Chicken ileal bile acid binding protein: a promising target of investigation to understand binding cooperativity across the protein family
Mara Guariento, Michael Assfalg, Serena Zanzoni, Dimitrios Fessas, Renato Longhi, Henriette Molinari

To cite this version:

HAL Id: hal-00479243
https://hal.archives-ouvertes.fr/hal-00479243
Submitted on 30 Apr 2010

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L’archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d’enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.
Chicken ileal bile acid binding protein: a promising target of investigation to understand binding cooperativity across the protein family

Mara Guariento*, Michael Assfalg*, Serena Zanzoni*, Dimitrios Fessas†, Renato Longhi‡, and Henriette Molinari*

*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; ‡Istituto di Chimica del Riconoscimento Molecolare, Consiglio Nazionale delle Ricerche, Via Mario Bianco 9, 20131 Milano, Italy.

1Corresponding Author:
Henriette Molinari, Dipartimento di Biotecnologie, Università degli Studi di Verona, Strada Le Grazie 15, 37134 Verona, Italy, e-mail: henriette.molinari@univr.it, Phone: +39 045 8027901, Fax: +39 045 8027929

Short title: Binding Mechanism of Chicken Ileal Bile Acid Binding Protein

Keywords: Bile Acid Binding Protein, Nuclear Magnetic Resonance, Multi-site Interaction, Chicken Ileum, Lipid Transport, Cooperativity

Abbreviations used: Abbreviations used: BABP, bile acid-binding protein; FABP, fatty acid-binding protein; L-BABP, liver bile acid binding protein; I-BABP, ileal bile acid-binding protein; BA, bile acid; GCDA, glycochenodeoxycholic acid; GCA, glycocholic acid; CSI, chemical shift index; CSP, chemical shift perturbation; HSQC, heteronuclear single quantum coherence; ITC, isothermal titration calorimetry; PDB, Protein Data Bank; RMSD, root-mean-square deviation.
ABSTRACT

Protein-bile acid interactions are crucial microscopic events at the basis of both physiological and pathological biochemical pathways. Bile acid binding proteins are intracellular transporters able to bind ligands with different stoichiometry, selectivity, and cooperativity. The molecular determinants and energetics of interaction are the observables that connect the microscopic to the macroscopic frameworks. This paper addresses the study and proposes a mechanism for the multi-site interaction of bile acids with chicken ileal bile acid binding protein (I-BABP) with the aim of elucidating the determinants of ligand binding in comparison to 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 cooperativity. 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 described results point to the presence of a protein scaffold which is able to establish long-range communication networks, but does not manifest positive binding cooperativity, 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.
INTRODUCTION

Bile acids (BAs) 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 bile acid synthesis which is critical for the maintenance of whole body cholesterol homeostasis [2]. The coordinate 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 BA binding proteins (BABPs) involved in lipid trafficking belong to the fatty acid binding protein (FABP) 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 ileal BABP (I-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 liver BABP (L-BABP) [9] showing a protein:ligand stoichiometry of 1:2. The chicken L-BABP/cholate is the only NMR structure so far reported of a ternary BABP complex. Very recent studies addressing the binding stoichiometry of rabbit and zebrafish I-BABPs with different bile acids, using calorimetry combined with mass spectrometry [10] and X-ray crystallography [11], respectively, suggested that these proteins may have the capability of binding two ligands inside their cavity and additional one or two ligands on the molecular surface. As for the binding mechanism, the most comprehensive work was reported on human I-BABP [12, 13]. The latter protein binds two molecules of glycocholic acid with low intrinsic affinity but an extraordinarily high degree of positive cooperativity. 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 cooperativity has not been elucidated yet [8].

A recent study has shown that chicken L-BABP, similarly to human I-BABP, displays very high binding cooperativity towards bile salts [14], and it was suggested that “a population shift model” [15] was responsible for the observed cooperativity. 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 may derive from the comparative analysis of similar proteins from different tissues and organisms.

Here we address the characterisation of the binding interactions of chicken I-BABP with bile salts using isothermal titration calorimetry (ITC) 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] has been mutagenized using the Quickchange Site-Directed Mutagenesis kit (Stratagene). Forward and a reverse primers (MWG Biotech) have been 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 cI-BABP coding sequence). The shorter construct was then coding for chicken I-BABP fused with the same tag that would have been found after the proteolytic cut of the longer product. The short construct has been used to transform E. coli XL1Blue competent cells (Stratagene). After a miniprep extraction (Sigma-Aldrich), the mutagenized plasmid sequence has been verified by DNA sequencing. The recombinant vector has been then used to transform E. coli SG expression strain (Qiagen) and transformed cells were selected on LB agar plates containing 100 μg/ml of ampicillin (Sigma-Aldrich).

Protein expression and purification
Chicken I-BABP has been expressed as soluble protein inoculating 1L of LB medium with a 10 ml starter culture and incubating at 37°C until cells reached an OD₆₀₀ of 0.6. Protein expression was induced by addition of isopropylthiogalactopyranoside (IPTG) 0.5 mM and incubation overnight at 20°C. Due to the presence of the C-terminal tag plus the four residues 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 1 reports the constructs deriving 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. ¹⁵N and ¹³C isotope labelling was achieved using M9 minimal media containing 1 g/L NH₄Cl and 4 g/L ¹³C-enriched glucose and incubating the cultures overnight at 28°C. The final yield of purified proteins was 30 mg/L of culture for both ¹⁵N and ¹⁵N-¹³C chicken I-BABP. The purity of the samples has been verified by SDS-PAGE and mass spectroscopy analysis has been used to quantify the isotopic enrichment (close to 100% in both ¹⁵N and ¹⁵N-¹³C 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, 0.05% NaN₃ and stored as lyophilized protein.

Determination of the protein aggregation state
The state of aggregation of chicken I-BABP has been 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 glycochenodeoxycholic acid (GCDA) at a protein-ligand ratio of 1:6). Measurements of the centrifuged NMR samples were carried out at 298K using a DYNA-PRO 801 dynamic light-scattering/molecular-sizing instrument (Protein Solutions). Data were processed with Dynamics software.

Isothermal Titration Calorimetry (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) 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 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 are described in the literature[19-22]. Briefly, the observable enthalpy is given by:

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

and the degree of association, i.e the concentration ratio $\tilde{x} = [\text{bound ligand}] / [\text{total protein}]$ is given by:

$$\tilde{x} = R T \left[ \frac{\partial \ln Q}{\partial \mu_L} \right]_{T, p} = \left[ \frac{\partial \ln Q}{\partial \ln [L]} \right]_{T, p}$$

where $R$ is the universal gas constant, $\mu_L$ 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 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, $P_0$:

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

$Q$ 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 $\Delta H_b$. 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 this 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 lyophilized protein in 90%H$_2$O/10%D$_2$O obtaining a final concentration of 0.4 mM. Unenriched GCDA and glycocholic acid (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 correspondent bile salts by addition of 2.15 µl of 1.1 M KOH for each mg of bile acid. Both unenriched and $^{15}$N-enriched bile salts have been then used to prepare stock solutions in a buffer 20 mM potassium phosphate pH 6.5, 135 mM KCl, 10 mM NaCl, 0.05% NaN$_3$. Titration experiments using 15-labelled ligand consisted in fourteen titration points measured at ligand:protein ratios $r = 0.3, 0.6, 0.7, 0.9, 1.1, 1.5, 1.8, 2.2, 2.7, 3.6, 5.0, 6.5, 8.5, 10$.

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 5mm 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 $^1$H dimension), followed by zero filling and Fourier transform.

Sequential assignment and assignment of C$_\alpha$, C$_\beta$, H$_\alpha$, H$_\beta$ and carbonyl carbons resonances have been obtained performing HNCACB, CBCA(CO)NH, HNCO, HBHANH and HBHA(CO)NH experiments on $^{15}$N-$^{13}$C chicken I-BABP. Each 2D $^1$H-$^{15}$N HSQC 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 obtained with TOPSPIN 2.1 (Bruker) while assignment 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 few months ten peaks (A1, F2, T3, Q42, N43, G44, I65, G88, K89 and I90) were shifted to a slightly different position, whereas three other peaks (G-2, S-1 and M0) were missing. These residues are all localized near the N-terminal tag MRGSM added during the cloning: G-2, S-1 and M0 belong to the tag, A1, F2 and T3 were directly linked to it, Q42, N43, G44, I65, G88, K89 and I90 were all positioned in loops surrounding the N-terminus. These data suggest that chicken I-BABP has the tendency to lose the five N-terminal residues, an hypothesis confirmed by mass spectrometry analysis. Possibly this tag has been removed by a contaminant protease or by an enzyme such as methionine aminopeptidase (MAP1), which is responsible of the cleavage of the first methionine in *E. coli* proteins [26]. However this modification did not affect the binding of bile acids to chicken I-BABP, as verified by isothermal titration calorimetry 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 result session, using the software Rosetta [27], as implemented on the e-grid.
RESULTS

Protein NMR resonances assignment and backbone structure analysis

A series of heteronuclear 3D 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 glycocholic acid (GCA) or with glycochenodeoxycholic acid (GCDA). A complete assignment has been 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 E72-A81, 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 chemical shift index (CSI) [28] analysis is consistent with the expected secondary structure (Figure S2). The experimental chemical shifts of $^{13}$C$_\alpha$, $^{13}$C$_\beta$, $^{13}$C', $^{15}$N, H$_\alpha$ and H$_N$ of 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 is based, as reported, on how well the coordinates of the lowest energy structures agree with one another [29]. A plot of ROSETTA all-atom energy as a function of the C$_\alpha$ RMSD relative to the model with the lowest energy shows a “funnelling” distribution (Figure S3), identified by Baker [29] as a convergence criterion. The 10 lowest energy structures are within 2Å from the best model (Figure S4).

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 (Figure 1a,c), and a superimposition of the spectra of the two holo forms (Figure 1e). The residue-specific chemical shift perturbations (CSP) derived from the comparison of apo and holo chicken I-BABPs, in complex with GCDA and GCA are reported in Figure 1b,d, respectively, while Figure 1f 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 here reported refers to the cI-BABP/GCDA complex.

Isothermal titration calorimetry experiments

Figure 2 reports the experimental ITC data of GCDA binding to chicken I-BABP in a $\Delta H$-r plot ($\Delta H =$ cumulative enthalpy per mole 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 Materials and Methods) 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 observe 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, while the simplest model proposed well represents 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 provides 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 coordinate. A titration experiment has been performed on \(^{15}\)N 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 additional twenty-six residues only the peaks relative to the holo form could be analyzed, 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 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 \( r = 0, 1, 2, 4 \) were also analysed in the assumption that carbon chemical shifts may select more localised conformational changes and offer the possibility to localize the binding sites. Selected regions of \(^1\)H-\(^{13}\)C CT-HSQC spectra registered at varying molar ratios are reported 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 varying protein:ligand ratios was performed on sixty 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 4, thus indicating that they were not affected by binding. These residues are mainly located at the bottom of the protein cavity, and in regions of strands I-J, adjacent to loop II; ii) a second group of residues, evenly distributed on all the strands of the proteins, reached their final chemical shift at \( r = 1 \); iii) few residues, mainly located in helix I, still exhibited a minor chemical shift change on going from \( r = 1 \) to 4.

**Binding site occupancies and site selectivity**

An NMR titration has been 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 reported in Figure 4. It appears that a single resonance (\( \delta H = 7.84 \) ppm, \( \delta N = 119.29 \) ppm) is observable up to a \( r = 1.8 \), whereas a second resonance appears at \( r = 2.2 \) (\( \delta H = 7.21 \) ppm, \( \delta N = 112.06 \) ppm), 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 unbound GCDA (U-GCDA), determined separately in a control experiment. Interestingly, the peak resonating at lower field ($\delta H = 7.72$ ppm, $\delta N = 119.17$ ppm) changes position and becomes sharper (Figure 4g) during the titration, reaching, at high $r$ values, chemical shift ($\delta H = 7.88$ ppm, $\delta N = 119.48$ ppm) 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 bile acid 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 ppm 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 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 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 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 analyzed, 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 about 100 µM). In Figure 7 the averaged normalized intensity of 22 analyzed 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 to almost 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 analyzed set of resonances corresponds to the only unbound protein without mixing of the singly bound form. A similar analysis was performed also 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 choose 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 the Supplementary Data (Figure S5).

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 to $^{15}$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 confirmed also by an even distribution in all 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).

**Comparative analysis of binding properties in ileal BABPs**

The binding mechanisms, stoichiometries, and affinities of few ileal BABP/bile acid 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 between the two ligands and protein residues pointing towards the internal cavity (Figure S6), 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 cooperativity can be described by the free energy of interaction between binding sites, \( \Delta \Delta G \), which provides a more convenient measure of cooperativity 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 analyzed protein-ligand adducts, it emerges that the human protein displays the largest cooperativity, the rabbit and zebrafish BABPs have intermediate cooperativity, 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 ileal 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 9b together with the calculated identities among the proteins (Figure 9a). Non-conservative aminoacid substitutions within the binding cavity, among the four investigated proteins, are highlighted on the ribbon structure of the chicken protein in Figure 9c. These substitutions evidence subtle changes in both the cavity dimensions and in potential hydrophobic and 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 cooperativity. In the sequence alignment reported in Figure 9b, underlined aminoacids 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 (3EM0). It is worth noting that T38, W49, Q51, N61, Q99 and E110, making contacts with the ligands in the human protein, are conserved in all the sequences, with the only exceptions of positions 51 and 99 exhibiting H to Q mutations without any direct correlation with cooperativity. Indeed both human (highly cooperative) and chicken (poor cooperativity) display a glutamine at position 51, while human and zebrafish (intermediate cooperativity) display a glutamine at position 99. The residues making contacts with the ligands in the zebrafish structure are Y14, I21, K30, Y53, V74, L90, Y97, Q99, T110 and R125. All the aminoacids occupying these positions are either conserved or exhibit conservative mutations in the analysed sub-family, with the only exception of K30 and V74, again without any clear correlation to the presence of cooperativity.

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 cooperativity [12]. However both W49 and N61 are conserved in the chicken ileal protein. It was further suggested that these residues are part of a H-bonding network in the protein-bile salt complex that connects the two sites, giving rise to two cooperativity networks in the doubly ligated system: an upper network involving the steroid ring hydroxyl group at position C-12 and a lower network involving the steroid ring hydroxyl group at positions C-3 and C-7 [12]. The structural data do not allow to
confirm this hypothesis and indeed only a careful structural correlation of the H-bonding network in the four proteins 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 cooperativity.

The comparison of the enthalpic contributions to binding of the characterized ileal BABPs from human, rabbit, zebrafish and chicken, revealed that the here described protein exhibited a much larger enthalpic change of \(-73.6\, \text{KJ mol}^{-1}\) on binding (Table 1). Interestingly a large enthalpy change of \(-109.5\, \text{KJ mol}^{-1}\) was recently measured for a mutant of rabbit I-BABP (Δα-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 here measured enthalpy change 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 (E72-A81) 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 promotes a major structural rearrangement, in agreement with the independent conclusion derived from the analysis of protein \(^{15}\text{N}-\text{H}\) and \(^{13}\text{C}-\text{H}\) chemical shift perturbations.

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 determined by localized enthalpic effects and could be removed by mutation of a glutamine 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 characterisation 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 cooperativity was established by NMR investigations, although no site-selectivity was observed \([16, 33]\).

**Conclusions**

In summary, the here described analysis 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 cooperativity, as observed for the human protein. These considerations make chicken I-BABP a potentially invaluable model system for the understanding of cooperativity among BABPs and may provide clues about the development of this efficient feature during species evolution.
ACKNOWLEDGEMENTS

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

FUNDING

This research was supported by FIRB 2003 (Project No. RBNE03FX83) and Fondazione Cariverona (Bando di ricerca scientifica 2007: Nuove applicazioni della Risonanza Magnetica Nucleare allo sviluppo di nano biotecnologie).
REFERENCES


FIGURE LEGENDS

Figure 1 Chemical shift changes of chicken I-BABP backbone $^{15}$N-H resonances on ligand binding
Left 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 (in blue); c) apo $^{15}$N-I-BABP (in red) and $^{15}$N-I-BABP/GCA 1:5 (in green); e) $^{15}$N-I-BABP/GCDA 1:5 (in blue) and $^{15}$N-I-BABP/GCA 1:5 (in green). Right panels: plots of the chemical shift differences relative to b) unbound protein and bound to GCDA; d) unbound protein and bound to GCA; f) protein bound to GCDA and bound to GCA. Residues exhibiting a chemical shift perturbation above the displayed threshold line are G22, E29, M30, V37, T50, G55, R57, F63, E68, T113, G116, V118, and T122 for both complexes; F8, D11, I28, G31, C34, F53, T58, A69, V83, and S114 for the complex with GCDA; I21, E39, and G56 for the complex with GCA. The missing bars refer to residues for which the assignment of the unbound form is missing, or to prolines.

Figure 2 Isothermal titration calorimetry 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 $\Delta H$-vs-$r$ plot, where $\Delta H$ is the cumulative enthalpy (sum of the peak areas in the insert) expressed per mole 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 (dotted line); two consecutive sites (continuous line).

Figure 3 Protein $^{13}$C-H chemical shift perturbations on ligand binding
Selected regions of $^1$H-$^{13}$C CT-HSQC spectra registered at 600 MHz, 25°C, at increasing ligand:protein ratios, $r$. Upper panels: residues showing two distinct resonances in slow exchange, from left to right: $r=0$, $r=1$, $r=2$, $r=4$. 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.

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.

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 depicted to the right of the corresponding spectra.

Figure 6. Population fractions of unbound and bound protein species
The population fractions $\chi$ are calculated according to the binding model parameters presented 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
Continuous line: population fraction of unbound protein (P); dashed line: sum of population fractions of unbound and singly bound protein (PL); circles: experimental average intensity decay
of one set of $^{15}$N-H protein signals (see text). The error bar represents the variability of the 22 analysed resonances.

**Figure 8 Evidence of ligand-induced global structural rearrangement**
The protein structure, displayed in cartoon representation, corresponds to the coordinates 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.

**Figure 9 Sequence alignment of four ileal proteins of the BABP family**
Sequence alignment of chicken ileal 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 an empty dot. Residues underlined in the sequences are those reported to be involved in binding in hI-BABP, on the basis of the first generation NMR structures, and zI-BABP, as obtained from the X-ray structure (3EM0).
### Table 1. Thermodynamic parameters obtained from ITC measurements considering two or three stepwise binding models for the interaction of bile acids and BABPs from different species\(^a\). All measurements were performed at 25°C.

<table>
<thead>
<tr>
<th></th>
<th>(K_1) ((\times10^4)M(^{-1}))</th>
<th>(\Delta H_1) (kJmol(^{-1}))</th>
<th>(T\Delta S_1) (kJmol(^{-1}))</th>
<th>(K_2) ((\times10^4)M(^{-1}))</th>
<th>(\Delta H_2) (kJmol(^{-1}))</th>
<th>(T\Delta S_2) (kJmol(^{-1}))</th>
<th>(K_3) ((\times10^4)M(^{-1}))</th>
<th>(\Delta H_3) (kJmol(^{-1}))</th>
<th>(T\Delta S_3) (kJmol(^{-1}))</th>
<th>(\Delta \Delta G^e) (kJmol(^{-1}))</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chicken</td>
<td>1.55±0.15</td>
<td>-50.3±5.0</td>
<td>-26.4±2.7</td>
<td>0.55±0.05</td>
<td>-23.3±2.3</td>
<td>-1.95±0.2</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-0.87</td>
</tr>
<tr>
<td>Human(^b)</td>
<td>0.13±0.03</td>
<td>10.0±6.2</td>
<td>6.7±1.5</td>
<td>16.9±3.2</td>
<td>-11.1±0.1</td>
<td>-4.0±0.1</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-15.50</td>
</tr>
<tr>
<td>Rabbit(^c)</td>
<td>0.47±0.4</td>
<td>-0.2±0.3</td>
<td>-</td>
<td>0.6±0.1</td>
<td>-9.0±2.1</td>
<td>-</td>
<td>2.5±0.2</td>
<td>18.7±2.0</td>
<td>-</td>
<td>-6.96</td>
</tr>
<tr>
<td>Zebrafish(^d)</td>
<td>0.63±0.1</td>
<td>-5.1±0.5</td>
<td>-</td>
<td>1.02±0.1</td>
<td>-1.6±0.2</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-4.55</td>
</tr>
</tbody>
</table>

\(^a\) The reported data refer to the following ligands: GCDA (chicken and rabbit), GCA (human), CA (zebrafish), which were shown to display the highest affinities for the corresponding BABP.

\(^b\) See reference [12]

\(^c\) 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].

\(^d\) The ITC data were also consistent with a model of three independent binding sites [11].

\(^e\) 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].
Figure 1
Figure 2
Figure 3
Figure 4
Figure 5
Figure 7
Figure 8
<table>
<thead>
<tr>
<th></th>
<th>cl-BABP</th>
<th>hI-BABP</th>
<th>rI-BABP</th>
<th>zI-BABP</th>
</tr>
</thead>
<tbody>
<tr>
<td>cl-BABP</td>
<td>100</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>hI-BABP</td>
<td>62</td>
<td>100</td>
<td></td>
<td></td>
</tr>
<tr>
<td>rI-BABP</td>
<td>59</td>
<td>82</td>
<td>100</td>
<td></td>
</tr>
<tr>
<td>zI-BABP</td>
<td>55</td>
<td>54</td>
<td>50</td>
<td>100</td>
</tr>
</tbody>
</table>

Figure 9