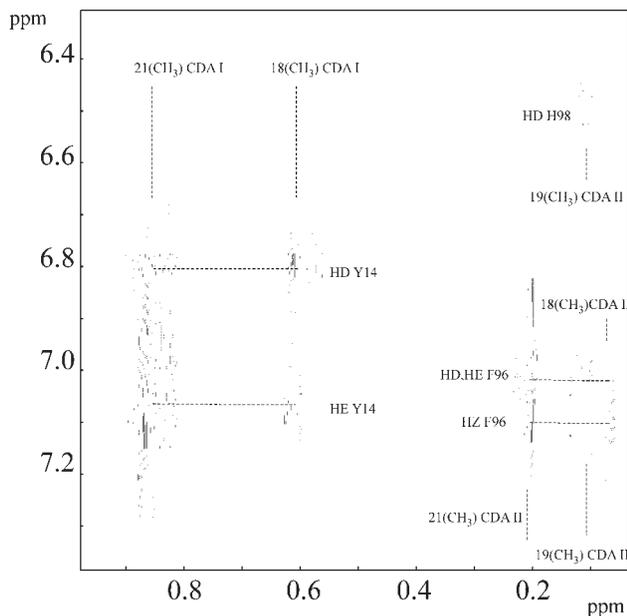
ously reported and deposited in the BioMagResBank (entry code 6642). <sup>1</sup>H and <sup>15</sup>N shifts changes were defined as<sup>34</sup>  $\Delta\delta(\text{HN},\text{N}) = [(\Delta\delta_{\text{HN}}^2 + \Delta\delta_{\text{N}}^2/25)/2]^{1/2}$  and residues showing protein chemical shift perturbation (CSP) higher than the average value plus one standard deviation (CSP > 0.4 ppm), namely R32, T53, T57, E67, A68, T72, D74, G75, K76, K79, C80, T81, L89, T91, F96, S97, H98, I111, were considered affected by ligand binding. To obtain information on the exchange regime bile salt interactions were further investigated by titrating <sup>15</sup>N cL-BABP with GCDA: successive addition of bile salt, from 0 to 3 molar equivalents, induced chemical shift changes occurring mostly in the slow exchange regime, with respect to the NMR timescale (Fig. 1): only peaks corresponding to the apo- and holo-states were observed during the titration. This result suggests that the dissociation constant  $K_d$  of cL-BABP with GCDA is relatively low (vide infra). It is worth mentioning that holo-proteins in complex with CDA or GCDA exhibit very similar HSQC spectra, except for few residues in the loops CD (R55), EF (M73), and GH (K95), due to their proximity to the glycine moiety, as suggested by docking results (vide infra). This observation indicates that the overall holo-structure is conserved upon binding of the bile salt or of its glycine conjugate.

### Intermolecular NMR NOE data

To identify protein–ligand intermolecular interactions two kinds of experiments were performed: (i) a 3D <sup>15</sup>N-NOESY-HSQC on the unlabeled cL-BABP:<sup>15</sup>N GCDA 1:2.5 complex and (ii) a 2D isotope-filtered experiment on the doubly enriched <sup>15</sup>N, <sup>13</sup>C cL-BABP:unlabeled CDA 1:2.5 complex. The first 3D experiment allowed the identification of NOEs between glycine amide group of GCDA and L21 H $\delta$  and F17 H $\epsilon$  protons. Figure 2 shows the spectral region of 2D isotope filtered experiment

**Figure 1**

2D <sup>1</sup>H-<sup>15</sup>N-HSQC spectra of 0.5 mM <sup>15</sup>N cL-BABP recorded at different protein:GCDA ratios: 1:0 (left panel); 1:0.25 (central panel); and 1:2 (right panel). Assignments are indicated on the spectra with the number of the corresponding amide. Subscripts F and B refer to free and bound species. Empty circles in the right panel indicate the position of the apo resonance of the indicated peak.



**Figure 2**

2D  $^1\text{H}$ - $^1\text{H}$  isotope-filtered NOESY performed on the doubly enriched  $^{13}\text{C}$ ,  $^{15}\text{N}$  cL-BABP in complex with unlabeled CDA (1:2.5 P:L ratio). The drawn spectral region exhibits intermolecular NOEs among protein aromatic residues and methyl groups belonging to the two CDA ligands. Only unambiguously assigned peaks were labeled.

exhibiting intermolecular NOEs among protein aromatic residues, namely Y14, F96, and H98, and ligand methyl groups. NOEs could be observed for methyl resonances corresponding to the two bound CDAs. The complete assignment of the holo-protein and ligands will be reported elsewhere (manuscript in preparation).

### Differences in dynamics between apo- and holo-cL-BABP

$^{15}\text{N}$  relaxation measurements were performed for cL-BABP, in its apo and holo-form, at three different magnetic fields.<sup>5</sup> In the apo protein, a few residues located in the C-terminal end of the protein, namely residue E67, A68, T72, T81, L89, S97, I111 exhibited motions on the millisecond time scale, as defined by the evaluation of  $R_{\text{ex}}$  contributions. These  $R_{\text{ex}}$  contributions were completely quenched upon ligand binding, therefore providing the information that such residues are actively involved in binding.

In addition, six cross peaks, corresponding to residues M73, V90, S93, K95, E99, Q100, missing in the  $^1\text{H}$ - $^{15}\text{N}$  HSQC spectra of apo protein appeared in the spectra of the complex with two CDA molecules, indicating that the ligands stabilize one conformation, thus quenching exchange broadening. These residues were additionally annotated as “active”.

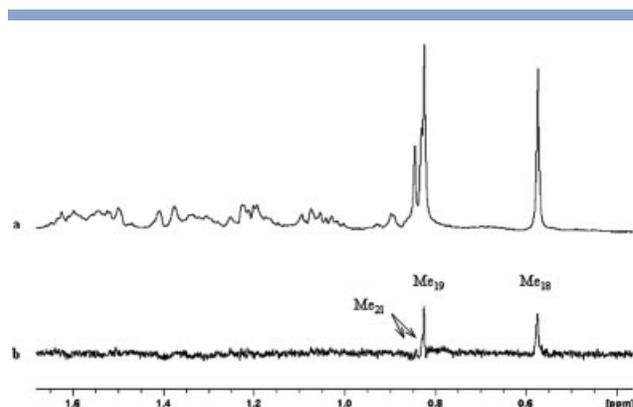
### STD experiments

To map the CDA protons in more intimate contact with the protein, 1D STD experiments were performed. The observed low dissociation constant is still suitable for STD analysis,<sup>19</sup> however the low molecular weight of cL-BABP (14 kDa) makes the setup of this experiment particularly demanding as a compromise should be found between the high power saturation pulses needed for full protein saturation by spin-diffusion and the low power saturation pulses needed to avoid direct irradiation of ligand resonances.

The STD spectrum showed the resonances of methyl protons bound to C18, C19, and C21 (Fig. 3). The selective saturation of the protein was performed by irradiating an aromatic resonance 1500 Hz apart from any ligand resonance. A STD control spectrum in the same experimental conditions on the ligand alone did not show any signal, thus excluding direct excitation of the bile acid. The results obtained from this experiment are in good agreement with the intermolecular NOEs between the methyl groups of the ligands and aromatic protons of Y14 and F96 observed in isotope filtered NMR experiments (Fig. 2).

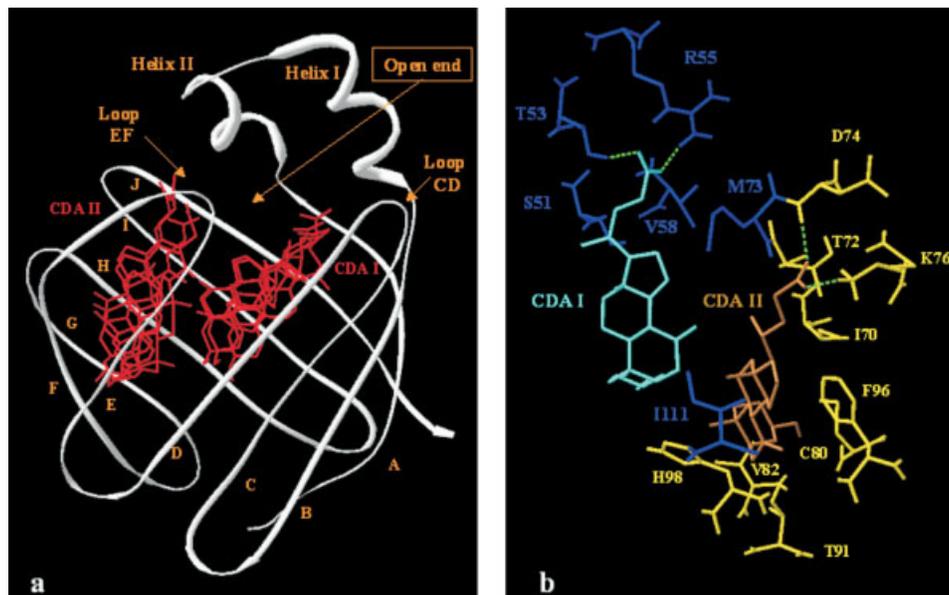
### Docking studies

HADDOCK calculations were started with the coordinates of apo cL-BABP (PDB id: 1MVG). An ensemble of 10 NMR structures was used to account for loops flexibility and different orientations of solvent accessible side-chains. To force the ligands to enter the cavity of the protein, additional distance restraints with an upper bound distance of 3 Å were defined between the centers of mass of the protein ( $\text{C}\alpha$  atoms only) and of the ligands (all atoms). These were only used during the rigid body docking stage to allow the two ligands to



**Figure 3**

Reference 1D NMR spectrum of cL-BABP (70  $\mu\text{M}$ ) in the presence of 1.6 mM CDA recorded at 500 MHz at 298 K (a). STD spectrum of the same sample (b). The assignment of the signals is reported.



**Figure 4**

HADDOCK models of cL-BABP complexed with two CDAs. Superposition of the best four structures. Secondary structural elements are labeled (Panel a). Side-chains of residues involved in CDA interaction are shown in yellow and blue for the internal<sup>35</sup> and superficial (cyan) CDA molecules, respectively. Green dotted lines indicate H-bonds. Residue labeling is reported (Panel b). The two views are different for clarity purposes.

freely explore the cavity and find their best arrangement during the subsequent semiflexible refinement.

Docking of a ternary complex requires definition of AIR between the protein and both CDA molecules. The used experimental techniques, with the only exception of NOE data, did not allow to distinguish the effect of one ligand molecule from the other. To avoid introducing any structural bias we did not differentiate the two binding sites and performed the docking with AIRs defined simultaneously to both ligands, except for three AIR involving Y14, F96, and H98, attributed unambiguously to one single ligand (see Fig. 2). HADDOCK successfully positioned the two ligands in the internal cavity of the protein, showing one CDA molecule more buried in the  $\beta$ -barrel with the carboxyl group pointing towards EF loop (internal binding site), so forth called CDA II, and the second positioned in a more solvent accessible region, close to the portal area (superficial binding site), so forth called CDA I.

The four best solutions of the best ranking cluster (Cluster no. 2, see supplementary Table) are visualized as a bundle in Figure 4 while the structural statistics of the ternary models are reported in Table II. The superposition of the four HADDOCK models indicates that the inner molecule is better defined with an average RMSD of 1.9 Å, with respect to the lowest energy structure, while the CDA located in the more superficial site presents a higher variability with a RMSD of 3.0 Å. Pro-

tein/ligand hydrophobic and H-bond interactions were analyzed and a list of cL-BABP residues involved in binding is reported in Table III. All residues making interactions in HADDOCK models belong to active or passive residues, with the exception of S51 (C strand), for which we did not observe any significant CSP. A possible expla-

**Table II**

Structural Statistics of the Four Best cL-BABP/CDA Model Structures

Backbone RMSD (Å) with respect to mean	
Flexible interface backbone	1.04 ± 0.5
All backbone	1.00 ± 0.5
Number of ambiguous interaction restraints (AIRs)	
Total AIRs	42
Number of AIR violations	
Total AIR violations	13
Intermolecular energies after water refinement	
$E_{vdw}$ (kcal mol <sup>-1</sup> )	-33 ± 7
$E_{elec}$ (kcal mol <sup>-1</sup> )	-316 ± 63
Buried surface area (Å <sup>2</sup> )	1525 ± 29
RMSD from idealized covalent geometry	
Bonds (Å)	0.004 ± 0.00
Angles (°)	0.70 ± 0.02
Impropers (°)	0.59 ± 0.04
Dihedrals (°)	22.6 ± 1.4
Ramchandran analysis	
Residues in the favoured region (%)	84
Residues in the additional allowed regions (%)	14
Residues in generously allowed regions (%)	2
Residues in disallowed regions (%)	0

**Table III**

Hydrophobic Interactions, H-Bond and Salt Bridges Present in at least Three of the Best Four Structures of the Complex cL-BABP/CDA Calculated by HADDOCK

	H-bond/salt bridges	Hydrophobic interactions
Inner CDA	T72 (EF loop)	<b>T72<sup>a</sup></b> (EF loop)
	D74 (EF loop)	<b>C80</b> (F strand)
	K76 (EF loop)	V82 (F strand)
Superficial CDA		T91 (G strand)
		F96 (H strand)
		<b>H98</b> (H strand)
	T53 (CD loop)	S51 (C strand)
		V58 (D strand)
	R55 (CD loop)	M73 (EF loop)
		<b>H98</b> (H strand)
	I111 (I strand)	

<sup>a</sup>“Spine” residues are highlighted in bold.

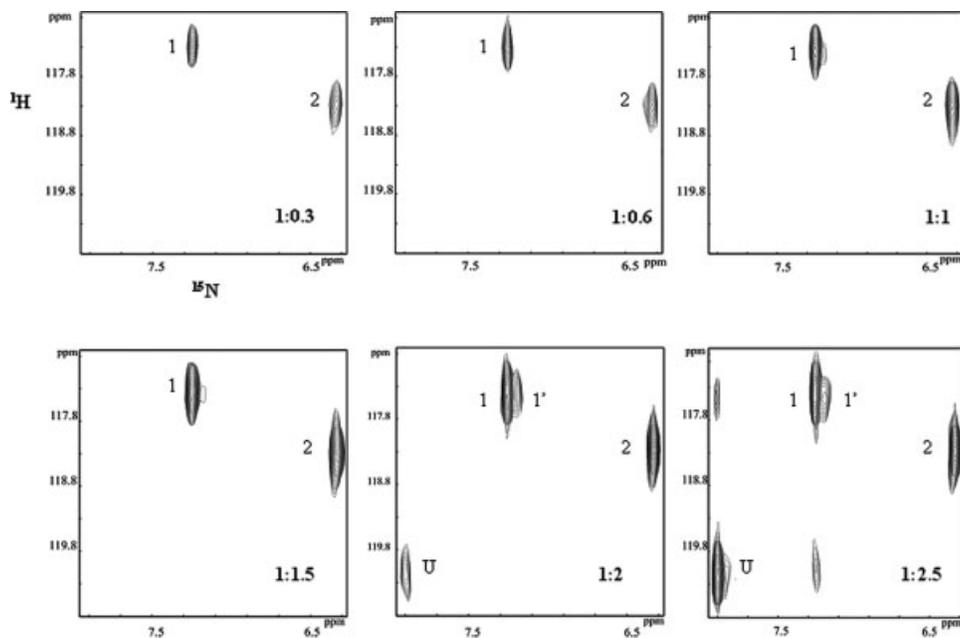
nation is that, as deduced from the models, S51 contacts the ligand via its side-chain, while the measured CSP are relative to backbone amide atoms.

### Bile salts site selectivity

The binding properties of recombinant cL-BABP have been further investigated upon titration of cL-BABP with

the two most abundant physiological glycine conjugates: GCA and GCDA, which differ only for the presence of an hydroxyl group at position 12 of the steroid ring C. These glycine conjugates were synthesized using a <sup>15</sup>N enriched glycine, to follow the titration behaviour from the ligand perspective, monitoring the occupancy of individual binding sites in the protein. A titration experiment of unlabeled cL-BABP with increasing amounts of <sup>15</sup>N-GCDA has been performed investigating nine protein:GCDA ratios (1:0.3, 1:0.6, 1:1, 1:1.5, 1:2, 1:2.5, 1:3, 1:5, 1:7.5). <sup>1</sup>H-<sup>15</sup>N-HSQC spectra of <sup>15</sup>N-GCDA acquired at 298 K at protein:ligand (P:L) ratios in the range 1:0.3 up to 1:2.5 are shown in Figure 5.

The inspection of 2D <sup>1</sup>H-<sup>15</sup>N HSQC spectra revealed the presence of up to four different resonances, depending on the analyzed P:L molar ratio. The peak occurring at 7.9 and 120.2 ppm in the <sup>1</sup>H and <sup>15</sup>N dimensions, respectively, represents unbound <sup>15</sup>N-GCDA, as assigned by control spectra obtained in the absence of the protein. Cross-peaks annotated as 1 (7.3 and 117.4 ppm) and 2 (6.3 and 118.3 ppm) (Fig. 5) represent GCDA bound to two distinct environments of the protein, arbitrarily denoted as “Site 1” and “Site 2”. The cross-peak annotated as 1' (7.2 and 117.4 ppm) only appears at P:L molar ratios equal or higher than 1:2 and it is likely due to the presence of two slightly different populations of

**Figure 5**

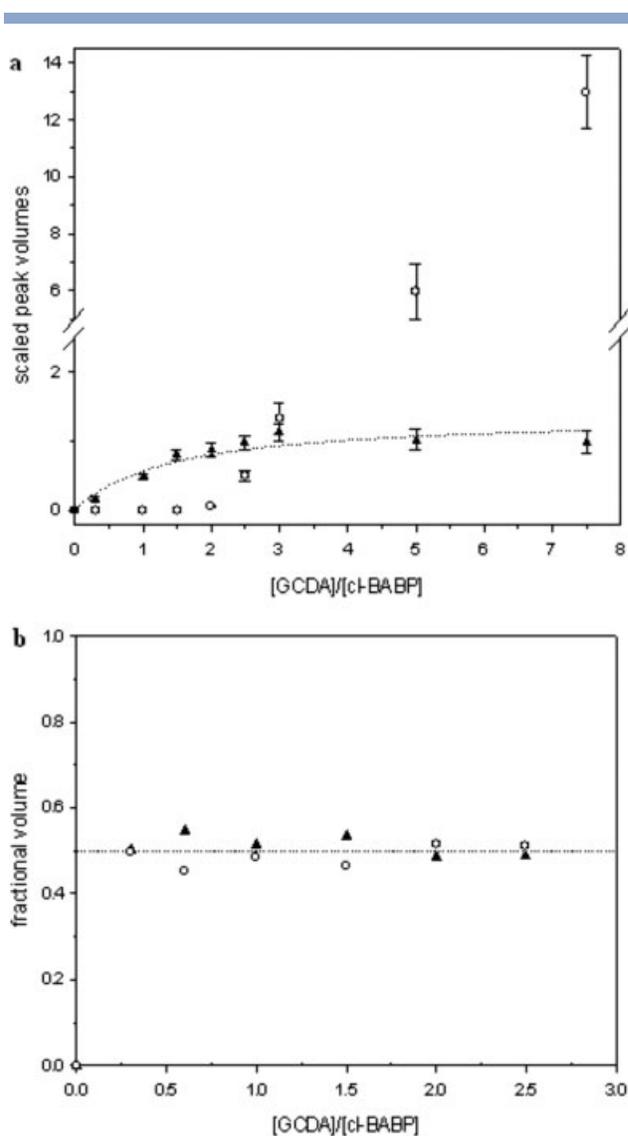
2D <sup>1</sup>H-<sup>15</sup>N-HSQC spectra of <sup>15</sup>N GCDA acquired at 298 K at six different protein:GCDA ratios: 1:0.3, 1:0.6, 1:1, 1:1.5, 1:2, 1:2.5. All the shown spectra were acquired with eight scans except for ratio 1:0.3 where 32 scans were necessary to get a comparable signal to noise ratio. The resonances corresponding to the unbound GCDA and to the bile acid bound to Sites 1, 1', and 2 are designated U, 1, 1', and 2, respectively. The small peak at 7.3 and 120.2 ppm in the spectrum at 1:2.5 protein:GCDA molar ratio results from chemical exchange of ligand from Site 1 and unbound pool.

GCDA at “Site 1”, given the similarity of its chemical shift with that of peak 1. This hypothesis is further confirmed by the analysis of peak volumes as a function of P:L molar ratio. At P:L 1:2 the volume of peak 1 decreases and that of peak 1' increases of the same amount. Figure 5 also shows that at P:L ratios higher than 1:2 an exchange peak between Site 1 and the free bile salt is present. This is a clear indication that Site 1 is more solvent accessible, and therefore hereafter referred to as “superficial site”. The two binding sites can be further distinguished on the basis of the chemical shift analysis of  $H_{\alpha}$  glycine protons ligands in 2D  $^1H$  TOCSY and 3D  $^1H$ - $^{15}N$  HSQC-NOESY:  $^{15}N$  enriched glycine moiety exhibits two separate resonances for the diastereotopic alpha protons only for ligand bound to Site 2 (data not shown), which should consequently be buried in the protein cavity.

A quantitative volume analysis was essayed and the binding curve, obtained by plotting the sum of peak volumes for Site 1, Site 1', and Site 2 versus ligand:protein molar ratio [Fig. 6(a)] indicated a 2:1 stoichiometry. The plot of the fraction of ligand bound to each site determined as  $V(\text{Site } i)/[V(\text{Site } 1) + V(\text{Site } 2) + V(\text{Site } 1')]$ , versus L:P molar ratios is reported in Figure 6(b). This plot shows that the two sites are equally populated at all L:P ratios, indicating a comparable affinity for GCDA and/or a highly cooperative binding mechanism. A commonly used index of cooperativity is the Hill coefficient which ranges, for a two-site system, from 1 (noncooperative systems) to 2 (highly positive cooperative systems). A rough estimate of the Hill coefficient ( $1.8 \pm 0.5$ ) was obtained by fitting the binding curve. We are aware that for systems with very high degree of positive cooperativity an analysis based on NMR data alone is not well amenable to derive reliable Hill coefficients<sup>9</sup>; the calculated coefficient, however, clearly exceeds 1 thus suggesting a high positive cooperativity.

Titration experiments of unlabeled cL-BABP were also performed with  $^{15}N$ -GCA investigating six protein:GCA ratios (1:0.3, 1:0.6, 1:1, 1:1.5, 1:2, 1:3.) at 298 K (Fig. 7). The resonance at 7.3 and 117.5 ppm, annotated as Site 1, is not present at low ligand concentrations and the severe line broadening makes this resonance unobservable also at 1:2 and 1:3 P:L ratios. The free resonance (7.9 and 120.3 ppm) is very broad as well (30 Hz) suggesting the presence of chemical exchange between ligand bound in Site 1 and free ligand. The resonance at 6.3 and 119.3 ppm, annotated as Site 2, exhibits a shift change during titration up to 1:2 P:L molar ratio [Fig. 8(a,b)] indicating a 1:2 stoichiometry also for this ligand. Its chemical shift, at every titration point, corresponds to the weighted average of the two cL-BABP species, complexed with one or two GCA molecules. Altogether these data suggest that cL-BABP binds GCA in the two sites with different  $K_d$ s.

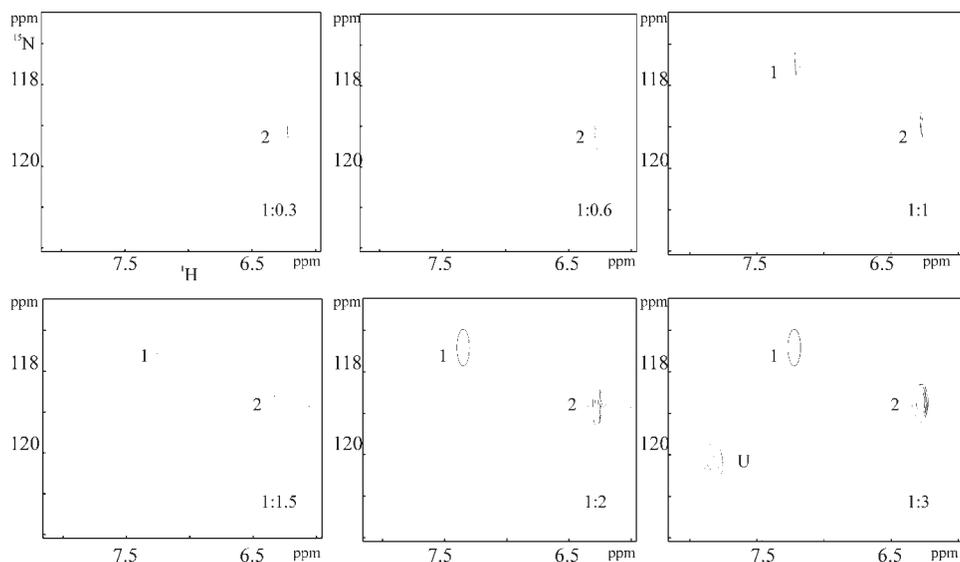
The temperature dependence of GCDA and GCA resonances was further investigated in the range 280–305 K



**Figure 6**

Sum of peak volumes of Site 1, Site 1', Site 2 and exchange peaks from 2D  $^1H^{15}N$ -HSQC spectra of  $^{15}N$  GCDA acquired at 298 K ( $\blacktriangle$ ) and the free peak volume ( $\circ$ ) are plotted versus ligand:protein molar ratio. The error bar is reported (a). The fraction of ligand bound to each site, determined as  $V(\text{Site } i)/[V(\text{Site } 1) + V(\text{Site } 2) + V(\text{Site } 1')]$ , is plotted versus ligand:protein molar ratio: Site 1 ( $\blacktriangle$ ) and Site 2 ( $\circ$ ) (b).

on 1:3 P:L samples [Fig. 9(a,b)]. In both cases a slow exchange regime is observed, allowing an estimate of the upper limit for  $k_{\text{off}}$  of about 2000 and 5000  $s^{-1}$  for ligand at Sites 1 and 2, respectively, based on chemical shift difference between bound and unbound resonances. Site 2 is sharpened upon temperature increase, reflecting the shorter protein correlation time. On the contrary, upon temperature increase, Site 1 and free resonances became selectively broadened and, in the case of GCA, Site 1 disappeared at temperatures higher than 298 K. This behaviour is indicative of a variation from slow to intermediate exchange regime.



**Figure 7**

2D  $^1\text{H}$ - $^{15}\text{N}$ -HSQC spectra of  $^{15}\text{N}$  GCA acquired at 298 K at six different protein:GCA ratios: 1:0.3, 1:0.6, 1:1, 1:1.5, 1:2, 1:3. At molar ratios 1:0.3 and 1:0.6 the spectra were acquired with 32 scans whereas for all the other molar ratios eight scans were enough to obtain a good signal to noise ratio. The resonances corresponding to the unbound GCA and to the bile acid bound to Sites 1 and 2 are designated U, 1, and 2, respectively. Empty circles indicate the position of the missing resonance of Site 1.

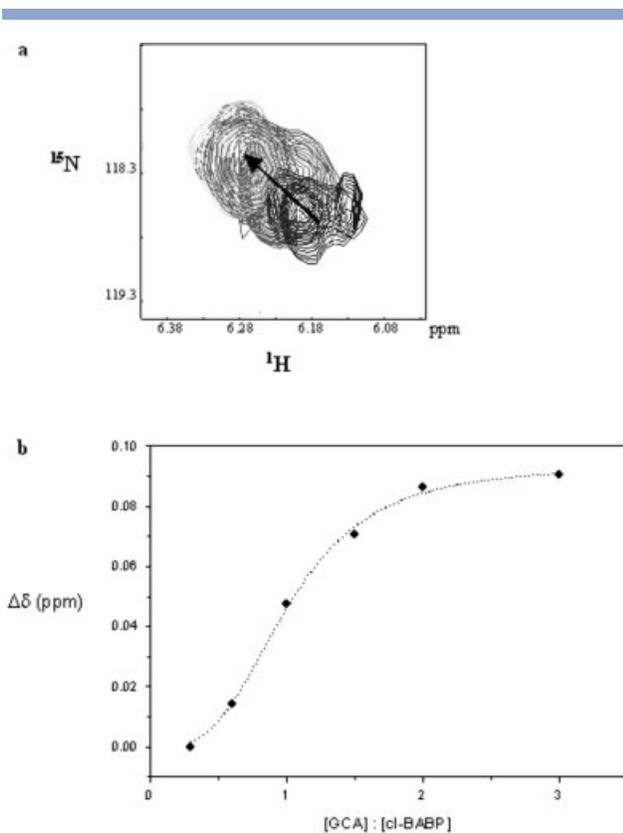
Investigation of site selectivity requires competition experiments and two moles of  $^{13}\text{C}$ -GCA were added to a solution containing 1:2 molar ratio cL-BABP:  $^{15}\text{N}$ -GCDA. The recorded  $^1\text{H}$  $^{15}\text{N}$ -HSQC spectrum showed that the volume of peaks corresponding to Sites 1 and 2 decreased of about 30% indicating that GCA binds to both sites of the protein with a lower affinity with respect to GCDA. This result is confirmed by the acquisition of 2D  $^1\text{H}$ - $^{13}\text{C}$ -HMBC spectra, where the long range correlations between carbonyl carbons and alpha protons ( $^2J_{\text{C-H}}$ ) were used for the observation of  $^{13}\text{C}$ -GCA. The obtained spectrum (Fig. 10) showed the presence of peaks at two carbonyl resonances (176.4, and 182.5 ppm), confirming that GCA binds to both sites. The assignment of “Site 1” and “Site 2” to superficial and internal binding site, respectively, could be obtained (i) on the basis of glycine alpha protons chemical shifts, which are diastereotopic only for Site 2; (ii) from the observed line broadening of the glycine amide resonances corresponding to “Site 1” and to the free ligand upon temperature increase (Fig. 9), indicating the presence of exchange between the ligand free and bound in Site 1.

To further investigate site selective interactions, complementary competition experiments were acquired and two moles of GCDA were added to a solution containing 1:2 molar ratio cL-BABP:  $^{15}\text{N}$ -GCA. The recorded  $^1\text{H}$   $^{15}\text{N}$ -HSQC spectrum showed a 70% decrease of peak volume, indicating that GCDA (i) binds to the protein with higher affinity with respect to GCA and (ii) does not

show site selectivity. The percentages of volume decrease obtained from the two competition experiments are in very good agreement.

#### Bile salts affinity

NMR binding studies allowed a rough estimate of  $K_d$  for GCDA binding to cL-BABP. We are aware that a quantitative measure of  $K_d$  values is possible only within an order of magnitude of the concentration of the studied species<sup>36</sup> and the high affinity of this system precludes a precise estimate by NMR. However an upper limit for  $K_d$  could be deduced from different experiments: the temperature dependence of GCDA and GCA resonances in the holo-protein allowed to estimate  $k_{\text{off}}$  and consequently a value of  $K_D \ll 200 \mu\text{M}$  for Site 1 exhibiting the lowest affinity. Interestingly,  $^1\text{H}$ - $^{15}\text{N}$  HSQC experiments, performed on progressively diluted samples (with a P:L 1:2 ratio), indicated that up to a  $170 \mu\text{M}$  protein concentration, the free ligand resonance was not observable, thus suggesting an upper limit of  $170 \mu\text{M}$  for  $K_d$ . Finally, titration experiments, performed on  $^{15}\text{N}$  labeled samples of cL-BABP allowed to decrease the upper limit values of  $K_d$  to  $20 \mu\text{M}$ . Indeed, assuming a diffusion controlled association constant  $k_{\text{on}}$  of the order of  $10^7 \text{ s}^{-1} \text{ M}^{-1}$ ,  $K_d$  could be derived by measuring the lowest chemical shift differences observed between apo and holo-peaks ( $<30 \text{ Hz}$ ).



**Figure 8**

Superposition of 2D  $^1\text{H}$ - $^{15}\text{N}$ -HSQC spectra of  $^{15}\text{N}$  GCA recorded for cL-BABP- $^{15}\text{N}$  GCA complexes with different P:L ratios in the range 1:0.3–1:3. Only the selected region displaying Site 2 is reported and an arrow shows the shift of resonance 2 upon titration (a). Chemical shift variation  $\Delta\delta = [(\Delta\delta_{\text{HN}})^2 + \Delta\delta_{\text{N}}^2/25]/2^{1/2}$  of peak 2 versus the GCA concentration is plotted (b).

## DISCUSSION

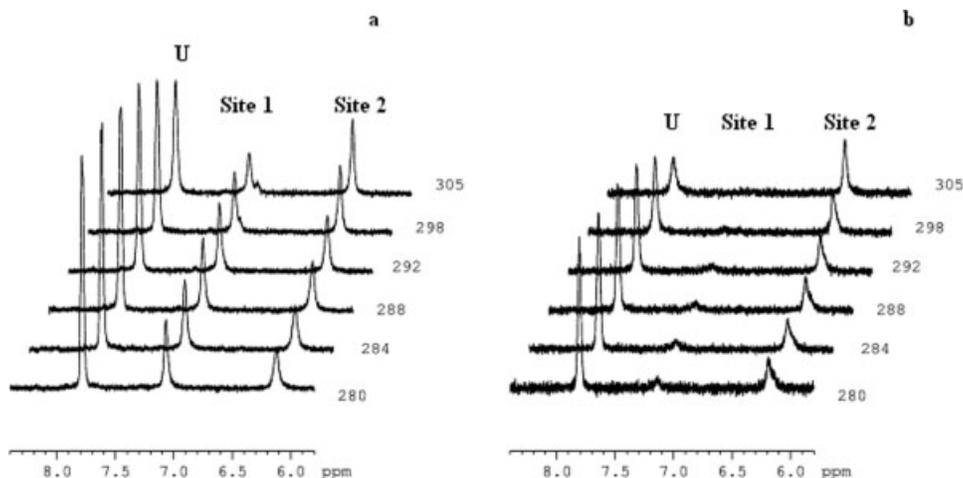
In the present article we address the modeling of a ternary complex between cL-BABP and two CDA molecules. This modeling procedure is highly demanding, given the complexity and the flexibility of the system under study. At the same time it should be noted that the full 3D structure determination based on the assignment of intermolecular NOEs is not trivial, due to the presence of two identical ligand molecules and of resonance overlap in the aliphatic region. We used for this docking a new version of HADDOCK (2.0\_devel), which supports simultaneous multibody ( $N > 2$ ) docking. The following experimental data, derived from different biochemical and biophysical techniques, were used to drive the docking: (i) NMR chemical shift perturbation in the presence of ligand; (ii)  $^{15}\text{N}$ -relaxation data on apo and holo-protein; (iii) ligand/protein NOEs; (iv) NMR STD data; (v) limited proteolysis results.

Docking with HADDOCK resulted in good models that satisfy all the experimental restraints, with the excep-

tion of the AIR involving R32, located in the helix II. It should be noted that no other active residue is close to the side-chain of R32, which protrudes into the solvent. The superposition of the four selected HADDOCK models indicates that CDA located in the more superficial site presents a higher variability ( $\langle \text{RMSD} \rangle$  of 3.0 Å) than the inner molecule ( $\langle \text{RMSD} \rangle$  of 1.9 Å), nicely paralleling the NMR data showing the presence of different populations at superficial site (Sites 1 and 1') in the  $^1\text{H}$ - $^{15}\text{N}$  HSQC spectra of  $^{15}\text{N}$ -GCDA/cL-BABP complex (Fig. 5).

The structural analysis of HADDOCK models indicated a few residues involved in hydrophobic and/or H-bond interactions with CDAs (Table III). Residues interacting with ligands are all located in the C-terminal face of the  $\beta$ -barrel (E, F, G, H, I strands) and in the CD and EF loops, at the open end of the protein cavity (Table III, and Fig. 4). In all the selected models charged side-chains of K76, located in the EF loop, and R55, located in the CD loop, are making salt-bridges with the carboxylate group of the inner and superficial CDA, respectively. The more superficial CDA is also involved, through its carboxyl tail, in an H-bond with the side-chains of T53, located in the loop CD. Similar interactions are also observed in the X-ray structure, highlighting the importance of loops CD and EF for ligand binding. Indeed, as previously discussed,<sup>5</sup> the conformational flexibility, on the microsecond to millisecond time scales, mainly localized at the C-terminal face of the  $\beta$ -barrel and involving the loop EF, is functional to ligand binding. The observed dynamics is primarily caused by the protonation/deprotonation of a buried histidine residue, H98, located on this flexible face and is mediated by a network of polar buried side-chains (T72, C80, S93, H98, E109, R120) defining a spine going from E to J strand. Interestingly most “spine” residues seem to be involved in crucial ligand/protein interactions on the basis of the analysis of HADDOCK models. In particular T72, C80, and H98 make interactions with the ligands in all the structures, while E109 and R120 are picked out in some of them. These results suggest the importance of the spine region for ligand interaction as well as for communication of the motion in the apo protein.

Our models are in agreement with the X-ray structure of cL-BABP in complex with two CA molecules (PDB id: 1tw4). Average RMSD values of 5.8 and 5.2 Å, measured for the ligands in the superficial and in the internal binding site, respectively, are considered as reasonable values since they refer to the differences in ligand coordinates after superposition of protein backbones. In docking studies challenging systems start with backbone RMSDs between free and bound forms of around 2 Å, thus small protein conformational changes can have a large effect on the results. This is the case of cL-BABP showing conformational changes mainly localized at the entrance of the protein cavity. It is important to stress that, even in



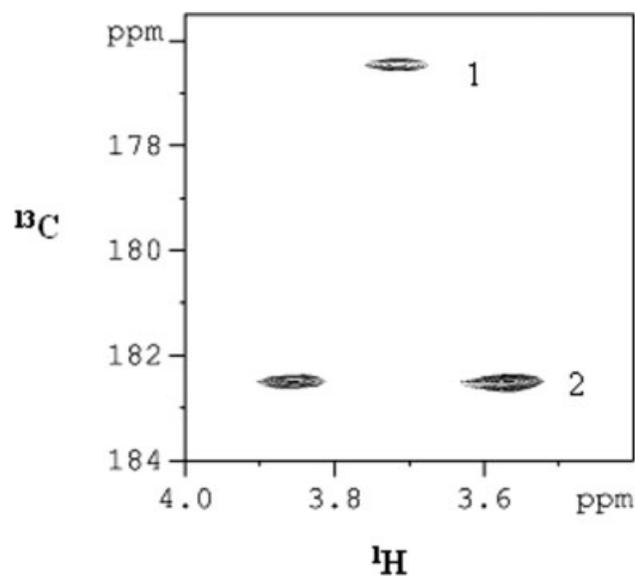
**Figure 9**

Stacked plot reporting the temperature dependence of the bile acid amide  $^1\text{H}$  resonances. One-dimensional first increment of the 2D  $^1\text{H}$ - $^{15}\text{N}$ -HSQC spectra collected on cL-BABP in complex with  $^{15}\text{N}$  GCDA (Panel a) and  $^{15}\text{N}$  GCA (Panel b) using a 1:2.5 P:L molar ratio.

the presence of an RMSD value around 5 Å, similar protein–ligand contacts are made, making the obtained ternary complex a good model for further structural work and for the design of mutants. Finally, it should be also pointed out that the X-ray structure is relative to cL-BABP in complex with two CA molecules, instead of CDAs. CDA and CA differ for the hydroxylation of position 12 on ring C, and the inspection of the X-ray structure reveals the presence of a hydrogen bond between hydroxyl groups in position 12 of the internal ligand with hydroxyl group in position 3 of the other ligand. The lack of this hydrogen bond in the CDA complex could be responsible for a slightly different orientation of the ligands.

NMR binding studies were also performed to address the issue of site selectivity for the binding of GCDA and GCA and its correlation with cooperativity. The evidence of cooperativity upon GCDA binding was indicated by the simultaneous binding of the two GCDA molecules and by the derived Hill coefficient, clearly exceeding one. GCDA/GCA competition experiments revealed a higher affinity of cL-BABP for GCDA, differently from the similar binding affinities measured in human I-BABP<sup>10</sup> for the two bile salts. These results nicely correlate with the natural abundance of bile salts in human and chicken species (vide infra). cL-BABP did not show any site-selectivity for the two bile salts at variance to what observed for hI-BABP, where GCDA binds nearly exclusively to one site, while GCA binds to the other site.<sup>12</sup> The absence of site selectivity, in spite of the highly cooperative system, suggests that site selectivity and cooperativity are not strongly correlated. The rationale of the different site-selectivity behaviour in cL-BABP and hI-BABP can be discussed on the basis of sequence alignment analysis

and of the data reported for hI-BABP mutants. Six single-residue mutations were introduced in the binding pocket of hI-BABP aiming at disrupting the hydrogen-bonding network possibly involved in the energetic communication between the two sites. Only the mutation at residue Q51 resulted in the loss of site preference for



**Figure 10**

2D  $^1\text{H}$ - $^{13}\text{C}$ -HMBC spectrum of cL-BABP in the presence of equimolar amount of  $^{13}\text{C}$ -GCA and unlabeled GCDA in a P:GCDA:GCA molar ratio equal to 1:2:2. The resonances corresponding to  $^{13}\text{C}$ -GCA bound to Sites 1, and 2 are designated as 1 and 2, respectively.



**Figure 11**

Sequence alignment of cL-BABP with ileal BABP sequences. For the ileal proteins only the name of the species is reported. cL-BABP residues involved in ligand binding are highlighted in bold.

both bile salts.<sup>12</sup> Interestingly, as apparent from the alignment of bile acid binding proteins from different organisms (Fig. 11), only rabbit and chicken lack a glutamine residue at this position. These data nicely correlate with the lack of site selectivity we have reported for cL-BABP on the basis of NMR data. It is worth mentioning that all of the HADDOCK solutions show that residue 51 is directly involved in ligand binding.

An analysis of the conservation of residues relevant for bile salt interaction, as highlighted by HADDOCK (in bold in Fig. 11), indicated that T53, R55, D74, T91, H98, I111 are not conserved within ileal-BABPs. The most significant changes, involving a net charge depletion, occur for R55 and D74, located in loops CD and EF, respectively, which are replaced by two glycines in ileal-BABPs. The side-chain of R55 is involved in a salt-bridge with the carboxylate tail of the superficial ligand, in all HADDOCK models. These observations point to a possible different role of the CD loop region in the binding mechanism of liver and ileal BABPs, which are likely to bind bile salts through different interaction networks.

The analysis of the HADDOCK solutions suggests that loop EF residues are important for the binding of both CDA molecules and therefore could be good candidates for the regulation of the system cooperativity.

The data here discussed indicate that docking simulations can be successfully used to obtain meaningful models of a ternary complex and represent a necessary step to help in the structure determination and in functional analysis of complex multisite systems. The obtained model, in conjunction with binding data and competition studies, provides a more complete picture of the binding mechanism in this protein family and to set up the basis for the rational design of mutants with selected binding properties, both in terms of site selectivity and binding cooperativity.

**Biological implications**

It has been reported that human I-BABP displays site-selectivity for GCDA and GCA and similar binding affinities for the two bile salts,<sup>10</sup> while we have shown here

that cL-BABP displays a high affinity for GCDA in both sites. There are ample evidences that the liver of lower vertebrates, such as chicken, frog, turtle, little skate, rainbow trout, has evolved specific transport proteins for mediating bile salt uptake and excretion,<sup>37</sup> although the molecular basis of this transport remains to be determined. Both for mammals and birds the primary site of de novo bile acid synthesis is the liver and approximately 90–95% of bile acids are absorbed from the intestines into portal blood and delivered to the liver, creating enterohepatic circulation. In avian species the principal organ for chemical digestion and absorption is the small intestine; in mammals, the vast majority of bile acids are absorbed in the ileum, whereas 5–15% of bile acids derive from absorption of unconjugated bile acids from the large intestine. Bile acid production is stimulated primarily 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 is important to stress that in humans primary acids include both CA (~55%) and CDA conjugates (~25%),<sup>10</sup> while in avians glycine and taurine CDA conjugates represent 70–95% of bile salt pool. The biliary bile acid profiles of domestic chickens, turkey and ducks have been determined by high performance liquid chromatography and fast atom bombardment mass spectrometry<sup>38</sup> and it has been found that chenodeoxycholytaurine is the major bile acid component. In agreement with these data taurine conjugated CDA was found to be the chief bile acid in both conventional and germ-free domestic fowls: in the latter, hydrolysis and esterification of bile salts produced 82% chenodeoxycholate.<sup>39</sup> The complete biliary bile acid composition of 25 species of fruit pigeons and doves was analyzed: in 23 of 25 species analyzed, CDA and its derivatives were the predominant bile acids present (in the range 69–94%).<sup>40</sup> Moreover, in avian families belonging to some ciconiiformes and herons, the dominant biliary bile acid is the 16 $\alpha$ -hydroxy derivative of CDA.<sup>41</sup>

Difference in the predominant bile acids in mammals and in different bird species deserves particular attention as far as the bile salt homeostasis is concerned.

In the terminal ileum the bile acid transport, which controls the active reabsorption of conjugated bile acids, plays an important role in modulating the bile acid pool size and exerting negative feedback regulation of cholesterol 7 $\alpha$ -hydroxylase, the enzyme that synthesizes bile acids starting from cholesterol. As a consequence, site-selectivity for GCDA and GCA and similar binding affinity for the two bile acids found in the human I-BABP are of considerable importance for the maintenance of the bile acid pool size, since in the mammalian intestine, the two bile acids are present in almost equal amounts. In the light of this discussion it is clear that the asymmetry and site selectivity of human I-BABP ensure that the majority of the protein complexes, *in vivo*, will contain both the

trihydroxy and dihydroxy bile salts.<sup>11</sup> In avians CDA is the predominant primary bile acid, hence there is probably no need for site-selectivity for other bile salts. Moreover, the binding affinity exhibited by cL-BABP for GCDA, higher than for GCA (70/30 as observed by NMR), seems to be sufficient to ensure that the majority of the protein complexes, *in vivo*, will contain mostly GCDA.

*In vitro*<sup>42,43</sup> 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 in vertebrates the addition of a third hydroxyl group has evolved, through six types of nuclear hydroxylation pathways, as a means of detoxifying CDA.<sup>40</sup> The toxicity of high concentration of CDA suggests a further role of cL-BABP in the regulation of bile salts concentration in the cell. These results enlarge the body of knowledge of biological function if related to the differences in physiological pathways and bile salt pools in human and avian species.

## ACKNOWLEDGMENTS

We thank Marianna Luppi, Serena Zanzoni, Simone Bonfanti, Massimo Pedò, and Mara Guariento for cL-BABP expression and purification. We would like to thank Dr. Roberto K. Salinas for the precious help in the early stages of the docking approach and for the interesting discussions.

## REFERENCES

1. Makishima M, Okamoto AY, Repa JJ, Tu H, Learned RM, Luk A, Hull MV, Lustig KD, Mangelsdorf DJ, Shan B. Identification of a nuclear receptor for bile acids. *Science* 1999;284:1362–1365.
2. Hofmann AF. The continuing importance of bile acids in liver and intestinal disease. *Arch Intern Med* 1999;159:2647–2658.
3. Nakahara M, Furuya N, Takagaki K, Sugaya T, Hirota K, Fukamizu A, Kanda T, Fujii H, Sato R. 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* 2005;280:42283–42289.
4. Sessler RJ, Noy N. A ligand-activated nuclear localization signal in cellular retinoic acid binding protein-II. *Mol Cell* 2005;18:343–353.
5. 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.
6. Nichesola D, Perduca M, Capaldi S, Carrizo ME, Righetti PG, Monaco HL. Crystal structure of chicken liver basic fatty acid-binding protein complexed with cholic acid. *Biochemistry* 2004;43:14072–14079.
7. Kanda T, Foucand L, Nakamura Y, Niot I, Besnard P, Fujita M, Sakai Y, Hatakeyama K, Ono T, Fujii H. Regulation of expression of human intestinal bile acid-binding protein in Caco-2 cells. *Biochem J* 1998;330(Part 1):261–265.
8. Martin GG, Atshaves BP, McIntosh AL, Mackie JT, Kier AB, Schroeder F. Liver fatty acid binding protein gene ablation potentiates hepatic cholesterol accumulation in cholesterol-fed female

- mice. *Am J Physiol Gastrointest Liver Physiol* 2006;290:G36–G48.
9. Tochtrop GP, Richter C, Tang C, Toner JT, Covey DF, Cistola DP. Energetics by NMR: site-specific binding in a positively cooperative system. *Proc Natl Acad Sci USA* 2002;99:1847–1852.
  10. Tochtrop GP, Bruns JL, Tang C, Covey DF, Cistola DP. Steroid ring hydroxylation patterns govern cooperativity in human bile acid binding protein. *Biochemistry* 2003;42:11561–11567.
  11. Tochtrop GP, DeKoster GT, Covey DF, Cistola DP. A single hydroxyl group governs ligand site selectivity in human ileal bile acid binding protein. *J Am Chem Soc* 2004;126:11024–11029.
  12. Toke O, Monsey JD, DeKoster GT, Tochtrop GP, Tang C, Cistola DP. Determinants of cooperativity and site selectivity in human ileal bile acid binding protein. *Biochemistry* 2006;45:727–737.
  13. Dominguez C, Boelens R, Bonvin AM. HADDOCK: a protein–protein docking approach based on biochemical or biophysical information. *J Am Chem Soc* 2003;125:1731–1737.
  14. Marley J, Lu M, Bracken C. A method for efficient isotopic labeling of recombinant proteins. *J Biomol NMR* 2001;20:71–75.
  15. Barlos K, Gatos D, Kaposos S, Poulos C, Schafer W, Yao WQ. Application of 2-chlorotrityl resin in solid phase synthesis of (Leu15)-gastrin I and unsulfated cholecystokinin octapeptide. Selective O-deprotection of tyrosine. *Int J Pept Protein Res* 1991;38:555–561.
  16. Kaiser E, Colecott RL, Bossinger CD, Cook PI. Color test for detection of free terminal amino groups in the solid-phase synthesis of peptides. *Anal Biochem* 1970;34:595–598.
  17. Bax A, Summers MF. 1H and 13C assignments from sensitivity-enhanced detection of heteronuclear multiple-bond connectivity by 2D multiple quantum NMR. *J Am Chem Soc* 1986;108:2093–2094.
  18. Marion D, Driscoll PC, Kay LE, Wingfield PT, Bax A, Gronenborn AM, Clore GM. 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 1  $\beta$ . *Biochemistry* 1989;28:6150–6156.
  19. Meyer B, Peters T. NMR spectroscopy techniques for screening and identifying ligand binding to protein receptors. *Angew Chem Int Ed Engl* 2003;42:864–890.
  20. Breeze AL. Isotope-filtered NMR methods for the study of biomolecular structure and interactions. *Prog Nucl Magn Reson Spectrosc* 2000;36:323–372.
  21. Delaglio F, Grzesiek S, Vuister GW, Zhu G, Pfeifer J, Bax A. NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J Biomol NMR* 1995;6:277–293.
  22. Johnson BA. Using NMRView to visualize and analyze the NMR spectra of macromolecules. *Methods Mol Biol* 2004;278:313–352.
  23. Wishart DS, Bigam CG, Holm A, Hodges RS, Sykes BD. 1H, 13C and 15N random coil NMR chemical shifts of the common amino acids. I. Investigations of nearest-neighbor effects. *J Biomol NMR* 1995;5:67–81.
  24. Brunger AT, Adams PD, Clore GM, DeLano WL, Gros P, Grosse-Kunstleve RW, Jiang JS, Kuszewski J, Nilges M, Pannu NS, Read RJ, Rice LM, Simonson T, Warren GL. Crystallography & NMR system: a new software suite for macromolecular structure determination. *Acta Crystallogr D Biol Crystallogr* 1998;54(Part 5):905–921.
  25. Vasile F, Ragona L, Catalano M, Zetta L, Perduca M, Monaco H, Molinari H. Solution structure of chicken liver basic fatty acid binding protein. *J Biomol NMR* 2003;25:157–160.
  26. Eufri D, Sironi A. SMILE–shaded molecular imaging on low-cost equipment. *J Mol Graph* 1989;7:158,165–169.
  27. 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.
  28. Jorgensen WL, Tirado-Rives J. The OPLS. Potential functions for proteins. Energy minimizations for crystals of cyclin peptides and crambin. *J Am Chem Soc* 1988;110:1657–1666.
  29. Linge JP, Williams MA, Spronk CA, Bonvin AM, Nilges M. Refinement of protein structures in explicit solvent. *Proteins* 2003;50:496–506.
  30. Schuttelkopf AW, van Aalten DM. PRODRG: a tool for high-throughput crystallography of protein–ligand complexes. *Acta Crystallogr D Biol Crystallogr* 2004;60:1355–1363.
  31. Fernandez-Recio J, Totrov M, Abagyan R. Identification of protein–protein interaction sites from docking energy landscapes. *J Mol Biol* 2004;335:843–865.
  32. Daura X, Gademan K, Jaun B, Seebach D, van Gusteren WF, Mark AE. Peptide folding: when simulation meets experiment. *Angew Chem Int Ed* 1999;38:236–240.
  33. Wallace AC, Laskowski RA, Thornton JM. LIGPLOT: a program to generate schematic diagrams of protein–ligand interactions. *Protein Eng* 1995;8:127–134.
  34. Farrow NA, Muhandiram R, Singer AU, Pascal SM, Kay CM, Gish G, Shoelson SE, Pawson T, Forman-Kay JD, Kay LE. Backbone dynamics of a free and phosphopeptide-complexed Src homology 2 domain studied by 15N NMR relaxation. *Biochemistry* 1994;33:5984–6003.
  35. Brown EM, Pfeffer PE, Kumosinski TF, Greenberg R. Accessibility and mobility of lysine residues in  $\beta$ -lactoglobulin. *Biochemistry* 1988;27:5601–5610.
  36. Zuiderweg ER. Mapping protein–protein interactions in solution by NMR spectroscopy. *Biochemistry* 2002;41:1–7.
  37. Ballatori N, Rebbor JF, Connolly GC, Seward DJ, Lenth BE, Henson JH, Sundaram P, Boyer JL. Bile salt excretion in skate liver is mediated by a functional analog of Bsep/Spgp, the bile salt export pump. *Am J Physiol Gastrointest Liver Physiol* 2000;278:G57–G63.
  38. Elkin RG, Wood KV, Hagey LR. Biliary bile acid profiles of domestic fowl as determined by high performance liquid chromatography and fast atom bombardment mass spectrometry. *Comp Biochem Physiol B* 1990;96:157–161.
  39. Haslewood GA. Bile salts of germ-free domestic fowl and pigs. *Biochem J* 1971;123:15–18.
  40. Hagey LR, Scheingart CD, Ton-Nu HT, Hofmann AF. Biliary bile acids of fruit pigeons and doves (Columbiformes): presence of 1- $\beta$ -hydroxychenodeoxycholic acid and conjugation with glycine as well as taurine. *J Lipid Res* 1994;35:2041–2048.
  41. Hagey LR, Scheingart CD, Ton-Nu HT, Hofmann AF. A novel primary bile acid in the Shoebill stork and herons and its phylogenetic significance. *J Lipid Res* 2002;43:685–690.
  42. Heuman DM, Mills AS, McCall J, Hylemon PB, Pandak WM, Vlahcevic ZR. Conjugates of ursodeoxycholate protect against cholestasis and hepatocellular necrosis caused by more hydrophobic bile salts. In vivo studies in the rat. *Gastroenterology* 1991;100:203–211.
  43. Heuman DM, Pandak WM, Hylemon PB, Vlahcevic ZR. Conjugates of ursodeoxycholate protect against cytotoxicity of more hydrophobic bile salts: in vitro studies in rat hepatocytes and human erythrocytes. *Hepatology* 1991;14:920–926.