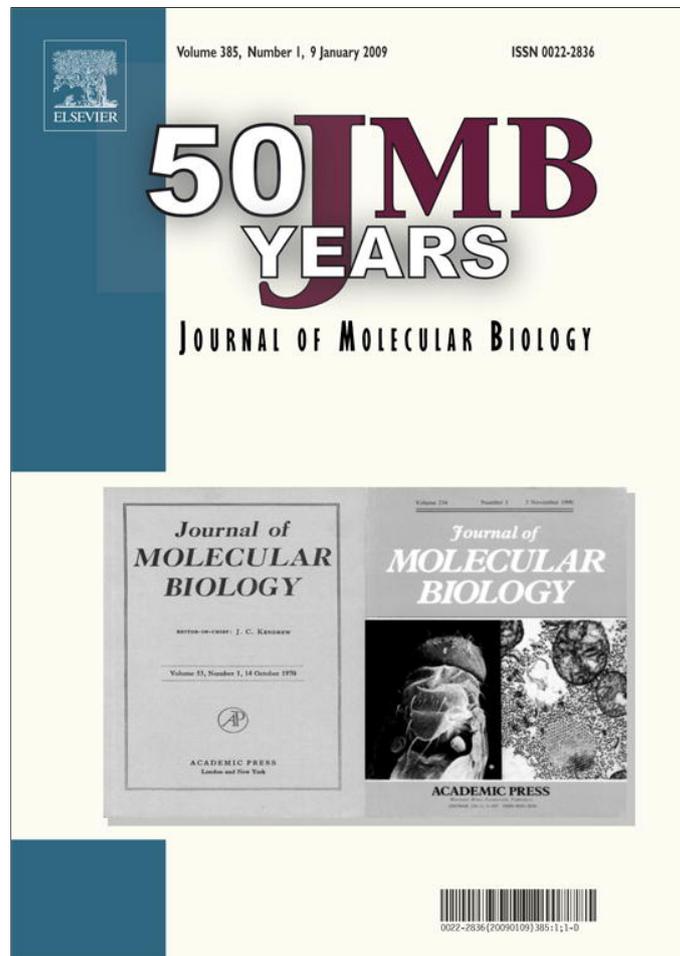


Provided for non-commercial research and education use.  
Not for reproduction, distribution or commercial use.



This article appeared in a journal published by Elsevier. The attached copy is furnished to the author for internal non-commercial research and education use, including for instruction at the authors institution and sharing with colleagues.

Other uses, including reproduction and distribution, or selling or licensing copies, or posting to personal, institutional or third party websites are prohibited.

In most cases authors are permitted to post their version of the article (e.g. in Word or Tex form) to their personal website or institutional repository. Authors requiring further information regarding Elsevier's archiving and manuscript policies are encouraged to visit:

<http://www.elsevier.com/copyright>

**JMB**Available online at [www.sciencedirect.com](http://www.sciencedirect.com) ScienceDirect

# The X-Ray Structure of Zebrafish (*Danio rerio*) Ileal Bile Acid-Binding Protein Reveals the Presence of Binding Sites on the Surface of the Protein Molecule

Stefano Capaldi<sup>1</sup>, Gianmaria Saccomani<sup>1</sup>, Dimitrios Fessas<sup>2</sup>,  
Marco Signorelli<sup>2</sup>, Massimiliano Perduca<sup>1</sup> and Hugo L. Monaco<sup>1\*</sup>

<sup>1</sup>*Biocrystallography Laboratory,  
Department of Biotechnology,  
University of Verona, Ca Vignal  
1, Strada Le Grazie 15,  
37134 Verona, Italy*

<sup>2</sup>*Department of Food Science,  
Technology and Microbiology  
(DISTAM), University of  
Milan, via Celoria 2,  
20133 Milan, Italy*

Received 24 June 2008;  
received in revised form  
24 September 2008;  
accepted 1 October 2008  
Available online  
14 October 2008

The ileal bile acid-binding proteins (I-BABPs), also called ileal lipid-binding proteins or gastrotropins, belong to the family of the fatty acid-binding proteins and play an important role in the solubilization and transport of bile acids in the enterocyte. This article describes the expression, purification, crystallization, and three-dimensional structure determination of zebrafish (*Danio rerio*) I-BABP both in its apo form and bound to cholic acid. This is the first X-ray structure of an I-BABP. The structure of the apoprotein was determined to a resolution of 1.6 Å, and two different monoclinic crystal forms of the holoprotein were solved and refined to 2.2 Å resolution. Three protein molecules are present in the asymmetric unit of one of the co-crystal forms and two in the other, and therefore, the results of this study refer to observations made on five different protein molecules in the crystalline state. In every case, two cholate ligands were found bound in approximately the same position in the internal cavity of the protein molecules, but an unexpected result is the presence of clear and unambiguous electron density for several cholate molecules bound on hydrophobic patches on the surface of all the five independent protein molecules examined. Isothermal titration calorimetry was used for the thermodynamic characterization of the binding mechanism and has yielded results that are consistent with the X-ray data. Ligand binding is described in detail, and the conformational changes undergone by the protein molecule in the apo-to-holo transition are examined by superposition of the apo- and holoprotein models. The structure of the holoprotein is also compared with that of the liver BABP from the same species and those of other I-BABPs determined by NMR.

© 2008 Elsevier Ltd. All rights reserved.

**Keywords:** ileal bile acid-binding protein (I-BABP); zebrafish (*Danio rerio*); cholic acid; X-ray structure; isothermal titration calorimetry (ITC)

Edited by R. Huber

## Introduction

Bile acids are synthesized from cholesterol in the liver, secreted into the bile, and then, to a large extent, reabsorbed in the intestine to return via the portal blood to the liver. Bile acid homeostasis and transport have generated a great deal of interest,

among other reasons, because their synthesis accounts for the catabolism of about 50% of the body cholesterol.<sup>1,2</sup> In the ileum, bile acids are actively absorbed in the apical membrane of the enterocyte, and once in the cell, they bind to the ileal bile acid-binding protein (I-BABP), an abundant 14-kDa cytosolic protein. The reabsorbed bile acids are known to activate the nuclear farnesoid X receptor in the enterocyte, which, in turn, stimulates the expression of I-BABP. Recent evidence indicates that there is an interaction that appears to be regulated by bile acids between the two proteins.<sup>3</sup> The I-BABPs, also called ileal lipid-binding proteins or gastrotropins, belong to the conserved multigene family of the fatty acid-binding proteins (FABPs).<sup>4–9</sup> The members of this family were originally named according to the tissue

\*Corresponding author. E-mail address:  
[monaco@sci.univr.it](mailto:monaco@sci.univr.it).

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; ITC, isothermal titration calorimetry; PDB, Protein Data Bank.

from which they were first isolated, but later on, when it became clear that different FABP types can be present in the same tissue, an alternative nomenclature was proposed.<sup>10</sup> All the members of the family share the same fold, a 10-stranded  $\beta$ -barrel in which two short helices are inserted between the first and the second strand of antiparallel  $\beta$ -sheet.

The first I-BABP to be structurally characterized was porcine I-BABP, at the time called gastrotrypin.<sup>11,12</sup> Other structural studies followed, which made use of NMR spectroscopy to examine the interaction of bile acids with the carrier protein from different species.<sup>13–15</sup> In an article focusing on human I-BABP, the stoichiometry of this interaction was subsequently established in two bile acid ligands per protein molecule.<sup>16</sup> Cooperativity and selectivity of ligand binding were also studied using NMR spectroscopy,<sup>17–19</sup> and the kinetic mechanism was examined using stopped-flow fluorescence analysis.<sup>20</sup> Recombinant rabbit I-BABP was photo-labeled with derivatives of cholytaurine and the amino acids located at the attachment sites identified.<sup>21</sup> In addition to the modifications expected for ligand binding in the internal cavity, this study identified reacting amino acids located on the surface of the protein molecule, suggesting the presence of secondary binding sites for bile acids. More recently, an engineered helixless variant of the same rabbit I-BABP was examined using isothermal titration calorimetry (ITC), and the data were interpreted with a model in which the stoichiometry of ligand binding was three cholates per protein molecule.<sup>22,23</sup>

Within the FABP family, the protein subgroup that appears to be closest to the I-BABPs, structurally and possibly functionally, is that of the liver BABPs (L-BABPs). This group, originally called the liver "basic" FABPs, is present in the liver of several vertebrates: birds, reptiles, amphibians, and fish but not in mammals. It is suspected, although not proved, that the function that L-BABPs perform may be carried out in mammals by the other paralogous protein in the liver, L-FABP. The structures of the L-BABPs from several species are known in detail. The first L-BABP whose three-dimensional structure was determined by X-ray diffraction was that of chicken L-BABP<sup>24,25</sup> followed by toad<sup>26</sup> and axolotl L-BABP.<sup>27</sup> NMR and other biophysical techniques have also been used to study chicken L-BABP, which remains the best characterized member of the L-BABP family.<sup>28,29</sup> We have recently reported the X-ray structure of zebrafish L-BABP complexed with cholate.<sup>30</sup> Whereas in all the other species studied so far it was found that the internal ligand-binding site accommodated two cholate molecules, in the zebrafish wild-type protein, only one ligand molecule was found. This unusual stoichiometry of ligand binding could, however, be changed to the usual one of two ligands per binding site by mutating a single amino acid, Cys91, to Thr.

In this article, we report the expression, purification, crystallization, and three-dimensional structure determination of zebrafish I-BABP both in its apo form and complexed with cholic acid in two

different crystal forms. In addition to the two ligands in the expected internal binding sites, in all the five molecules present in the asymmetric units of the two co-crystals, cholate molecules were found to bind on the surface of the protein. ITC was used for the thermodynamic characterization of the binding mechanism and has yielded results that are consistent with the X-ray data.

## Results and Discussion

### Structure of zebrafish apo I-BABP

The final model of the apoprotein corresponds to the full-length 130-amino-acid chain plus two amino acids in the N terminus deriving from the cloning procedure and two amino acids in the C terminus, part of the thrombin cleavage site. The model contains 1030 non-hydrogen protein atoms, one ethylene glycol, and 133 water molecules. The conventional  $R$ -factor is 20.2% and  $R_{\text{free}}$  is 24.0% (Table 1). The  $R$ -factors and r.m.s.d. values of Table 1 were calculated with the program REFMAC.<sup>31</sup> The stereochemical quality of the protein model was assessed with the program PROCHECK,<sup>32</sup> 95.6% of the residues are in the most favorable region of the Ramachandran plot and the remaining 4.4% are in the additionally allowed region. The overall fold consists of the canonical  $\beta$ -barrel with 10 strands of antiparallel  $\beta$ -chain and the two  $\alpha$ -helices inserted in between the first and the second strand. The secondary-structure assignments for the  $\beta$ -strands are the following: strand A, residues 4–12; B, residues 37–43; C, residues 46–53; D, residues 57–64; E, residues 69–72; F, residues 78–86; G, residues 89–93; H, residues 98–104; I, residues 107–115; and J, residues 120–129. The two  $\alpha$ -helices span residues 14–20 and 25–30.

The resolution of the trigonal crystal form of the apoprotein is the highest of the three forms and is quite adequate for the analysis of the structure of the solvent molecules within the ligand-binding cavity of the protein (see Table 1). Therefore, the solvent structure was initially studied in this crystal form. The single molecule present in the asymmetric unit contains five very well defined solvent molecules in the cavity that were later on found in four out of the five protein molecules present in the two forms of the co-crystals. In the fifth protein molecule, there was more uncertainty because of the lower quality of the map in that region (see below). In addition, an ethylene glycol molecule, a compound present in the cryoprotectant, binds to solvent molecules in the center of the cavity, in the region which in the holoforms is occupied by one of the cholate molecules (number 150 in our notation). These as well as the solvent molecules bound to the cholates of the holoforms are discussed below.

### Structure of zebrafish holo I-BABP

The two different crystal forms of the holoprotein were grown under different conditions with a ratio

**Table 1.** Data collection and refinement statistics

	I-BABP apo	I-BABP + cholate	I-BABP + cholate
Space group	<i>P</i> 3 <sub>1</sub> 21	C2 (form A)	C2 (form B)
<i>a</i> (Å)	54.42	149.19	96.29
<i>b</i> (Å)	54.42	80.29	85.29
<i>c</i> (Å)	82.70	39.72	75.06
$\alpha$ (°)	90.0	90.0	90.0
$\beta$ (°)	90.0	96.12	127.65
$\gamma$ (°)	120.0	90.0	90.0
Protein molecules in asymmetric unit	1	3	2
Resolution range (Å)	41.0–1.6	40.0–2.2	30.0–2.2
Observed reflections	98,855	65,571	73,395
Independent reflections	18,986	22,845	24,125
Multiplicity	5.2 (5.3)	2.9 (2.5)	3.1 (3.1)
$R_{\text{merge}}$ (%) <sup>a</sup>	6.9 (28.1)	6.0 (34.7)	5.9 (31.2)
$I/\sigma$	16.9 (6.4)	16.3 (2.3)	9.1 (1.3)
Completeness (%)	99.1 (100.0)	96.4 (84.4)	98.6 (97.3)
Wilson <i>B</i> -factor (Å <sup>2</sup> )	15.8	37.7	43.6
Reflections in refinement	18,001	21,673	22,788
$R_{\text{cryst}}$ (%) <sup>b</sup>	20.2 (21.7)	22.2 (27.0)	22.6 (27.2)
$R_{\text{free}}$ (%) (test set 5%) <sup>c</sup>	24.0 (22.8)	26.4 (33.6)	25.4 (28.8)
Protein atoms	1030	3088	2054
Ligand atoms	4 (ethylene glycol)	377	261
Water molecules, total	133	69	70
Water molecules in the cavity	15	23 (10+12+1)	16 (6+10)
r.m.s.d. on bond lengths (Å) <sup>d</sup>	0.008	0.010	0.012
r.m.s.d. on bond angles (°)	1.155	1.462	1.536
Planar groups (Å)	0.004	0.003	0.004
Chiral volume deviation (Å <sup>3</sup> )	0.076	0.091	0.096
Average <i>B</i> -factor (Å <sup>2</sup> )			
Protein atoms	15.6	A: 35.7, B: 35.6, C: 36.9	A: 39.1, B: 39.1
Ligand atoms	25.5	40.1	59.4
Solvent atoms	28.8	28.0	39.4

The values in parentheses refer to the highest-resolution shells. For the data collection of apo I-BABP, the highest resolution interval is 1.69–1.60 Å, whereas for the two co-crystals of I-BABP complexed with cholate, it is 2.32–2.20 Å.

The highest-resolution shells used in the refinements are 1.64–1.60 Å for the apoprotein and 2.26–2.20 Å for the complexes with cholate.

The ligand of the apo form is an ethylene glycol molecule.

<sup>a</sup>  $R_{\text{merge}} = \frac{\sum_h \sum_i |I_{ih} - \langle I_h \rangle|}{\sum_h \sum_i \langle I_h \rangle}$ , where  $\langle I_h \rangle$  is the mean intensity of the *i* observations of reflection *h*.

<sup>b</sup>  $R_{\text{cryst}} = \frac{\sum ||F_{\text{obs}}| - |F_{\text{calc}}||}{\sum |F_{\text{obs}}|}$ , where  $|F_{\text{obs}}|$  and  $|F_{\text{calc}}|$  are the observed and calculated structure factor amplitudes, respectively. Summation includes all reflections used in the refinement.

<sup>c</sup>  $R_{\text{free}} = \frac{\sum ||F_{\text{obs}}| - |F_{\text{calc}}||}{\sum |F_{\text{obs}}|}$ , evaluated for a randomly chosen subset of 5% of the diffraction data not included in the refinement.

<sup>d</sup> r.m.s.d. from ideal values.

of approximately 10 molecules of cholate per protein molecule in the crystallization liquor. Both forms belong to space group C2 and were identified with the labels A and B (see Table 1). Form A grows in sodium citrate at pH7.0 and contains three molecules in the asymmetric unit, whereas form B grows in ammonium sulfate buffered with Hepes at pH7.5 and contains two molecules in the asymmetric unit. Although the crystals grow under rather different conditions, they diffract to approximately the same resolution, 2.2 Å.

The co-crystals of form A grow under very similar conditions as the apoprotein but they belong to a different space group. Data collection statistics are summarized in Table 1. The structure was solved by molecular replacement and refined as described in the appropriate section. The refinement statistics are also given in Table 1. The Ramachandran plot of this model has 93.2% of the residues in the most favorable region of the plot and the remaining 6.8% are in the additionally allowed region. The conventional *R*-factor of the model is 22.0% and  $R_{\text{free}}$  is 26.4%.

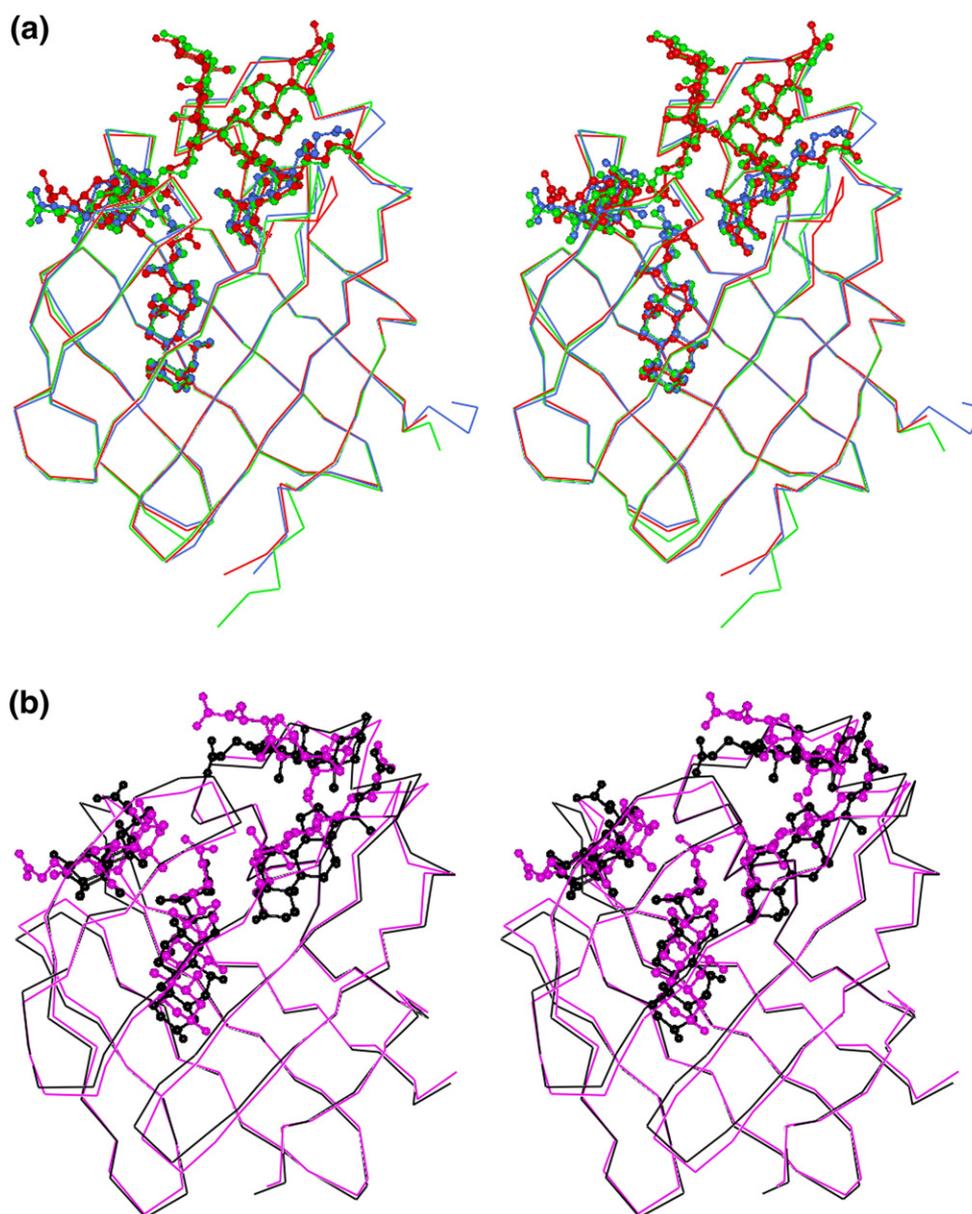
The co-crystals of form B grow at pH7.5 in the presence of ammonium sulfate as the precipitant instead. The refinement statistics are also given in Table 1. In this case, the Ramachandran plot has 94.6% of the residues in the most favorable region of the plot and the remaining 5.4% are in the additionally allowed region. The conventional *R*-factor of the model is 22.6% and  $R_{\text{free}}$  is 25.4%.

In all the five protein molecules examined in the two crystal forms, two cholate molecules are present in the internal ligand-binding site. Furthermore, the two cholate molecules overlap almost perfectly in the binding site of four out of the five protein molecules, and in the fifth (molecule B of form B), one overlaps with the other four, whereas the other is in the same position but rotated approximately 90° about the longest cholate axis.

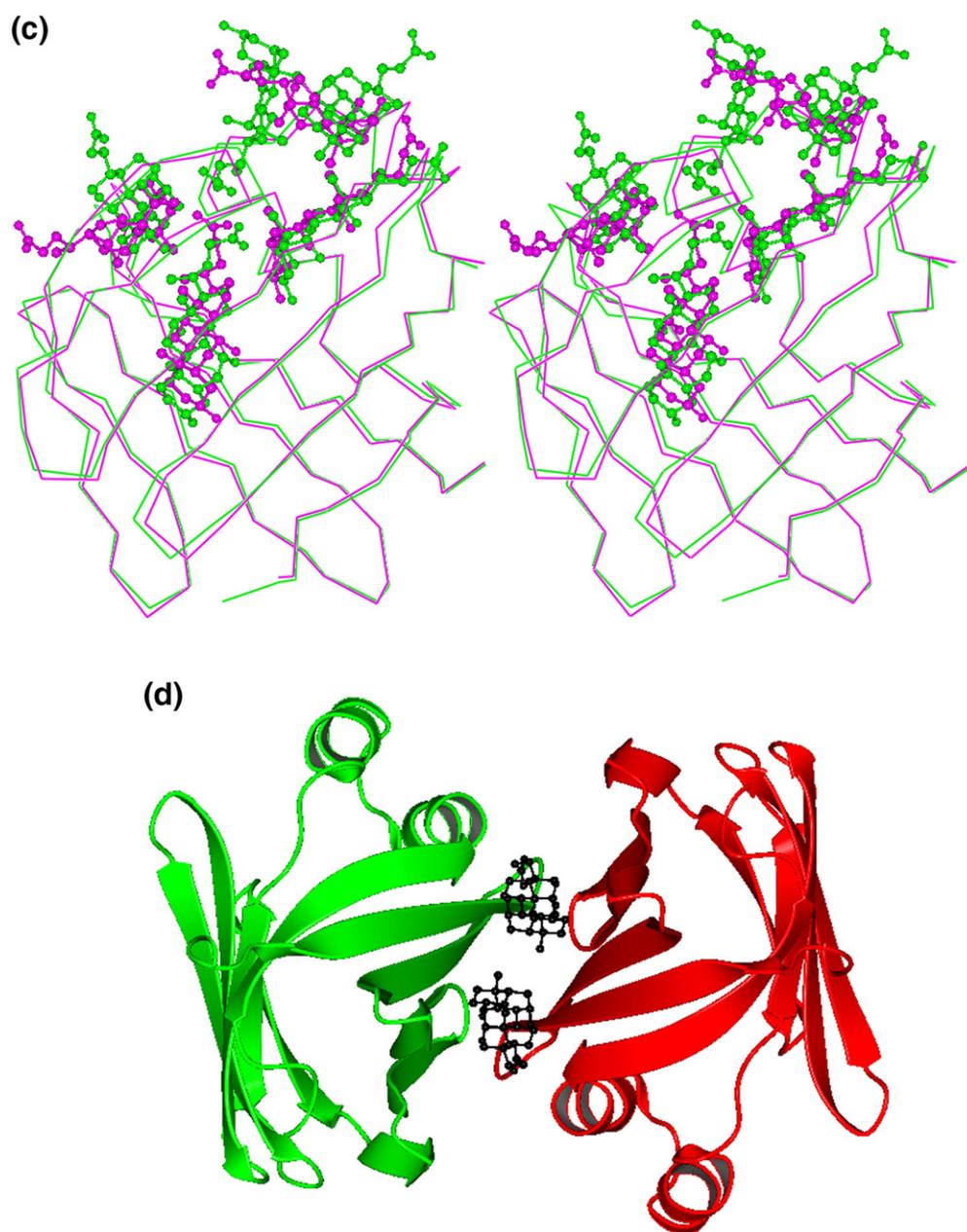
An unexpected finding is the presence of cholate molecules bound on the surface of all the five protein molecules examined in the two crystal forms. Figure 1a is a stereodiagram of the superposition of the three molecules present in the asymmetric unit of form A. The coordinates were superimposed

using the program LSQKAB.<sup>33</sup> Cholate models were introduced in the asymmetric unit only if their density was completely clear and unambiguous. In two out of the three protein molecules present in the asymmetric unit, there are three cholate molecules in approximately identical position bound on the surface of the protein. In the third protein molecule (molecule C in our notation), only one cholate mole-

cule on the surface of the protein is sufficiently convincing and overlaps with one of the cholates present in the other two. Figure 1b is an analogous diagram in which the two molecules superimposed are those present in the asymmetric unit of form B. In this case, only two cholates were found bound on the surface of the protein and more variability in their orientation was observed. Remarkably, an



**Fig. 1.** Crystal structure of zebrafish I-BABP complexed with cholate. (a) Stereo view of the superposition of the three molecules in the asymmetric unit of crystal form A (see Table 1). The electron density of two cholate molecules on the surface in one of the I-BABP molecules in the asymmetric unit (molecule C) was not convincing enough to be modeled. (b) Superposition of the two molecules present in the asymmetric unit of crystal form B of the complex (see Table 1). Note that in this case, the cholate molecules occupy approximately the same position but their overlap is less extensive than in form A. (c) Superposition of one molecule from each of the two crystal forms. The model in green is molecule A from crystal form A and the model in magenta is molecule A from crystal form B. (d) Ribbon representation of two symmetry-related molecules of crystal form B. The crystallographic 2-fold axis is approximately perpendicular to the plane of the figure. The additional cholate molecule, not represented in the other diagrams, is colored black. The coordinates were superimposed using the program LSQKAB.<sup>33</sup> The figures were prepared using the visualization program CCP4 mg.<sup>34</sup>



**Fig. 1** (legend on previous page)

important difference was also found in one of the ligand molecules bound in the internal cavity of one of the protein molecules (molecule B in our notation), which is rotated about  $90^\circ$  with respect to the position found in all the other protein molecules examined in the crystals. Figure 1c shows the superposition of one protein model from each of the two crystal forms (in both cases, molecule A in our notation). Note that in this case, there is more variability in the orientation of the molecular axes of the cholates bound on the protein surface, although the area where they bind is approximately the same. Note also that the third cholate molecule absent in form B and in one of the protein molecules of form A makes important contacts with the other two cholates on the protein surface.

An additional cholate molecule was found bound only on the surface on molecule B of crystal form B. Its electron density is very clear, but it is found at the interface between two crystallographically related B protein molecules and interacts with both. We have interpreted its presence in the asymmetric unit as the result of a crystal packing artifact. Figure 1d shows a model of two molecules B of form B related by the crystallographic dyad with these cholate molecules at their interface.

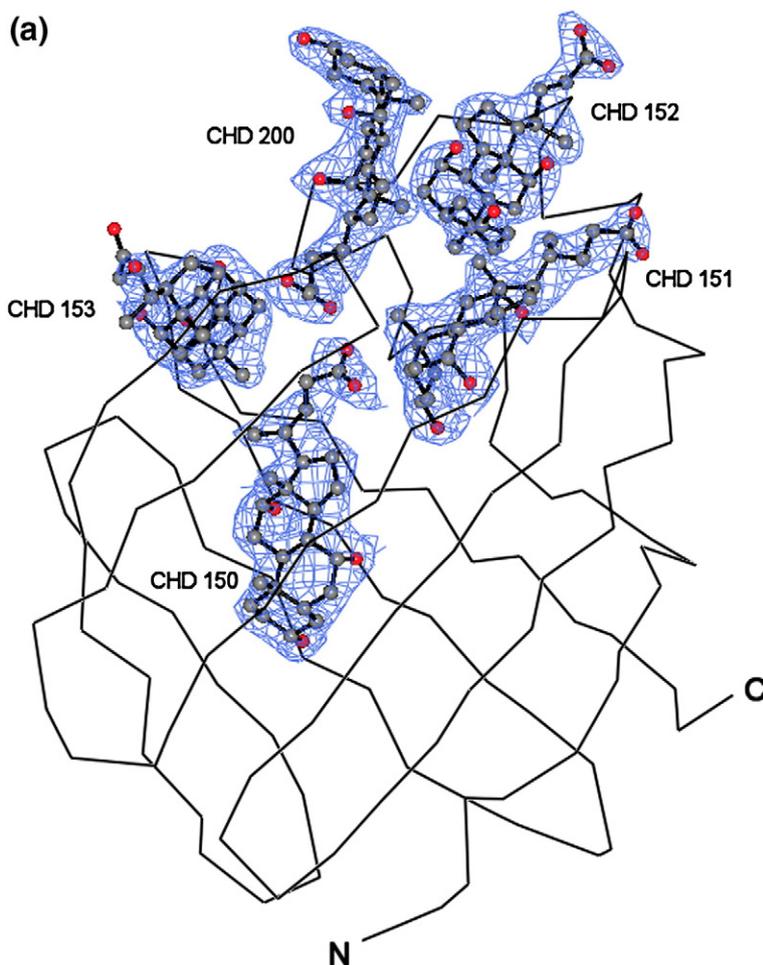
#### Ligand binding

For the sake of clarity, our discussion will focus on the binding to molecule A of crystal form A. Figure 2a shows the electron density of the five

ligands bound to the protein. As the figure shows, the electron density is very clear for all five molecules. A similar situation is observed in the case of molecule B of this crystal form, but the third molecule in the asymmetric unit (molecule C) does not show an electron density that is convincing enough to be modeled for two out of the three cholates bound on the protein surface. Figure 2b is a diagram representing the interactions of the two ligands within the internal cavity of the protein. Table 2 lists the distances between side chains and cholate molecules, and Table 3 lists the contacts between structured solvent molecules and cholates and protein side chains in the cavity. In addition to the hydrophobic contacts, the main direct interactions of the cholates with side chains are established with three tyrosines: the OH of Tyr97 with O12 of cholate molecule number 150 and the OH of Tyrosine 14 and 53 with O12 and O7, respectively, of cholate 151. The other

contacts with side chains are bridged by hydrogen-bonded water molecules in the cavity, which interact with Gln51, Gln99, Glu110, and Arg125 and with the O7 atoms of the two ligands and the carboxylate of cholate 150. Important hydrophobic contacts are established with Ile23, Gly31, Trp49, Phe63, Met71, Val83, Leu90, Ile92, and Thr101.

There is considerably more variability in the alignment of the cholate molecules bound on the surface of the protein. Cholates 152 and 153 are present in the two crystal forms, but their precise orientation on the protein surface is somewhat different (see Fig. 1a–c). This may be due to the lack of strong hydrogen bond interactions that could fix the ligand with its axis in a well-defined position. The cholate molecule that we have labeled 200 is present only in molecules A and B of crystal form A. It interacts with both molecules 152 and 153, with the cholates bound in the internal cavity, 150 and 151, and with



**Fig. 2.** Cholate binding to zebrafish I-BABP. (a)  $\alpha$ -Carbon chain trace of the molecule showing the experimental electron density of the five cholate molecules bound to molecule A of crystal form A. The  $2F_o - F_c$  map was contoured at a  $1.5 \sigma$  level. (b) Diagram representing the amino acid side chains and solvent molecules in contact with the two molecules bound in the internal cavity of the protein molecule. Hydrogen bonds are indicated with green dotted lines, whereas the amino acids that make hydrophobic contacts are only indicated but not represented as ball-and-stick models. (c) The same type of diagram as (b) but with the three cholate molecules bound on the surface of the protein molecule. (d) The three cholate molecules bound on the surface of molecule A of crystal form A. The protein molecular surface is represented yellow in the hydrophobic patches while polar regions are red (acid) or blue (basic). (a) was prepared using the program CCP4 mg, whereas (b) and (c) were prepared using the visualization program LIGPLOT.<sup>35</sup>

(b)

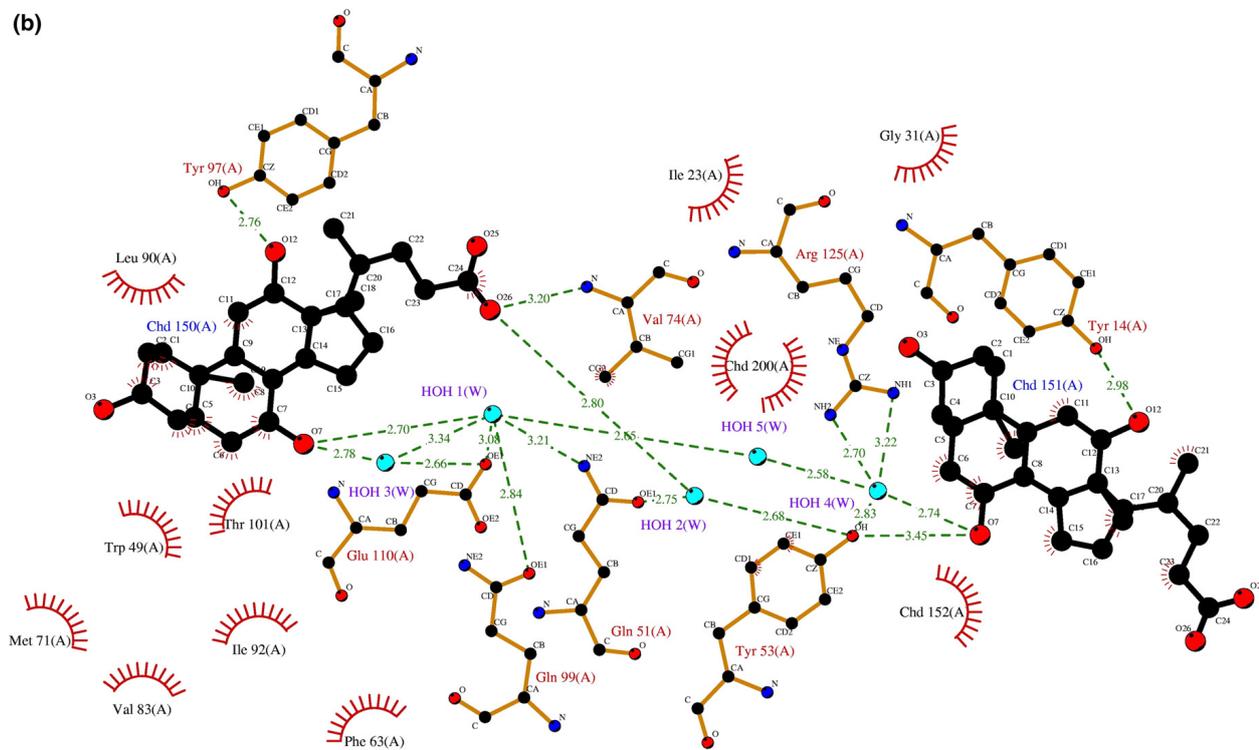


Fig. 2 (legend on page 104)

(c)

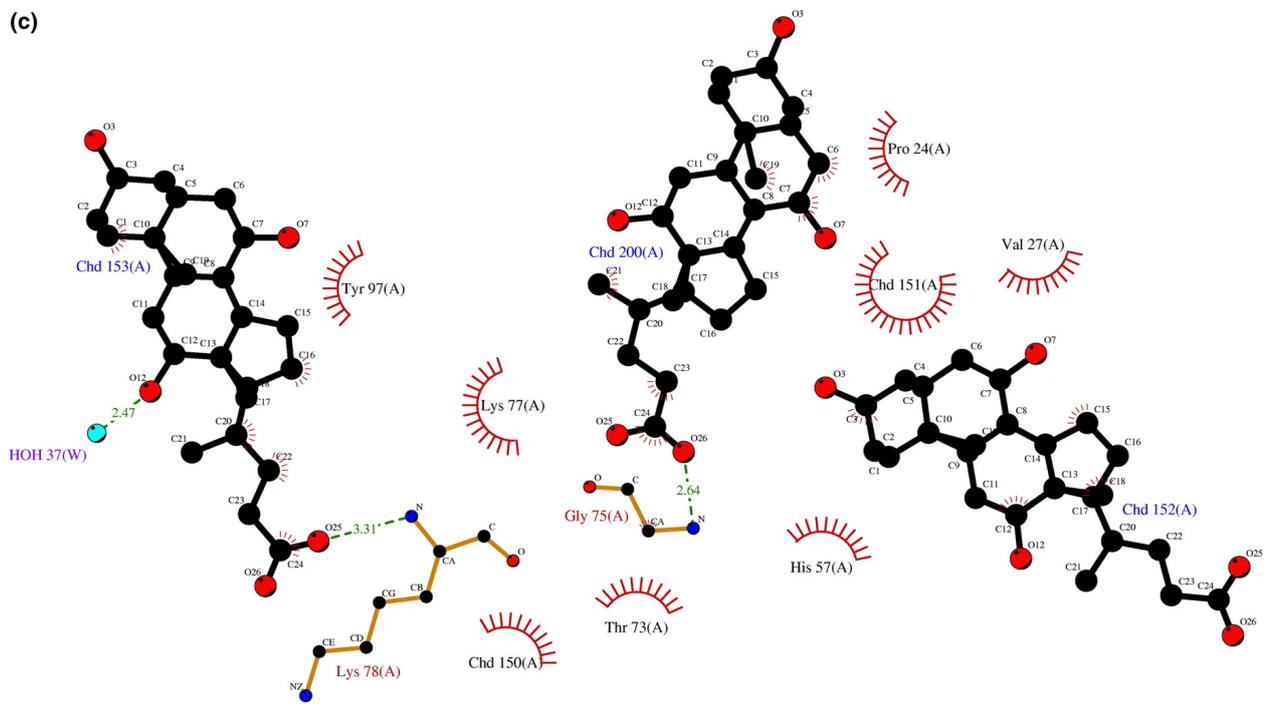


Fig. 2 (legend on page 104)

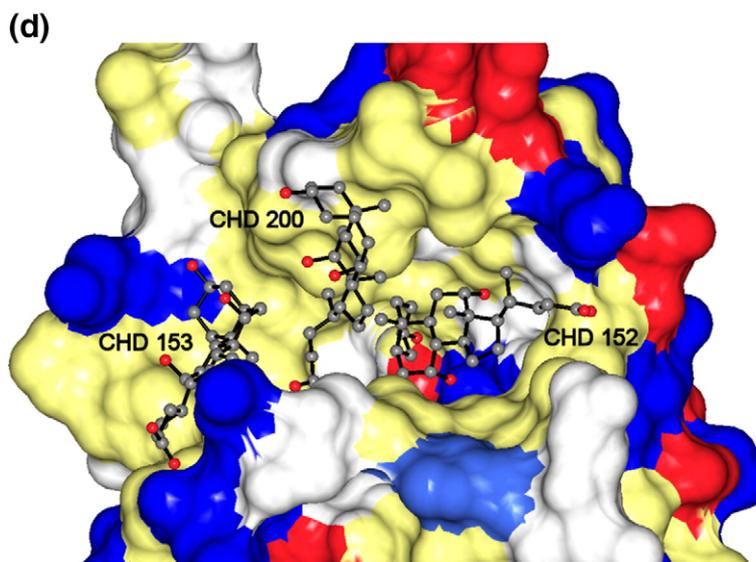


Fig. 2 (legend on page 104)

the following protein side chains: Ile23, Pro24, Thr73, and Gly75 (see Table 2). The presence of both cholates 152 and 153 seems to be a necessary although not sufficient condition for the binding of molecule 200. The main hydrophobic contacts of the protein molecule with the cholates on its surface are represented in Fig. 2c. They involve the following residues: Pro24, Val27, His57, Thr73, Lys77, and Tyr97. Figure 2d shows the protein surface with cholates 152, 153, and 200 bound to it. The polar residues are represented red (acid) and blue (basic) while the hydrophobic surface is yellow. Note that the region with the bound cholates is virtually completely hydrophobic.

### ITC experiments

Figure 3 reports the experimental ITC data in a  $\Delta H-r$  plot ( $\Delta H$ =cumulative enthalpy per mole of protein;  $r$ =concentration ratio=total titrated ligand/total protein) at two different temperatures: 25 and 35 °C.

Several binding models were tested to interpret these data. The best-fit trials using the simplest model (one binding site) and the model that implies two independent (but not necessarily equivalent) binding sites did not yield satisfactory results.

A good fit was instead obtained at both temperatures with the model that implies three independent (but not necessarily equivalent) binding sites (see Fig. 3). The relevant values (see Table 4) of the thermodynamic parameters (binding constant, etc.) suggest the presence of two binding sites, presumably located within the cavity (significant enthalpy values) and an extra entropy-driven binding site (very low enthalpic contribution), which is compatible with an adhesion process on the surface of the molecule. The external binding might indeed have no precise location, mainly being a physical adhesion of ligand molecules to

low polarity regions of the protein surface. This is in line with the X-ray results (see above) that reveal a poorly defined position of the surface-bound ligands. It must, however, be stressed that the local environment during the formation of protein crystals is rather different from that met in the dilute solution used in ITC experiments. The possibility of external adhesion of ligands could nonetheless affect their actual concentration in the solution bulk in subsaturation conditions and the overall partition of accessible states. This is indeed the effect of adding such extra binding in the simple model of two independent sites that improves the fitting of the ITC data.

Increasing the temperature from 25 to 35 °C substantially increased the overall enthalpic contribution without changing significantly that of the third site (see Table 4).

However, a satisfactory fit was also obtained at both temperatures with a fourth model that implies the consecutive binding to a couple of not necessarily equivalent sites (see Fig. 3). In this case, the protein states are three: free, first site occupied, and both sites occupied. The relevant partition function in this case is:  $Q=1+K_{b1}\cdot[L]+K_{b1}\cdot K_{b2}\cdot[L]^2$ . The values of the best-fit thermodynamic parameters (Table 4) suggest that both ligands should be located within the cavity. This model is simpler insofar as it does not imply any external binding and corresponds to a commonly met situation.

In conclusion, the available calorimetric data confirm the presence of two molecules bound within the cavity but do not allow a clear-cut choice between the latter two models. The temperature change did not allow discriminating between the two alternatives. The presence of surface binding sites, with a considerable effect on the overall binding mechanism in diluted protein solutions, remains a thermodynamic compatible possibility that needs further confirmation. In any case, such surface binding should be

**Table 2.** Distances between the closest z-I-BABP residues of molecule A of crystal form A and the cholates molecules separated in two groups: those bound in the internal cavity are labeled 150 and 151, whereas those on the surface are labeled 152, 153, and 200

Main contacts between the cholic acid molecules and z-I-BABP residues				
Cholate molecule	Atom	BABP residue	Atom	Distance (Å)
Molecules bound in the internal cavity				
150	O26	Val74	N	3.20
150	O3	Leu90	O	2.80
150	O12	Tyr97	OH	2.76
150	O7	Gln99	NE2	3.39
150	O12	Gln99	NE2	2.65
150	O3	Thr101	OG1	3.41
150	O3	Thr101	N	3.47
151	O12	Tyr14	OH	2.98
151	C12	Tyr14	OH	3.81
151	C11	Ile23	CD1	3.75
151	O26	Lys30	O	3.97
151	O7	Tyr53	OH	3.45
151	C15	Tyr53	CE1	3.59
151	C7	Val74	CG2	3.70
151	O7	Arg125	NH2	3.31
151	O3	Arg125	NH2	4.04
Molecules bound on the surface of the protein				
152	C18	Val27	CG1	3.84
152	C20	Val27	CG2	4.00
152	C21	Lys30	CD	3.91
152	O25	Lys30	NZ	4.15
152	C3	His57	CG	3.65
152	C19	Val74	CB	4.07
152	C5	Val74	O	3.93
153	C4	Ile21	CG2	4.00
153	C16	Lys77	CG	3.79
153	C22	Lys77	CB	3.82
153	O25	Lys78	N	3.31
153	C21	Phe79	CD1	3.95
153	C1	Tyr97	CB	3.82
200	C21	Ile23	CD1	4.08
200	C19	Pro24	CD	3.83
200	C24	Thr73	CG2	2.98
200	O25	Thr73	CG2	2.60
200	O26	Thr73	CG2	2.85
200	C24	Gly75	N	3.32
Other contacts of the central cholic acid molecule on the surface				
Cholate molecule	Atom	Cholate molecule	Atom	Distance (Å)
200	C23	150	O25	3.04
200	C24	150	O25	3.26
200	O26	150	O26	3.59
200	C21	151	C19	3.73
200	C7	152	C15	3.80
200	O25	153	C15	3.63

Cholate number 200 is not visible in the other crystal form and in molecule C of form A.

entropy driven and therefore strongly dependent on the environmental conditions, that is, solvent and other solutes.

### Structural differences between the apo- and the holoprotein

The program LSQKAB<sup>33</sup> was used to superimpose the  $\alpha$ -carbons of residues 1–127 of the model of the

apoprotein and chains A of the two crystal forms of the complex with cholates, and the distances between equivalent  $\alpha$ -carbons were calculated. The mean r.m.s.d. values were 1.04 and 1.09 Å for crystal forms A and B, respectively. In addition, and to be used as a control, a similar computation was done for molecules A and B of crystal forms A (mean r.m.s.d.=0.35 Å) and B (mean r.m.s.d.=0.73 Å) and molecules A of the two crystal forms of the co-crystals (mean r.m.s.d.=0.58 Å). The results are represented in Fig. 4a as a function of the amino acid number. As the figure shows, there are areas of larger variability in the conformation of the protein: the two helices and the connections between strands C–D, E–F, and I–J. In all these areas, variations are observed when two molecules are compared, which, in every case, have bound cholates, but are present in different positions in the crystal asymmetric unit and/or crystal form. Somewhat larger variations are observed in the comparison of the apo and holo models, but in this case, an additional very large peak, almost identical in the two crystal forms, is found in the region between amino acids 90 and 100, that is, in the hairpin connecting strands G–H.

Figure 4b shows a holo molecule (red) superimposed to an apo molecule (blue). Arrows are used to indicate the areas where the two chains are more distant and a green arrow points to the hairpin connecting strands G–H, which, as the figure shows, is clearly in a more open conformation.

It is also worth mentioning that the side chains of several amino acids in these areas are involved in ligand binding in the internal cavity of the protein, especially to the cholates molecule that we have labeled 150.

The solvent accessible volume of the ligand-binding cavity of all the molecules present in the three crystal forms was calculated with the program CASTp.<sup>36</sup> The results are the following: for the apoprotein, it was 314.0 Å<sup>3</sup>, whereas for the three molecules in the asymmetric unit of form A of the co-crystals, the values were 1015.1, 905.3, and 948.9 Å<sup>3</sup>, and for the two molecules in form B, they were 1028.8 and 635.8 Å<sup>3</sup>. The increase in volume in the apo-to-holo transition shows that the conformational change takes place, as expected, with an increase in the volume of the ligand-binding cavity. In general, all the values of the models with cholates bound give comparable figures with the exception of molecule B of crystal form B (635.8 Å<sup>3</sup>), which, interestingly, is the only one that has one of the cholates molecules bound in the cavity in a different orientation (see Fig. 1b). These values, when compared with the equivalent for similar proteins, are somewhat larger; for example, the solvent accessible volume of the C91T mutant of zebrafish L-BABP is 675.4 Å<sup>3</sup> and that for chicken L-BABP is 627.0 Å<sup>3</sup>. This difference may be due to the presence of four extra amino acids inserted in the sequence of zebrafish I-BABP in the loop connecting the last two strands of the  $\beta$ -structure (see below).

**Table 3.** Distances between solvent molecules present in the crystals, the cholates bound in the internal cavity, and protein side chains

Main contacts between solvent molecules, cholic acid molecules, and side chains in molecule A of crystal form A				
Cholate molecule or side chain	Atom	Solvent molecule holoprotein	Atom	Distance (Å)
CHD 150	O7	HOH 1	O	2.70
Gln99	OE1	HOH 1	O	2.84
Glu110	OE1	HOH 1	O	3.08
HOH 3	O	HOH 1	O	2.65
CHD 150	O26	HOH 2	O	2.80
Tyr53	OH	HOH 2	O	2.68
Gln51	OE1	HOH 2	O	2.75
CHD 150	O7	HOH 3	O	2.78
Glu110	OE1	HOH 3	O	2.66
CHD 151	O7	HOH 4	O	2.74
Arg125	NH2	HOH 4	O	2.70
Tyr53	OH	HOH 4	O	2.83
HOH 3	O	HOH 4	O	2.58
HOH 2	O	HOH 5	O	2.65
HOH 4	O	HOH 5	O	2.58
Amino acid side chain	Atom	Solvent molecules that are in the same position in the cavity of the apoprotein	Atom	Distance (Å)
Leu108	O	HOH 6	O	3.04
Glu110	OE2	HOH 6	O	2.45
Ser127	OG	HOH 6	O	2.71
Glu110	O	HOH 7	O	2.89
Ser112	OG	HOH 7	O	2.63
Trp49	O	HOH 8	O	2.83
Gln51	OE1	HOH 8	O	2.89
Asn61	OD1	HOH 8	O	2.92
Thr38	OG	HOH 9	O	2.90
Gln51	NE2	HOH 9	O	2.77
Glu110	OE1	HOH 9	O	2.72
Leu36	O	HOH 10	O	2.94
Thr38	OG	HOH 10	O	2.55
Ser127	OG	HOH 10	O	2.90

The notation is the same as in Table 2.

### Comparison with other BABPs

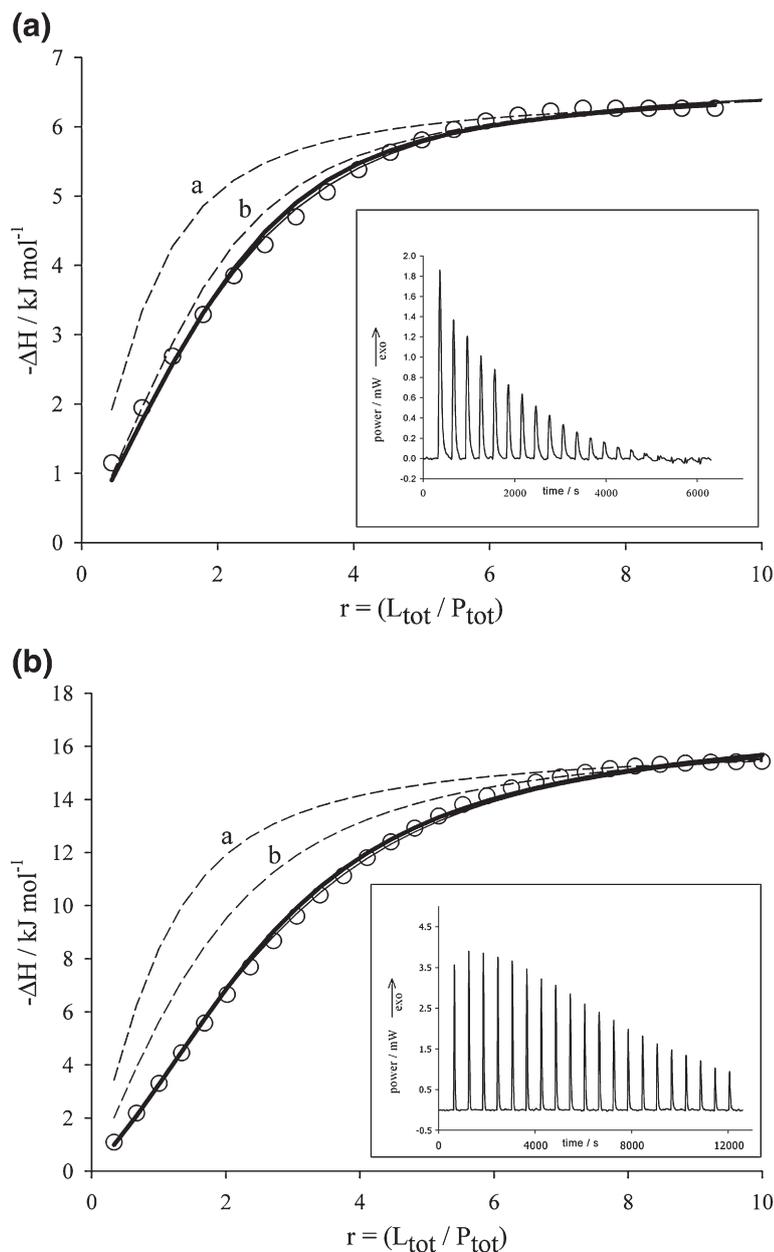
The liver of fish, amphibians, reptiles, and birds expresses a BABP that belongs to the same structural family as the I-BABPs. These proteins were originally called the liver "basic" FABPs, and there are currently several X-ray structures available,<sup>25–27</sup> including that of the wild type and two mutants of zebrafish L-BABP.<sup>30</sup> Unlike all the other members of the family, wild-type zebrafish L-BABP can accommodate a single cholate molecule in its internal binding site, but the mutation of a single amino acid, Cys91, changes this stoichiometry of ligand binding to the one found in all the other members of the family: two cholates per site. Figure 5a is a stereodiagram that superimposes the model of this mutant of zebrafish L-BABP and the model of zebrafish I-BABP. Note that even though the two cholates bind in approximately the same region, their exact position and orientation are not identical.

NMR spectroscopy was used to examine the binding of glycocholate<sup>14</sup> and taurocholate<sup>15</sup> to porcine and human I-BABPs, respectively. The two lowest-energy models that contain a single bile acid molecule in the ligand-binding site were super-

imposed to the model of zebrafish I-BABP. The result of this superposition is represented in Fig. 5b. Note that in both cases, the single molecule included in the NMR structures occupies approximately the position of one of the molecules in the X-ray model, but the precise position and orientation of the ligand are different.

The sequences of five I-BABPs from different species were aligned using the program CLUSTALW<sup>37</sup> and are represented in Fig. 6. Note that after amino acid 118, there is an insertion of four amino acids: GTAV present only in zebrafish I-BABP. The last column on the right hand indicates the percentage identity of the amino acid sequences with zebrafish I-BABP, which is in every case lower than about 60%. A black dot identifies the residues involved in ligand binding of the cholates present in the internal cavity of zebrafish I-BABP, and a gray dot identifies those on the protein surface. In general, the residues involved in ligand binding are either conserved or substituted by acceptable alternatives.

The bottom part of the figure aligns zebrafish I-BABP and L-BABP. Note that the four amino acids absent in the other I-BABPs are also absent in L-BABP and that the percentage identity between the



**Fig. 3.** ITC experiments. Sodium cholate binding to zebrafish I-BABP at 25 °C (a) and 35 °C (b). The raw data for the first 20 injections are shown in the inset. Experimental data (circles) in the  $\Delta H-r$  plot where  $\Delta H$  is the cumulative enthalpy (sum of the peak areas in the inset) expressed per mole of protein versus the concentration ratio and  $r$  is the concentration ratio  $r = \text{total titrated ligand} / \text{total protein}$ . The lines represent the best fit of the data using four thermodynamic binding models: single site (dotted thin line a), two independent sites (dotted thin line b), three independent sites (continuous bold line), and two consecutive sites (continuous thin line almost superimposed to the bold line).

two sequences is only about 44%. The black arrows identify the residues involved in cholate binding in the cavity of I-BABP, whereas the gray arrows identify those that participate in cholate binding in

the cavity of the C91T mutant of zebrafish L-BABP. About half the residues that participate in ligand binding align in the two sequences and the other half do not.

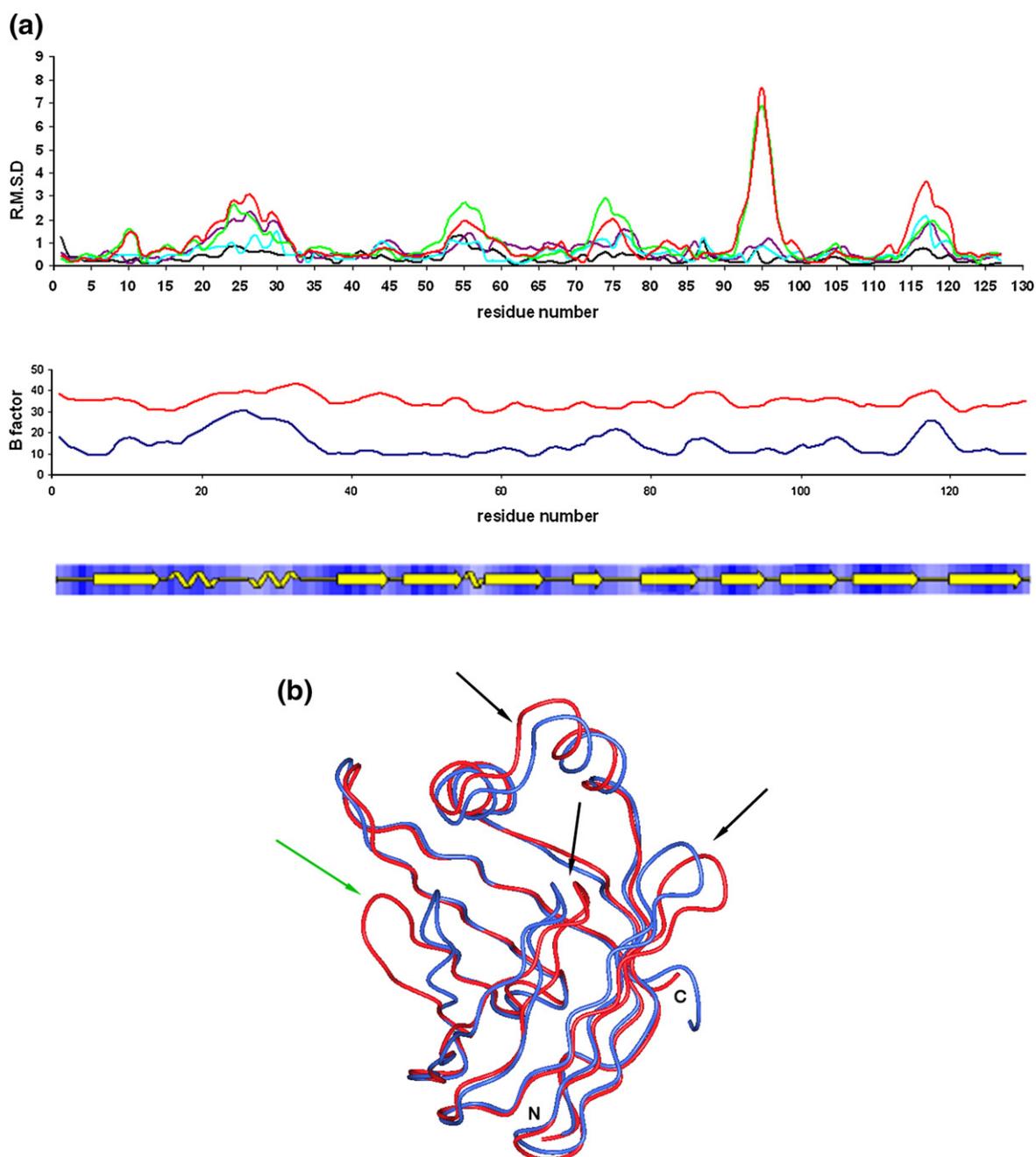
**Table 4.** Best-fit parameters obtained using the three independent sites binding model and the two consecutive sites binding model on the ITC data

$T$ (°C)	$K_{b1}$	$K_{b2}$	$K_{b3}$	$\Delta H_{b1}^\circ$	$\Delta H_{b2}^\circ$	$\Delta H_{b3}^\circ$	$T\Delta S_{b1}^\circ$	$T\Delta S_{b2}^\circ$	$T\Delta S_{b3}^\circ$	$\Delta G_{b1}^\circ$	$\Delta G_{b2}^\circ$	$\Delta G_{b3}^\circ$
Three independent sites binding model												
25	3.45	1.45	3.50	-4.7	-1.8	-0.2	21.20	21.95	25.74	-25.90	-23.75	-25.94
35	0.93	0.65	3.00	-9.4	-8.2	-0.3	14.01	14.29	26.11	-23.41	-22.49	-26.41
Two consecutive binding sites model												
25	0.63	1.02	—	-5.1	-1.6	—	16.59	21.28	—	-21.69	-22.88	—
35	0.21	0.84	—	-10.6	-6.4	—	9.00	16.75	—	-19.60	-23.15	—

$K_b$  is expressed in  $10^4 \text{ M}^{-1}$ . All the other parameters are expressed in kilojoules per mole.

In both the models,  $n=1$  fixed stoichiometry was assumed for each binding site.

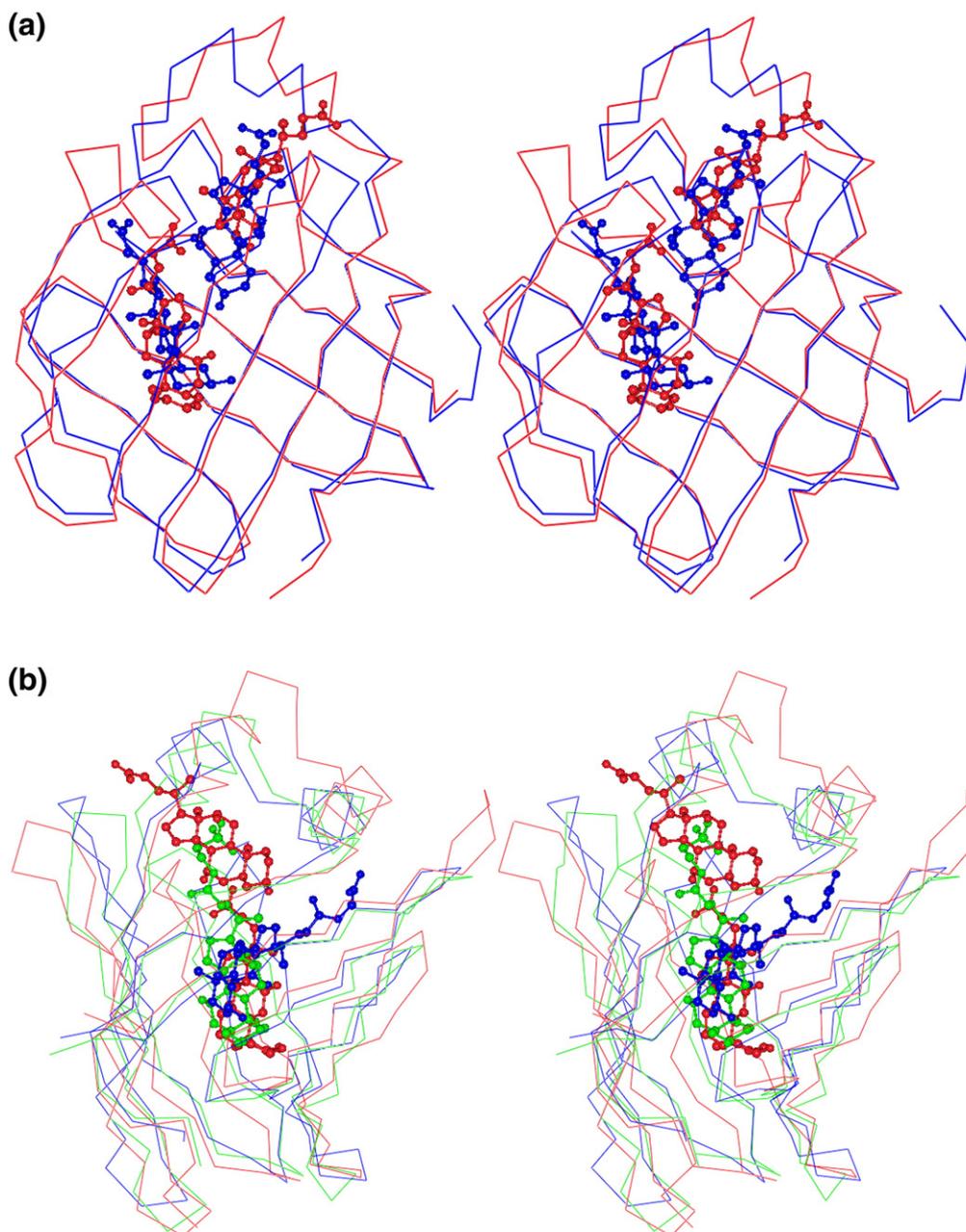
Errors are within 10% of the reported values, lower than 15% if an eventual error of 10% is taken into account in the protein concentration.



**Fig. 4.** Comparison of the apo- and holoprotein models. (a) Upper panel: r.m.s.d. between  $\alpha$ -carbon atoms of the apoprotein model and the models of the holoprotein: (i) A chain of form A (red), (ii) A chain of form B (green). The black trace compares chains A and B of crystal form A, the violet trace compares chains A and B of crystal form B, and the cyan trace compares chains A of the two crystal forms. Lower panel: main-chain *B*-factors of the apo form (blue) and molecule A of crystal form A (red) as a function of the amino acid number. The strip at the bottom of the figure represents the elements of secondary structure. (b) The models of the apoprotein (blue) and the holoprotein (red) superimposed using the program LSQKAB.<sup>33</sup> The model of the holoprotein used is molecule A of crystal form A. Note that the cavity covered by the two helices is more open in the holoprotein. The regions where the two polypeptide chains are more distant are indicated with arrows.

It is worth mentioning that the existence of a superficial binding site for cholate in zebrafish L-BABP was revealed by ITC although electron density for it was not present in the maps.<sup>30</sup> The observation that rabbit I-BABP is also photolabeled on its surface by derivatives of taurocholate further strength-

ens this result.<sup>21</sup> Experimental data to support the presence of specific binding sites on hydrophobic surfaces of a protein molecule are not easy to acquire, and in the case of X-ray crystallography, the presence of the ligand may interfere in the crystallization process. This may explain the fact that, up



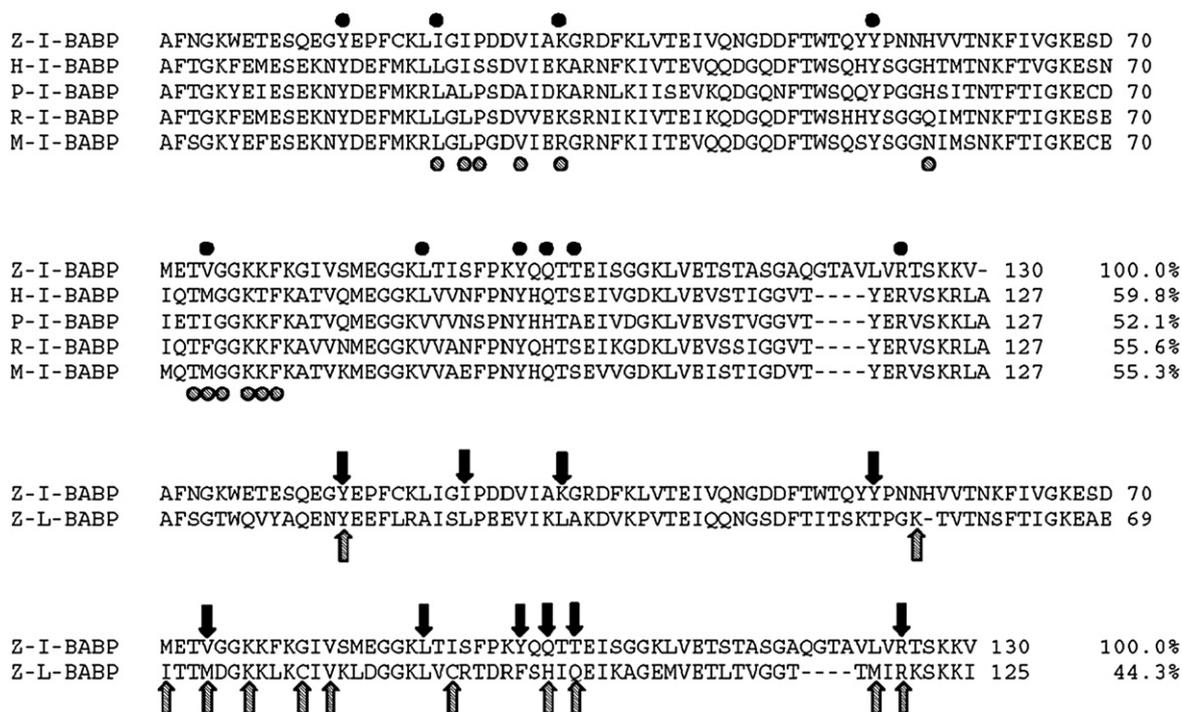
**Fig. 5.** Comparison of the model of zebrafish holo I-BABP (molecule A of crystal form A) and the models of other BABPs. (a) Stereo diagram superimposing zebrafish I-BABP (red) and the C91T mutant of zebrafish L-BABP (blue, Ref. [30], PDB code: 2QO5). (b) Stereo diagram of the model of zebrafish holo I-BABP (red) superimposed to the NMR models of porcine I-BABP complexed with glycocholate (blue, Ref. [14], PDB code: 1EIO) and human I-BABP complexed with taurocholate (green, Ref. [15], PDB code: 1O1V). The coordinates used for both NMR structures were the first sets listed in the PDB files. The models were superimposed by using the program LSQKAB<sup>33</sup> and only the two cholate molecules bound in the internal cavity of zebrafish I-BABP are represented in the diagrams.

to now, no I-BABP crystals from other species could be grown. With the information gathered so far, the physiological role of these extra sites on the protein surface can only be a matter of speculation: they could help guide the molecules to the internal sites or perhaps could improve the ligand-binding capacity of a small protein molecule in the presence of a large excess of the ligand.

## Materials and Methods

### Protein expression, purification, complex formation, and crystallization

The cDNA coding for zebrafish I-BABP (IMAGE ID 7039169), obtained from RZPD (Deutsches Ressourcen-zentrum fuer Genomforschung GmbH), was amplified by



**Fig. 6.** Sequence alignment of five I-BABPs and zebrafish I-BABP and L-BABP. The abbreviations used are as follows: Z, zebrafish; H, human; P, porcine; R, rat; M, mouse; L-BABP, liver BABP (formerly called liver "Basic" FABP). The last column on the right-hand side gives the percentage identity of each sequence and that of zebrafish I-BABP. A black dot identifies the residues involved in the binding of cholate to zebrafish I-BABP in the internal cavity, whereas a gray dot identifies those involved in the binding on the surface of the molecule. The black arrows identify the residues involved in cholate binding in the cavity of zebrafish I-BABP, and the gray arrows identify those involved in cholate binding in the cavity of the C91T mutant of zebrafish L-BABP.

PCR using primers designed to introduce restriction sites for BamHI and PstI endonucleases and a sequence coding for a digestion site for thrombin in the C-terminal end in the amplified fragment. After purification, the fragment and the expression vector pQE50 (Qiagen) were digested with the restriction enzymes mentioned above and incubated with ligase to insert the cDNA in the vector respecting the reading frame. SG13009 *Escherichia coli* cells were transformed with the resulting vector, grown at 37 °C, and protein synthesis was induced overnight at 20 °C with 0.5 mM IPTG (isopropyl  $\beta$ -D-1 thiogalactopyranoside). Under these conditions of subcloning in pQE50, the expressed intracellular domain is fused to a histidine tag through its C terminus. The presence of the tag allowed the affinity purification of the fused protein by passing the bacterial extracts through a nickel-Sepharose column. The column was equilibrated with 20 mM Tris-HCl, pH7.5, 0.5 M NaCl, 10 mM imidazole, and 0.02%  $\text{NaN}_3$ , and the bound protein was eluted with a linear gradient of imidazole from 10 to 500 mM. After the affinity column, the tag was removed by thrombin digestion and the protein was further purified by gel filtration in a Superdex G-75 column equilibrated with 20 mM Tris-HCl, pH7.5, 0.15 M NaCl, and 0.02%  $\text{NaN}_3$ . Complete removal of the tag was assessed by Western blot analysis using an anti-His-horseradish peroxidase-conjugated antibody (Sigma-Aldrich). The purified protein showed one band in SDS-PAGE. Ten times the molar protein concentration of sodium cholate was added to the apoprotein at a concentration of about 20 mg/mL in 20 mM Tris-HCl buffer, pH7.5, in order to prepare the complex with the ligand. The cholate concentration in the protein solution

was thus about 14 mM. The solution was stirred overnight at 20 °C and used at this concentration for the initial screen of crystallization conditions. Molecular Dimensions Structure Screens were used at 20 °C with the hanging-drop method, mixing 1  $\mu$ L of the protein solution with the same volume of the precipitating solution and equilibrating *versus* a volume of 0.3 mL of the latter in the reservoir. The conditions yielding small crystals were later refined, and the sitting-drop method with larger volumes was also tested until crystals that were large enough for data collection were obtained. The best crystals of apo I-BABP grow by mixing equal volumes of the protein solution and 0.1 M sodium Hepes, pH7.5, 1.4 M trisodium citrate dihydrate, and 1% ethylene glycol. The co-crystals of I-BABP with cholate were grown using a protocol identical with that of the apoprotein. In this case, two different crystal forms were obtained. Both are monoclinic, belong to space group C2, and are identified with the labels forms A and B. The first form (A) grows under conditions that are very similar to those of the apoprotein: 1.1 M trisodium citrate dihydrate, pH7.0, while the second form (B) grows in 0.1 M sodium Hepes, pH7.5, 2.0 M ammonium sulfate.

#### Data collection, structure solution, and refinement

The crystals of apo I-BABP are trigonal, belong to space group  $P3_221$ , with  $a=b=54.4$  Å and  $c=82.7$  Å, and contain one molecule in the asymmetric unit (see Table 1). The crystals of form A of the complex of I-BABP with cholate are monoclinic, belong to space group C2, with  $a=149.2$  Å,

$b=80.3$  Å,  $c=39.7$  Å, and  $\beta=96.2^\circ$ , and contain three molecules in the asymmetric unit. The crystals of form B belong to the same space group (C2), with  $a=96.3$  Å,  $b=85.3$  Å,  $c=75.1$  Å, and  $\beta=127.6^\circ$ , and contain two molecules in the asymmetric unit (Table 1).

The data for the crystals of the apoprotein and of form A of the complex with cholate were collected at the ID29 beamline of the European Synchrotron Radiation Facility in Grenoble ( $\lambda=0.97$  and  $0.93$  Å) at 100 K after a brief soaking in a mixture of 80% of the mother liquor and 20% glycerol. The data were indexed, integrated, and reduced using the programs MOSFLM and Scala.<sup>38,39</sup> The data for form B of the complex were collected at the XRD1 beamline of the Elettra Synchrotron in Trieste at 100 K with 80% mother solution–20% glycerol as cryoprotectant. The detector was a MarResearch CCD. The data were indexed, integrated, and reduced using the program AUTOMAR. The diffraction data statistics are summarized in Table 1.

The structure of form B of the complex I-BABP–cholate was solved first using the CCP4 suite of programs for crystallographic computing. The initial phases were calculated by the molecular replacement method as implemented in the program MOLREP,<sup>40</sup> with the coordinates of zebrafish L-BABP [Ref. [30], Protein Data Bank (PDB) accession code: 2QO4] as the search probe. The automatic search with data up to a resolution of 3.0 Å gave an unambiguous result with a correlation coefficient of 37.5 and an  $R$ -factor of 37.5%. The model was rigid body refined, moving initially the entire molecule and, in a second stage, the elements of secondary structure using the program REFMAC.<sup>31</sup> After the proper side chains had been introduced, the model was subjected to a series of rounds of positional refinement alternated with manual model revisions with the program XtalView<sup>41</sup> and the refinement program REFMAC. During the process of refinement and model building, the quality of the model was controlled with the program PROCHECK.<sup>32</sup> Ligand and solvent molecules were added to the model in the final stages of refinement, the latter according to hydrogen bond criteria and only if their  $B$ -factors refined to reasonable values and if they improved the  $R_{\text{free}}$ . The model was finally subjected to a final round of TLS refinement as implemented in the program REFMAC.

A similar procedure was followed to solve and refine the structure of the apoprotein and of form A of the complex. In both cases, one molecule of the I-BABP–cholate model in the asymmetric unit of form B without solvent molecules and the ligand was used as the search probe. Refinement of these structures was also carried out with the program REFMAC following essentially the same procedure described above. The final refinement statistics for the models of the three crystal forms are summarized in Table 1.

### Isothermal titration calorimetry

The proteins and sodium cholate were dissolved in the following buffer: 20 mM Tris–HCl, pH7.5, 0.15 M NaCl, 0.02% sodium azide. The sodium cholate concentration in the titrating solution was 9.3 mM, and the initial protein concentration in the measurement cell (determined spectrophotometrically) was 0.11 and 0.15 mM for the measurements performed at 25 and 35 °C, respectively. The titrations were performed using a CSC Nano Isothermal Titration Calorimeter III (model 5300). A total of 20 (25 °C) and 40 (35 °C) injections of 5- $\mu$ L aliquots of titrating solution were added to the 973- $\mu$ L protein solution cell. The heat of the injections was corrected for

the heat of dilution of the ligand into the buffer. Three replicas were performed. One replica was also performed using the MicroCal VP-ITC instrument.

Several binding models were tested to interpret the calorimetric data, and the fitting functions are described in the literature.<sup>42–44</sup>

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, that is, the concentration ratio  $\bar{x} = [\text{bound ligand}]/[\text{total protein}]$ , is given by

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

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.<sup>43</sup> 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

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

of the concentrations of all protein species,  $P_j$ , referred to the free protein,  $P_0$ .  $Q$  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 article.

The fit attempts based on the binding models were accomplished using the nonlinear Levenberg–Marquardt method.<sup>45</sup> The errors of each fitting parameter were calculated with a 95.4% confidence limit by the Monte Carlo simulation method. An eventual error of 10% in the protein concentration was also taken into account.

### Accession numbers

Coordinates and structure factors have been deposited in the PDB with accession numbers 3ELX, 3ELZ, and 3EM0.

### Acknowledgements

We are grateful to the staff of the European Synchrotron Radiation Facility in Grenoble (Proposal MX 456) and Elettra in Trieste for assistance during data collection. We also want to thank Prof. Alberto Schiraldi for a fruitful discussion of the thermodynamic data. This work was supported by a FIRB (Fondo per gli Investimenti della Ricerca di Base) grant from the Italian Ministry of the Universities and Scientific Research. Gianmaria Saccomani is the

recipient of a scholarship granted by Banco Popolare di Verona e Novara.

## References

- Trauner, M. & Boyer, J. L. (2003). Bile salt transporters: molecular characterization, function, and regulation. *Physiol. Rev.* **83**, 633–671.
- Alrefai, W. A. & Gill, R. K. (2007). Bile acid transporters: structure, function, regulation and pathophysiological implications. *Pharm. Res.* **24**, 1803–1823.
- Nakahara, M., Furuya, N., Takagaki, K., Sugaya, T., Hirota, K., Fukamizu, A. *et al.* (2005). Ileal bile acid-binding protein, functionally associated with the farnesoid X receptor or the ileal bile acid transporter, regulates bile acid activity in the small intestine. *J. Biol. Chem.* **280**, 42283–42289.
- Banaszak, L., Winter, N., Xu, Z., Bernlohr, D. A., Cowan, S. & Jones, T. A. (1994). Lipid-binding proteins: a family of fatty acid and retinoid transport proteins. *Adv. Protein Chem.* **45**, 89–151.
- Glatz, J. F. C. & van der Vusse, G. J. (1996). Cellular fatty acid-binding proteins: their function and physiological significance. *Prog. Lipid Res.* **35**, 243–282.
- Coe, N. R. & Bernlohr, D. A. (1998). Physiological properties and functions of intracellular fatty acid-binding proteins. *Biochim. Biophys. Acta*, **1391**, 287–306.
- Schaap, F. G., van der Vusse, G. J. & Glatz, J. F. (2002). Evolution of the family of intracellular lipid binding proteins in vertebrates. *Mol. Cell. Biochem.* **239**, 69–77.
- Zimmerman, A. W. & Veerkamp, J. H. (2002). New insights into the structure and function of fatty acid-binding proteins. *Cell. Mol. Life Sci.* **59**, 1096–1116.
- Hauerland, N. H. & Spener, F. (2004). Fatty acid-binding proteins—insights from genetic manipulations. *Prog. Lipid Res.* **43**, 328–349.
- Hertzel, A. V. & Bernlohr, D. A. (2000). The mammalian fatty acid-binding protein multigene family: molecular and genetic insights into function. *Trends Endocrinol. Metab.* **11**, 175–180.
- Walz, D. A., Wider, M. D., Snow, J. W., Dass, C. & Desiderio, D. M. (1988). The complete amino acid sequence of porcine gastrotropin, an ileal protein which stimulates gastric acid and pepsinogen secretion. *J. Biol. Chem.* **263**, 14189–14195.
- Gantz, I., Nothwehr, S. F., Lucey, M., Sacchettini, J. C., DelValle, J., Banaszak, L. J. *et al.* (1989). Gastrotropin: not an enteroxyntin but a member of a family of cytoplasmic hydrophobic ligand binding proteins. *J. Biol. Chem.* **264**, 20248–20254.
- Lücke, C., Zhang, F., Rüterjans, H., Hamilton, J. A. & Sacchettini, J. C. (1996). Flexibility is a likely determinant of binding specificity in the case of ileal lipid binding protein. *Structure*, **4**, 785–800.
- Lücke, C., Zhang, F., Hamilton, J. A., Sacchettini, J. C. & Rüterjans, H. (2000). Solution structure of ileal lipid binding protein in complex with glycocholate. *Eur. J. Biochem.* **267**, 2929–2938.
- Kurz, M., Brachvogel, V., Matter, H., Stengelin, S., Thuring, H. & Kramer, W. (2003). Insights into the bile acid transportation system: the human ileal lipid-binding protein–cholytaurine complex and its comparison with homologous structures. *Proteins: Struct., Funct., Genet.* **50**, 312–328.
- Tochtrop, G. P., Richter, K., Tang, C., Toner, J. J., Covey, D. F. & Cistola, D. P. (2002). Energetics by NMR: site-specific binding in a positively cooperative system. *Proc. Natl Acad. Sci. USA*, **99**, 1847–1852.
- Tochtrop, G. P., Bruns, J. L., Tang, C., Covey, D. F. & Cistola, D. P. (2003). Steroid ring hydroxylation patterns govern cooperativity in human bile acid binding protein. *Biochemistry*, **42**, 11561–11567.
- Tochtrop, G. P., DeKoster, G. T., Covey, D. F. & Cistola, D. P. (2004). A single hydroxyl group governs ligand site selectivity in human ileal bile acid binding protein. *J. Am. Chem. Soc.* **126**, 11024–11029.
- Toke, O., Monsey, J. D., DeKoster, G. T., Tochtrop, G. P., Tang, C. & Cistola, D. P. (2006). Determinants of cooperativity and site selectivity in human ileal bile acid binding protein. *Biochemistry*, **45**, 727–737.
- Toke, O., Monsey, J. D. & Cistola, D. P. (2007). Kinetic mechanism of ligand binding in human ileal bile acid binding protein as determined by stopped-flow fluorescence analysis. *Biochemistry*, **46**, 5427–5436.
- Kramer, W., Sauber, K., Baringhaus, K. -H., Kurz, M., Stengelin, S., Lange, G. *et al.* (2001). Identification of the bile acid-binding site of the ileal lipid-binding protein by photoaffinity labeling. *J. Biol. Chem.* **276**, 7291–7301.
- Kouvatsos, N., Meldrum, J. K., Searle, M. S. & Thomas, N. R. (2006). Coupling ligand recognition to protein folding in an engineered variant of rabbit ileal lipid binding protein. *Chem. Commun. (Cambridge)*, **44**, 4623–4625.
- Kouvatsos, N., Thurston, V., Ball, K., Oldham, N. J., Thomas, N. R. & Searle, M. S. (2007). Bile acid interactions with rabbit ileal lipid binding protein and an engineered helixless variant reveal novel ligand binding properties of a versatile beta-clam shell protein scaffold. *J. Mol. Biol.* **371**, 1365–1377.
- Scapin, G., Spadon, P., Mammi, M., Zanotti, G. & Monaco, H. L. (1990). Crystal structure of chicken liver basic fatty acid-binding protein at 2.7 Å resolution. *Mol. Cell. Biochem.* **98**, 95–99.
- Nichesola, D., Perduca, M., Capaldi, S., Carrizo, M. E., Righetti, P. G. & Monaco, H. L. (2004). Crystal structure of chicken liver basic fatty acid-binding protein complexed with cholic acid. *Biochemistry*, **43**, 14072–14079.
- Di Pietro, S. M., Córscico, B., Perduca, M., Monaco, H. L. & Santomé, J. A. (2003). Structural and biochemical characterization of toad liver fatty acid-binding protein. *Biochemistry*, **42**, 8192–8203.
- Capaldi, S., Guariento, M., Perduca, M., Di Pietro, S. M., Santomé, J. A. & Monaco, H. L. (2006). Crystal structure of axolotl (*Ambystoma mexicanum*) liver bile acid-binding protein bound to cholic and oleic acid. *Proteins*, **64**, 79–88.
- Schievano, E., Quarzago, D., Spadon, P., Monaco, H. L., Zanotti, G. & Peggion, E. (1994). Conformational and binding properties of chicken liver basic fatty acid-binding protein in solution. *Biopolymers*, **34**, 879–887.
- Ragona, L., Catalano, M., Luppi, M., Cicero, D., Eliseo, T., Foote, J. *et al.* (2006). NMR dynamic studies suggest that allosteric activation regulates ligand binding in chicken liver bile acid-binding protein. *J. Biol. Chem.* **281**, 9697–9709.
- Capaldi, S., Guariento, M., Saccomani, G., Fessas, D., Perduca, M. & Monaco, H. L. (2007). A single amino acid mutation in zebrafish (*Danio rerio*) liver bile acid-binding protein can change the stoichiometry of ligand binding. *J. Biol. Chem.* **282**, 31008–31018.

31. Murshudov, G. N., Vagin, A. A. & Dodson, E. J. (1997). Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **53**, 240–255.
32. Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. (1993). PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Crystallogr.* **26**, 283–291.
33. Kabsch, W. (1978). A solution for the best rotation to relate two sets of vectors. *Acta Crystallogr., Sect. A: Found. Crystallogr.* **32**, 922–923.
34. Potterton, E., McNicholas, S., Krissinel, E., Cowtan, K. & Noble, M. (2002). The CCP4 molecular-graphics project. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **58**, 1955–1957.
35. Wallace, A. C., Laskowski, R. A. & Thornton, J. M. (1995). LIGPLOT: a program to generate schematic diagrams of protein–ligand interactions. *Protein Eng.* **8**, 127–134.
36. Liang, J., Edelsbrunner, H. & Woodward, C. (1998). Anatomy of protein pockets and cavities: measurement of binding site geometry and implications for ligand design. *Protein Sci.* **7**, 1884–1897.
37. Thompson, J. D., Higgins, D. G. & Gibson, T. J. (1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* **22**, 4673–4680.
38. Leslie, A. G. W. (1992). Recent changes to the MOSFLM package for processing film and image plate data. *Joint CCP4/ESF-EACMB Newsl. Protein Crystallogr.* **26**, 27–33.
39. Collaborative Computational Project Number 4. (1994). The CCP4 suite: programs for protein crystallography. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **50**, 760–767.
40. Vagin, A. & Teplyakov, A. (2000). An approach to multi-copy search in molecular replacement. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **56**, 1622–1624.
41. McRee, D. E. (1999). XtalView/Xfit—a versatile program for manipulating atomic coordinates and electron density. *J. Struct. Biol.* **125**, 156–165.
42. Castronuovo, G., Elia, V., Fessas, D., Velleca, F. & Viscardi, G. (1996). *Carbohydr. Res.* **287**, 127–138.
43. Wyman, J. & Gill, S. J. (1990). Binding and Linkage, chapt. 2, pp. 49–59, University Science Books, Mill Valley, CA.
44. Gill, S. J., Robert, C. H. & Wyman, J. (1988). In *Biochemical Thermodynamics* (Jones, M. N., ed.), 2nd edit., chapt. 4, pp. 145–181, Elsevier Science Publishers B.V., Amsterdam.
45. Press, W. H., Flannery, B. P., Teukolsky, S. A. & Vetterling, W. T. (1989). *Numerical Recipes, The Art of Scientific Computing*, pp. 521–538, Cambridge University Press, Cambridge, UK.