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

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

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

INTRODUCTION

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

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

A previous study has shown that chicken L-BABP, similarly to human I-BABP, displays very high binding co-operativity towards bile salts [14], and it was suggested that ‘a population-shift model’ [15] was responsible for the observed co-operativity. Subsequent studies on selected mutants of both human I-BABP [12] and chicken L-BABP [16] have started to identify protein residues that are functional to the binding mechanism. A further approach towards the description of the interaction mechanism assessing the absence of strong co-operativity. Both the measured energetics of binding and the distributed protein chemical-shift perturbations are in agreement with a first binding event triggering a global structural rearrangement. The enthalpic and entropic contributions associated with binding of the first ligand indicate that the interaction increases stability and order of the bound protein. The results described in the present study point to the presence of a protein scaffold which is able to establish long-range communication networks, but does not manifest positive-binding co-operativity, as observed for the human protein. We consider chicken I-BABP a suitable model to address the molecular basis for a gain-of-function on going from non-mammalian to mammalian species.

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

Protein–bile acid interactions are crucial microscopic events at the basis of both physiological and pathological biochemical pathways. BABPs (bile-acid-binding proteins) are intracellular transporters able to bind ligands with different stoichiometry, selectivity and co-operativity. The molecular determinants and energetics of interaction are the observables that connect the microscopic to the macroscopic frameworks. The present paper addresses the study and proposes a mechanism for the multi-site interaction of bile acids with chicken I-BABP (ileal BABP) with the aim of elucidating the determinants of ligand binding in comparison with homologous proteins from different species and tissues. A thermodynamic binding model describing two independent consecutive binding sites is derived from isothermal titration calorimetry experiments and validated on the basis of both protein-observed and ligand-observed NMR titration data. It emerges that a singly bound protein is relatively abundant at low ligand/protein molar ratios

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

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

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

Protein expression and purification
Chicken I-BABP has been expressed as a soluble protein inoculating 1 litre of LB medium with 10 ml of starter culture and incubating at 37°C until cells reached a *D*~600~ of 0.6. Protein expression was induced by the addition of 0.5 mM IPTG (isopropyl β-D-thiogalactoside) and incubation overnight at 20°C. Owing to the presence of the C-terminal tag plus the four residue N-terminal tag, derived from the BamHI cleavage site used for the cloning step, the predicted isoelectric point of chicken I-BABP shifted from the predicted isoelectric point of chicken I-BABP shifted from the BamHI cleavage site used for the cloning step, the predicted isoelectric point of chicken I-BABP shifted from the original value of 6.8 to 8.8. Supplementary Figure S1 (at http://www.BiochemJ.org/bj/425/bj4250413add.htm) shows the constructs derived from different cloning strategies. Given the increased value of pi, the same purification protocol previously described [14] was used, based on ionic exchange instead of affinity chromatography, thus increasing the yield of soluble protein from 40 to 80 mg/l of culture in rich medium (LB). This result was possibly due to the elimination of the thrombin cleavage step, one of the causes of protein precipitation during the purification process. 15N and 13C isotope labelling was achieved using M9 minimal media containing 1 g/l NH4Cl and 4 g/l 1C-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 13N-13C chicken I-BABP. The purity of the samples was verified by SDS/PAGE, and MS analysis was used to quantify the isotopic enrichment (close to 100 % in both 15N and 13N-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 (dithiothreitol) and 0.05 % sodium azide, and stored as freeze-dried protein.

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

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

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

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

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

\[
\bar{x} = \frac{R \ln Q}{\tilde{Q} \cdot \frac{\partial \ln Q}{\partial \ln [L]} \frac{1}{T} \frac{\partial \ln Q}{\partial \ln [L]} \frac{1}{T} \mu_L }
\]

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 as the free protein state. Since we can approximate these systems as diluted solutions, the thermodynamic activities of the solutes may be replaced with their molar concentrations. Under this assumption the partition function is the sum of the concentrations of all protein species, \( P_i \), referred to the free protein, \( P_0 \) (eqn 3):

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

\( Q \) 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, \( K_a \), and the binding enthalpy \( \Delta H_a \). The binding constant is a dimensionless quantity by definition. However, in order to stress the approximation of the thermodynamic activities with the molar concentrations, the use of M⁻¹ units for this parameter is widely used and was adopted in the present paper.
Data fitting to the binding models was performed using the nonlinear Levenberg–Marquardt method [23]. The errors of each fitting parameter were calculated with a 95.4% confidence limit by the Monte Carlo simulation method. An error of 5% in the protein concentration was also taken into account.

NMR spectroscopy

NMR samples of chicken I-BABP were prepared by resuspending the freeze-dried protein in 90% H2O/10% D2O obtaining a final concentration of 0.4 mM. Unenriched GCDA and GCA were purchased as sodium salts from Sigma–Aldrich. 15N-Enriched GCDA and GCA were prepared as previously described [4] and transformed into the corresponding bile salts by addition of 2.15 µl of 1.1 M KOH for each milligram of bile acid. Both unenriched and 15N-enriched bile salts were then used to prepare stock solutions in a buffer of 20 mM potassium phosphate, pH 6.5, 135 mM KCl, 10 mM NaCl and 0.05% sodium azide. Titration experiments using 15N-labelled ligand consisted of fourteen titration points measured at ligand/protein ratios of r = 0.3, 0.6, 0.7, 0.9, 1.1, 1.5, 1.8, 2.2, 2.7, 3.6, 5.0, 6.5, 8.5 and 10.

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

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

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

Combined chemical-shift changes were calculated as:

$$\Delta\delta_{HN} = \left(\Delta\delta_{1H}^2 + \Delta\delta_{15N}^2/25\right)^{1/2}$$

Considerations on protein stability

The superimposition of two 1H-15N HSQC spectra of the same sample of chicken I-BABP–GCDA complex (1:5 ratio) recorded at different time points indicated that after a few months ten peaks (Ala1, Phe2, Thr1, Gln42, Asn43, Gly44, Ile66, Gly88, Lys89 and Ile89) were shifted to a slightly different position, whereas three other peaks (Glu72, Ser92 and Met9) were missing. These residues are all localized near the N-terminal tag MRGSM added during the cloning: Gly2, Ser1 and Met0 belong to the tag; Ala1, Phe2 and Thr1 were directly linked to the tag; Gln42, Asn43, Gly44, Ile66, Gly88, Lys89 and Ile89 were all positioned in loops surrounding the N-terminus. These results suggest that chicken I-BABP has the tendency to lose the five N-terminal residues, an hypothesis confirmed by MS analysis. Possibly this tag has been removed by a contaminant protease or by an enzyme such as MAP1 (methionine aminopeptidase 1), which is responsible of the cleavage of the first methionine residue in E. coli proteins [26]. However, this modification did not affect the binding of bile acids to chicken I-BABP, as verified by ITC performed on a sample before and after separation of the two forms of protein.

Rosetta protein structure calculation

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

RESULTS

Protein NMR resonance assignment and backbone structure analysis

A series of heteronuclear three-dimensional NMR experiments were used for the assignment of 1H, 13C and 15N backbone resonances of chicken I-BABP in the unbound form as well as in complex either with GCA or with GCDA. A complete assignment was obtained for the GCDA and GCA complexes, while an assignment level of 83% was derived for the apo protein. For this protein form, spectral correlations were missing for the stretch Glu2–Ala11, as a result of fast exchange of amide protons with the solvent and/or conformational exchange. A similar behaviour was reported for apo chicken L-BABP [14]. The apparent flexibility of this specific region of the apo protein (E, F strand and EF loop) is consistent with the presence of an ensemble of native states. Conformational exchange of the apo form is quenched upon ligand binding, as observed for chicken L-BABP [9,14].

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

CSP (chemical-shift perturbation) mapping upon ligand binding

The 1H-15N correlation NMR spectrum of a protein and its specific pattern can be regarded as a fingerprint of its structure. Monitoring the perturbations of 15N and Hα 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 1H-15N HSQC spectrum of apo chicken I-BABP with the spectra of the protein in complex with GCDA and GCA (Figures 1a and 1c), and a superimposition of the spectra...
Figure 1 Chemical-shift changes of chicken I-BABP backbone $^{15}$N-H resonances on ligand binding

Left-hand panels: superimpositions of $^{1}H-^{15}N$ HSQC spectra of (a) apo $^{[15]N}$-I-BABP (in red) and $^{[15]N}$-I-BABP–GCDA, 1:5 ratio (in blue); (c) apo $^{[15]N}$-I-BABP (in red) and $^{[15]N}$-I-BABP–GCA, 1:5 ratio (in green); and (e) $^{[15]N}$-I-BABP–GCDA, 1:5 ratio (in blue) and $^{[15]N}$-I-BABP–GCA, 1:5 ratio (in green). Right-hand panels: plots of the chemical-shift differences relative to (b) unbound protein and protein bound to GCDA; (d) unbound protein and protein bound to GCA; and (f) protein bound to GCDA and bound to GCA. Residues exhibiting a chemical shift perturbation above the displayed threshold line are Gly23, Glu29, Met30, Val37, Thr50, Gly55, Arg57, Phe63, Glu68, Thr113, Gly116, Val118 and Thr122 for both complexes; Phe8, Asp11, Ile28, Gly31, Cys34, Phe53, Thr58, Ala69, Val83 and Ser114 for the complex with GCDA; and Ile21, Glu39 and Gly56 for the complex with GCA. The missing bars refer to residues for which the assignment of the unbound form is missing, or to proline residues.

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

ITC experiments

Figure 2 reports the experimental ITC data of GCDA binding to chicken I-BABP in a $\Delta H$ against $r$ plot ($\Delta H =$ cumulative...
enthalpy per mol of protein; \( r = \) concentration ratio \( = \) total titrated ligand/total protein). Thermodynamic binding models of increasing complexity, in terms of the number of fitting parameters and of protein states, were tested to interpret the calorimetry measurements, assuming that the interaction is best described by the simplest model in agreement with the experimental data. The best fit trials using the simplest model, i.e. one-step single-site binding, did not yield satisfactory results. An excellent fit (Figure 2) was instead achieved using the next thermodynamic model (according to the minimum complexity criterion) that implies the presence of two consecutive binding sites. The relevant partition function (see the Experimental section) in this case implies the presence of two consecutive binding sites. The relevant partition function (see the Experimental section) in this case:

\[
Q = 1 + K_{b1}[L] + K_{b2}[L]^2
\]

and implies three protein states: free, one binding site occupied and two sites occupied. The values of the corresponding best fit thermodynamic parameters are reported in Table 1. We observed that in both binding events the interaction is enthalpically driven. Models with additional complexity (for example the two independent binding site model, described by \( Q = 1 + K_{b1}[L] + K_{b2}[L]^2 + K_{b3}[L]^3 \), and implying four protein states) may also be considered appropriate. However, in the absence of additional experimental details their application is not justified, whereas the simplest model proposed represents well the average behaviour of all possible microstates.

**NMR \(^{15}\text{N}\) and \(^{13}\text{C}\) titration results**

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

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

<table>
<thead>
<tr>
<th>Ligand</th>
<th>$K_1$ (M$^{-1}$)</th>
<th>$ΔH_1$ (kJ mol$^{-1}$)</th>
<th>$ΔS_1$ (kJ mol$^{-1}$ K$^{-1}$)</th>
<th>$K_2$ (M$^{-1}$)</th>
<th>$ΔH_2$ (kJ mol$^{-1}$)</th>
<th>$ΔS_2$ (kJ mol$^{-1}$ K$^{-1}$)</th>
<th>$ΔG$ (kJ mol$^{-1}$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chicken GCDA</td>
<td>1.55 ± 0.15</td>
<td>-50.3 ± 5.0</td>
<td>223 ± 23</td>
<td>2.3 ± 0.2</td>
<td>-40 ± 0.1</td>
<td>2 ± 0.2</td>
<td>-12.7 ± 2.0</td>
</tr>
<tr>
<td>Human GCA</td>
<td>0.15 ± 0.03</td>
<td>5.0 ± 0.2</td>
<td>69.3 ± 3.2</td>
<td>0.12 ± 0.1</td>
<td>10.2 ± 0.1</td>
<td>2 ± 0.2</td>
<td>-11.0 ± 2.0</td>
</tr>
<tr>
<td>Zebrafish GCA</td>
<td>0.03 ± 0.01</td>
<td>214 ± 27</td>
<td>67 ± 15</td>
<td>0.03 ± 0.01</td>
<td>124 ± 2.2</td>
<td>2 ± 0.2</td>
<td>-14.2 ± 1.5</td>
</tr>
</tbody>
</table>

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 $ΔG = -RT ln(\frac{4K_2}{K_1})$.

The ITC data were also consistent with a model of three independent binding sites, as expected from the reported large enthalpic contributions.

The binding model of 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 GCA $^{15}$N-H signals. The pertinence of the proposed

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

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

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

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

In order to investigate the determinants of interaction between binding sites, detailed structural data would be helpful. However, up to now, the only available high-resolution structure of an I-BABP with two bile salts bound inside the cavity is that of the protein from zebrafish \cite{11}. We therefore focussed our attention on primary sequence data and structural models. A sequence alignment of the mentioned proteins is reported in Figure 9(B) together with the calculated identities among the proteins (Figure 9A). Non-conservative amino acid substitutions within the binding cavity, among the four investigated proteins, are highlighted on the ribbon structure of the chicken protein in Figure 9(C). These substitutions indicate subtle changes in both the cavity dimensions and in potential hydrophobic and
The population fractions $\chi$ are calculated according to the binding model parameters shown in Table 1. Fractions of the unbound, singly bound and double-bound protein forms correspond to curves a, b and c respectively.

Hydrophilic interactions with any ligand molecule bound inside the cavity. It is, however, not straightforward to correlate the occurring substitutions with the trends in binding co-operativity. In the sequence alignment reported in Figure 9(B), underlined amino acids mark residues establishing contacts with the ligands, as derived from the first generation NMR structure (never released) of the human protein and from the X-ray structure of the zebrafish protein (PDB code: 3EM0). It is worth noting that Thr$^{38}$, Trp$^{49}$, Gln$^{51}$, Asn$^{61}$, Gln$^{99}$ and Glu$^{110}$, which make contacts with the ligands in the human protein, are conserved in all the sequencies, with the only exceptions of positions 51 and 99 exhibiting histidine to glutamine mutations without any direct correlation with co-operativity. Indeed both human (highly co-operative) and chicken (poor co-operativity) display a glutamine at position 51, whereas human and zebrafish (intermediate co-operativity) display a glutamine at position 99. The residues making contacts with the ligands in the zebrafish structure are Tyr$^{14}$, Ile$^{21}$, Lys$^{80}$, Tyr$^{53}$, Val$^{74}$, Leu$^{90}$, Tyr$^{97}$, Gln$^{99}$, Thr$^{100}$ and Arg$^{125}$. All of the amino acids occupying these positions are either conserved or exhibit conservative mutations in the analysed sub-family, with the only exception being Lys$^{80}$ and Val$^{74}$, again without any clear correlation to the presence of co-operativity.

Mutations performed on the residues making contact with the ligands in the human protein indicated that W49A and N61A mutants result in loss of positive binding co-operativity [12]. However, both Trp$^{49}$ and Asn$^{61}$ are conserved in the chicken ileal protein. It was further suggested that these residues are part of a hydrogen-bonding network in the protein–bile salt complex that connects the two sites, giving rise to two co-operativity networks in the doubly ligated system: an upper network involving the steroid ring hydroxy group at position C-12, and a lower network involving the steroid ring hydroxy group at positions C-3 and C-7 [12]. The structural data do not allow us to confirm this hypothesis and indeed only a careful structural correlation of the hydrogen-bonding network in the four proteins making contacts with the ligands in the zebrafish structure are
Figure 8  Evidence of ligand-induced global structural rearrangement

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

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

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

The comparison of the enthalpic contributions with binding of the characterized I-BABPs from human, rabbit, zebrafish and chicken, revealed that the protein described in the present study exhibited a much larger enthalpic change of $-73.6 \text{ kJ} \cdot \text{mol}^{-1}$ on binding (Table 1). Interestingly a large enthalpy change of $-109.5 \text{ kJ} \cdot \text{mol}^{-1}$ 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 enthalpy change measured here may be ascribed to the presence of local folding accompanying the binding event. This conclusion is even more strongly supported by the largely negative entropic contribution (Table 1), indicating that the system is shifted to a more highly ordered state by ligand binding. It is then likely that missing NMR connectivities of residues in the EF strands and EF loop of the apo protein (Glu$_{72}$–Ala$_{11}$) 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 of the family, accompanied by further thermodynamic analysis and molecular dynamic simulations can lead to the design of appropriate mutations capable of providing a mechanistic insight into the determinants of co-operativity.
promotes a major structural rearrangement, in agreement with the independent conclusion derived from the analysis of protein \(^{15}\)N- and \(^{13}\)C-H CSPs.

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

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

**Conclusions**

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

**AUTHOR CONTRIBUTION**

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

**ACKNOWLEDGEMENTS**

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

**FUNDING**

This work was supported by the FIRB (Fondo per gli Investimenti della Ricerca di Base 2003 [project number RBNE3SPX03]; and the Fondazione Cariverona [Bando di ricerca scientifica 2007: Nuove applicazioni della Risonanza Magnetica nucleare allo sviluppo di nano biotecnologie].

**REFERENCES**


© The Authors Journal compilation © 2010 Biochemical Society
the identification of protein secondary structure using 13C chemical-shift data. J. Biomol.
NMR 4, 171–180
Singarapu, K. K., Lemak, A. et al. (2008) Consistent blind protein structure
4685–4690
a natively unfolded helixless variant of rabbit ILBP. Biochemistry 48, 7556–7564
31 Kramer, W., Sauber, K., Baringhaus, K. H., Kurz, M., Stengelin, S., Lange, G., Corsiero, D.,
the ileal lipid-binding protein by photoaffinity labeling, matrix-assisted laser desorption
ionization-mass spectrometry, and NMR structure. J. Biol. Chem. 276, 7291–7301
of calbindin D9k: a Monte Carlo simulation study. Biochemistry 30, 5209–5217
33 Tomaselli, S., Ragona, L., Zetta, L., Assafal, M., Ferranti, P., Longhi, R., Bonvin, A. M. and
Molinari, H. (2007) NMR-based modeling and binding studies of a ternary complex
between chicken liver bile acid binding protein and bile acids. Proteins 69, 177–191
SUPPLEMENTARY ONLINE DATA

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

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

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

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

From the top: the original sequence of chicken I-BABP (from Ala1 to Ala127) is represented by a black bar; the long construct (as reported in [1]) with the short N-terminal tag preceding the first methionine residue and the long C-terminal tag containing the thrombin-cleavage site and an His6-tag; the short construct (the present work) with the short N-terminal tag and the C-terminal tag cut after the first four residues.

1 To whom correspondence should be addressed (email henriette.molinari@univr.it).
Figure S2  Histogram representation of the chemical-shift index (CSI)

CSI obtained for $\alpha$-protons (HA), $\alpha$-carbons (CA), $\beta$-carbons (CB) and carbonyl carbons (CO) of: (a) chicken I-BABP in complex with GCDA, 1:5 ratio, and (b) chicken I-BABP in complex with GCA, 1:5 ratio. (c) Consensus secondary structure assignments obtained combining the CSIs for the complexes of chicken I-BABP with GCDA and GCA. The residues exhibiting a difference between the two complexes are indicated.
Figure S3  Plot of rescored Rosetta all-atom energy against Ca RMSD relative to the lowest-energy model (dot on vertical axis)

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

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

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

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

The lower field shifted ligand H_n signal labelled 1/U (Figure 4 in the main text) can be attributed to both unbound and bound ligand. The chemical-shift changes of the signal during the titration with chicken I-BABP can be measured with high precision and may be used to understand the nature of the observed signal. It is assumed that the ligand species are in a fast chemical-exchange regime and that the observed shift of the signal is a linear combination of the shifts of the unbound and some bound ligand. The latter could correspond to a molecule in a singly bound or a double-bound protein, or both. Ligand H_n chemical shifts for peak 1/U were simulated from linear combinations of the ITC-derived population fractions considering values of 7.62 and 7.92 p.p.m. for the bound and unbound forms respectively. It was also considered that the singly bound protein could exist in a variety of forms, not all of them contributing to the observed signal. In particular, after several simulation trials we considered that only 45% of the total single-bound protein could contribute to signal 1/U. The ligand and the remaining 55% singly bound protein may be bound less specifically or its signal be broadened by dynamic equilibria. The best agreement between ITC and NMR data was obtained considering that 1/U is produced by both unbound ligand and ligand present in approximately half of the singly ligated protein forms. The Figure shows the simulated shifts corresponding to a linear combination of the population fractions of unbound and: singly bound protein (PL), dashed line; sum of singly and double-bound protein (PL+PLL), dashed/dotted line; and NMR observable singly bound protein (45% of PL), continuous line. •, Experimental proton-shift changes of the H_n ligand signal 1/U.
Figure S6  Selected regions of two-dimensional $^1$H-$^1$H-NOESY (nuclear Overhauser enhancement spectroscopy) acquired at 600 MHz, 25°C, on a sample containing ci-BABP–GCDA at a 1:4 ratio in 30 mM Pi, pH 6.5

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

REFERENCES


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