Structural Requirements for Cooperativity in Ileal Bile Acid-binding Proteins

Serena Zanzoni, Michael Assfalg, Alejando Giorgetti, Mariapina D’Onofrio, and Henriette Molinari

From the Department of Biotechnology, University of Verona, 37134 Verona, Italy and the German Research School for Simulation Sciences, FZ-Jülich and RWTH Aachen, 52428 Jülich, Germany

Background: Ileal bile acid-binding proteins display different degrees of binding cooperativity. Molecular dynamics simulation studies of the wild type and mutant complexes and essential dynamics analysis of the trajectories supported the role of the identified amino acid residues as hot spot mediators of communication binding properties such as the degree of binding cooperativity. To investigate the molecular requirements for cooperativity we adopted a gain-of-function approach, exploring the possibility to turn a noncooperative chicken I-BABP (cI-BABP) into a cooperative mutant protein.

Results: To this aim we first solved the solution structure of cI-BABP in complex with two molecules of the physiological ligand glycochenodeoxycholate. The emerging picture is consistent with an extended conformational selection model of the binding mechanism that can be described as an extended conformational selection model.

Significance: Knowledge of the determinants of cooperativity in protein carriers is the key to understanding bile acid trafficking.

Intracellular bile acid-binding proteins (I-BABP), belonging to the family of intracellular lipid-binding proteins, control bile acid trafficking in enterocytes and participate in regulating the homeostasis of these cholesterol-derived metabolites. I-BABP orthologues share the same structural fold and are able to host up to two ligands in their large internal cavities. However variations in the primary sequences determine differences in binding properties such as the degree of binding cooperativity. To investigate the molecular requirements for cooperativity we adopted a gain-of-function approach, exploring the possibility to turn the noncooperative chicken I-BABP (cI-BABP) into a cooperative mutant protein.

To this aim we first solved the solution structure of cI-BABP in complex with two molecules of the physiological ligand glycochenodeoxycholate. A comparative structural analysis with closely related members of the same protein family provided the basis to design a double mutant (H99Q/A101S cI-BABP) capable of establishing a cooperative binding mechanism. Molecular dynamics simulation studies of the wild type and mutant complexes and essential dynamics analysis of the trajectories supported the role of the identified amino acid residues as hot spot mediators of communication between binding sites. The emerging picture is consistent with a binding mechanism that can be described as an extended conformational selection model.

The binding features of BABPs appear difficult to capture, although recent studies have set important milestones in the atomic level description of BABP-ligand interactions (7–10). It has been definitively ascertained that BABPs can form at least ternary complexes by hosting two ligand molecules inside a large protein cavity (11–14). One of the most intriguing features described for some members of the family relates to binding cooperativity, i.e. the energetic coupling between separate binding sites, a property that provides an additional level of signaling or regulation. Human ileal BABP (hI-BABP) and chicken liver BABP (cl-BABP) display an extraordinary degree of positive cooperativity associated with two internal binding sites, illustrated qualitatively by a Hill coefficient close to 2 (11, 15). The two proteins can thus effectively sequester ligands from the intracellular milieu, suggesting that they play a role in offering cytoprotection against high concentrations of bile acids. Cooperativity often arises from allosteric communication, a phenomenon that frequently escapes detection by common biochemical approaches (16). NMR and calorimetric studies performed on hI-BABP and on a series of mutants have attempted to identify the communication pathways between the two binding sites (9, 10). Further interaction studies reported for other members of the BABP family describe the binding stoichiometry and affinity of the rabbit ileal protein of small (14–15 kDa) β-barrel proteins that act as lipid chaperones (1–3). Specifically, BABPs are found abundantly expressed in the enterocytes and hepatocytes of various vertebrates and have been shown to play a pivotal role in the transcellular trafficking and enterohepatic circulation of bile salts (4). During the last decade a growing amount of evidence has been presented to show that the various FABPs that interact with lipids not only facilitate their transport in aqueous media but, by interacting with specific targets, also modulate their subsequent biological action or metabolism (5). Thus, structural and mechanistic knowledge of the interactions between lipids and their cognate binding proteins has become mandatory in understanding the action of lipids as signaling compounds and metabolic intermediates (6).

The binding features of BABPs appear difficult to capture, although recent studies have set important milestones in the atomic level description of BABP-ligand interactions (7–10). It has been definitively ascertained that BABPs can form at least ternary complexes by hosting two ligand molecules inside a large protein cavity (11–14). One of the most intriguing features described for some members of the family relates to binding cooperativity, i.e. the energetic coupling between separate binding sites, a property that provides an additional level of signaling or regulation. Human ileal BABP (hI-BABP) and chicken liver BABP (cl-BABP) display an extraordinary degree of positive cooperativity associated with two internal binding sites, illustrated qualitatively by a Hill coefficient close to 2 (11, 15). The two proteins can thus effectively sequester ligands from the intracellular milieu, suggesting that they play a role in offering cytoprotection against high concentrations of bile acids. Cooperativity often arises from allosteric communication, a phenomenon that frequently escapes detection by common biochemical approaches (16). NMR and calorimetric studies performed on hI-BABP and on a series of mutants have attempted to identify the communication pathways between the two binding sites (9, 10). Further interaction studies reported for other members of the BABP family describe the binding stoichiometry and affinity of the rabbit ileal protein
(r1-BABP) (17) as well as the binding thermodynamics and crystallographic structure of fully complexed zebrafish ileal protein (z1-BABP) (13).

Along this line of investigation, we reported previously on chicken I-BABP (cl-BABP) (18, 19). Isothermal titration calorimetry measurements supported by NMR titration experiments allowed us to establish a thermodynamic binding model describing two consecutive binding events. At variance with the human protein, a single bound cl-BABP was found to be relatively abundant at a low ligand/protein ratio, indicative of poor cooperativity. Indeed, a previous analysis of all available thermodynamic data for I-BABPs allowed us to place the proteins in an indicative cooperativity scale based on the estimated free energy of coupling ($\Delta G$) between the internal binding sites. The two extremes were defined by hl-BABP (maximum cooperativity, $\Delta G = -15.50 \text{ kJ/mol}^{-1}$) and cl-BABP (lowest cooperativity, $\Delta G = -0.87 \text{ kJ/mol}^{-1}$). In the case of cl-BABP, analysis of the energetics of binding and of chemical shift perturbations (19) further indicated that the first binding event triggers a global structural rearrangement associated with a substantial enthalpic contribution ($\Delta H = -50.3 \pm 5 \text{ kJ/mol}^{-1}$). The protein thus appears capable of establishing long-range interactions, although it displays poor energetic coupling between the binding sites. In line with the classical definition of cooperativity (20), we thus proposed that cl-BABP behaves as an allosteric system, making it a promising model for exploration of the determinants of cooperativity.

The goal of the present study was to gain more insight into the structural basis of positive cooperativity in I-BABPs through the comparative analysis of key interactions in proteins with different coupling energies. As a necessary starting point we determined the solution structure of cl-BABP in complex with two molecules of the physiologically prevalent bile acid glycochenodeoxycholic acid (GCDA). The structural details have been analyzed to guide the identification of key residues capable, upon mutation, of turning a low cooperative system into one displaying increased binding efficiency. The mutant analysis has the potential to single out functional interactions more directly than expected on the basis of structural comparisons between homologous proteins displaying a limited, but still significant, variability in amino acid composition. Single and double mutants were produced and ligand titration was monitored by NMR, as it was shown previously that NMR can capture the significant, variability in amino acid composition. Single and double mutants were produced and ligand titration was monitored by NMR, as it was shown previously that NMR can capture the significant variability in amino acid composition. The two extremes were defined by hl-BABP (maximum cooperativity, $\Delta G = -15.50 \text{ kJ/mol}^{-1}$) and cl-BABP (lowest cooperativity, $\Delta G = -0.87 \text{ kJ/mol}^{-1}$). In the case of cl-BABP, analysis of the energetics of binding and of chemical shift perturbations (19) further indicated that the first binding event triggers a global structural rearrangement associated with a substantial enthalpic contribution ($\Delta H = -50.3 \pm 5 \text{ kJ/mol}^{-1}$). The protein thus appears capable of establishing long-range interactions, although it displays poor energetic coupling between the binding sites. In line with the classical definition of cooperativity (20), we thus proposed that cl-BABP behaves as an allosteric system, making it a promising model for exploration of the determinants of cooperativity.

The goal of the present study was to gain more insight into the structural basis of positive cooperativity in I-BABPs through the comparative analysis of key interactions in proteins with different coupling energies. As a necessary starting point we determined the solution structure of cl-BABP in complex with two molecules of the physiologically prevalent bile acid glycochenodeoxycholic acid (GCDA). The structural details have been analyzed to guide the identification of key residues capable, upon mutation, of turning a low cooperative system into one displaying increased binding efficiency. The mutant analysis has the potential to single out functional interactions more directly than expected on the basis of structural comparisons between homologous proteins displaying a limited, but still significant, variability in amino acid composition. Single and double mutants were produced and ligand titration was monitored by NMR, as it was shown previously that NMR can capture the significant variability in amino acid composition. The two extremes were defined by hl-BABP (maximum cooperativity, $\Delta G = -15.50 \text{ kJ/mol}^{-1}$) and cl-BABP (lowest cooperativity, $\Delta G = -0.87 \text{ kJ/mol}^{-1}$). In the case of cl-BABP, analysis of the energetics of binding and of chemical shift perturbations (19) further indicated that the first binding event triggers a global structural rearrangement associated with a substantial enthalpic contribution ($\Delta H = -50.3 \pm 5 \text{ kJ/mol}^{-1}$). The protein thus appears capable of establishing long-range interactions, although it displays poor energetic coupling between the binding sites. In line with the classical definition of cooperativity (20), we thus proposed that cl-BABP behaves as an allosteric system, making it a promising model for exploration of the determinants of cooperativity.
Structure-Cooperativity Relationship in Ileal BABPs

**MD Simulations**—The starting coordinates for the wild type protein were from the calculated NMR structures. The H99Q/A101S mutant protein was modeled by point mutations of the residues using the Mutator plug-in in VMD (41). Both wild type and doubly mutated proteins were simulated in explicit aqueous solution inserted into a cubic box of water molecules, ensuring that the solvent shell would extend for at least 1.2 nm around them. Gromos96, in combination with the simple point charge (SPC) force field, was used for the simulation. Long-range electrostatic interactions were treated with the particle mesh Ewald (PME) method using a grid with a spacing of 0.12 nm (42). The cut-off radius for the Lennard-Jones interactions, as well as for the real part of the PME calculations, was set to 0.9 nm. The LINCS algorithm (43) was used to constrain all bond lengths involving hydrogen atoms, and the time step used was 2 fs. The systems were energy-minimized imposing harmonic position restraints of 1000 kJ mol\(^{-1}\) nm\(^{-2}\) on solute atoms, allowing the equilibration of the solvent without distorting the solution structure. After an energy minimization of the solvent and the solute without harmonic restraints, the temperature was gradually increased from 0 to 300 K in 200 ps of simulation. Each system was finally simulated for 50 ns. All simulations were performed with the GROMACS software package (44). The module g_cluster was used to clusterize the trajectory with an r.m.s.d. cut-off of 0.1 nm. The latter analysis was performed to extract the representative structures of the most populated clusters, that is, the most visited conformations along the entire length of MD simulations. The analysis of structural features was carried out on the representative structures. The large scale motions were investigated by calculating the eigenvectors of the covariance matrix of Cα atoms using the g_covar and g_anaeg modules of GROMACS. The residue-based root mean square fluctuations were calculated by projecting each trajectory on the first eigenvector. The hydrogen bonds formed during the simulations were extracted using the GROMACS program g_hbond.

**RESULTS**

Identification of a Singly Ligated cl-BABP—cl-BABP was shown able to bind at least two molecules of bile salt. However, previous NMR and thermodynamic data on cl-BABP (19) had predicted a considerable concentration of singly bound protein over a wide range of ligand/protein (L/P) molar ratios, accounting for up to 50% of all protein when \(1 < L/P < 2\). The structural analysis of the singly ligated species was expected to be very informative for understanding the sequence of binding events. Therefore, we performed an analysis of ligand resonances at various L/P values in order to identify the conditions suitable for such a characterization. At L/P = 2 ([supplemental Fig. S2]) an intense signal in the two-dimensional \(^{15}N\)-filtered NOESY experiment of a \([^{15}N,^{13}C]\)-cl-BABP-u-GCDA sample, connecting the resonances at 0.42 and 0.57 ppm, was assigned to an exchange cross-peak between C18 methyl groups of unbound and singly bound GCDA. This assignment was confirmed by the fact that on increasing ligand concentration this cross-peak disappeared. Complementary NMR experiments, namely \(F_1\)-edited, \(F_2\)-filtered, three-dimensional HMQC-NOESY (12) with mixing times ranging from 0 to 200 ms, were performed ([supplemental Fig. S3]), confirming the assignments.
of all the methyl groups of the doubly and singly ligated species. It was thus possible unambiguously to identify signals corresponding to ligand in the singly bound protein; however, the complexity of the spectra prevented a rigorous structural analysis, and we focused the study on the double bound species only.

Identification of Intermolecular Restraints for the Double Bound cI-BABP—cI-BABP forms a ternary complex with two ligand molecules (19). The major difficulty in the structure determination of the complex consists in the assignment of the two identically bound molecules of bile acids. The chemical shift dispersion of GCDA protons is very poor, but the presence of the three methyl groups, C18, C19, and C21 (see Scheme 1 for numbering), which appear in the unbound bile salt as two high intensity singlets (reported chemical shifts for C18 and C19 are 0.67 and 0.93 ppm, respectively) and as a doublet (C21, 0.95 ppm) resonating at high fields (45), constituted a good entry point for the assignment of the bound molecules. Two-dimensional F1/F2-15N,13C-filtered NOESY experiments performed on a [15N,13C]cI-BABP-u-GCDA sample with L/P = 4 were used to filter out 1H protein signals, allowing the observation of the GCDA resonances. Three high field resonating signals could be distinguished at 0.15, 0.19, and 0.42 ppm (Fig. 1). These resonances exhibited NOESY exchange peaks with intense signals at the chemical shift of free GCDA (in excess in the sample) and were assigned to the methyl groups of one bound GCDA molecule (hereby called GCDA-II). Specifically, the identification of C19 and C21, showing almost superimposed signals in the free ligand, was possible on the basis of a double filtered TOCSY experiment of the complex, where only the C21 methyl groups can display cross-peaks. If two mol-

cules of GCDA are bound to cI-BABP, then a total of six different methyl signals are expected in this region in addition to those of the free GCDA. The resonances of the ligand bound to site I (GCDA-I) resulted coincident with those of free GCDA based on the following observations: (i) The 1H-15N HSQC spectrum of u-cI-BABP-[15N]GCDA showed a single averaged resonance at 7.89 ppm for the amide group of free GCDA and GCDA-I, and (ii) the HACA and HNCA spectra of u-cI-BABP-[15N,13C]GCDA exhibited a single Cα chemical shift for the free and bound (to site I) ligand.

Few additional ligand resonances beyond those of the methyl groups were sufficiently resolved at L/P = 4 and could be unambiguously assigned, including the signals of H-7 (GCDA-I 3.85 ppm, GCDA-II 3.89 ppm) and H-3 of GCDA-I (3.46 ppm) as well as H-26 of both bound molecules (3.78 and 3.27 ppm for GCDA-I and GCDA-II, respectively). These assignments allowed the identification of 35 intermolecular NOEs (Fig. 2) used in the first run of structure calculations of the complex.

Structure Calculation of the Ternary Complex between cI-

BABP and Two Molecules of GCDA—First, the protein-only coordinates of the fully complexed species were determined on the basis of intramolecular experimental restraints. The computed family of conformers resulted of high quality r.m.s.d. 0.49 ± 0.08 Å for backbone atoms and 1.07 ± 0.08 Å for all heavy atoms (86% of the backbone angles in the most favored region and 14% in additionally favored regions of the Ramachandran plot with a G-factor for backbone/all dihedral angles of −0.62/−0.72). Subsequently, the ligand poses and conformations were determined with a data-driven docking approach (28, 46) based on the calculated structures. To identify the interaction surface, the average H3N and N chemical shift changes between the free and bound states (19) were analyzed. Resonances with significant chemical shift perturbations, deviating by more than twice the standard deviation from the mean chemical shift perturbation, included residues 53, 57−59, 97, and 98 residing in the CD loop and β-strand H. These residues were treated as ambiguous interaction restraints. Intermolecular NOEs, derived from edited-filtered experiments, were introduced as ambiguous interaction restraints in the first HADDOCK run. Subsequent rounds of the docking procedure coupled with spectral analysis led to the identification of 48 unambiguous restraints between the ligands and the protein from which the structure of the complex was derived. A complete list of ambiguous and unambiguous restraints is given in supplemental Table S1. The 20 lowest energy structures in the best energy score cluster were retained. The structural statistics of the cluster with the best average HADDOCK score are presented in

![Scheme 1. Glycochenodeoxycholic acid structure, showing numbering of carbon atoms.](image-url)
Table 1, and Fig. 3 displays the HADDOCK model that best represents the structure of the complex. The 20 best structures in the selected cluster have average r.m.s.d. values of 0.43 \pm 0.07 and 0.96 \pm 0.09 Å for the backbone and all heavy atoms, respectively. The average r.m.s.d. values of the ligands were 1.15 Å for GCDA-I and 0.72 Å for GCDA-II.

### Structural Correlates of Cooperativity: Analysis of I-BABP Subfamily

A phylogenetic analysis of the FABP family was performed as the starting point for a comparative functional analysis. The protein sequences closest to cI-BABP cluster in three well defined groups corresponding to liver FABPs, liver BABPs, and ileal BABPs (Fig. 4A). Thus the investigation of functional amino acids in cI-BABP was concentrated only on its closest relatives, the I-BABPs, including both mammalian and non-mammalian species.

Within the family of I-BABPs, only a few members have been studied in terms of ligand interaction properties. Some thermodynamic binding data have been presented in the literature for human (9, 11), rabbit (17), zebrafish (13), and chicken ileal BABPs (19), which allowed us to place these proteins in an indicative cooperativity scale as mentioned in the Introduction. Because it was expected that the most and the least cooperative proteins would present the largest functional differences, we focused on the structural comparison between cl-BABP (low cooperativity) and hi-BABP (high cooperativity). For the latter, only a single bound complex has been reported (Protein Data Bank code: 1O1V), as it was determined at temperatures at which one of the two ligands could not be observed because of exchange broadening of the NMR signals (47, 48). However, the protein scaffold can be predicted to represent the protein bound to two ligands with good approximation.

All of the available structural and binding data were used to point out the structural correlates of cooperativity according to the following approach. To identify functional residues, all the nonconserved amino acids pointing into the internal protein cavity were selected and mapped onto the sequence alignment plot, including the other homologous proteins (Fig. 4B). To point out the relevant interactions, pairs of amino acids were analyzed. The 11 nonconserved residues (residues 8, 24, 27, 30, 34, 53, 54, 59, 99, 101, and 114) within the protein cavity may give rise to 55 possible pairs. Among these pairs only eight were made of residues exhibiting a reciprocal average distance shorter than 4 Å and six (residues 53/54, 53/34, 54/34, 24/27, 27/30, and 30/34) were found in both human and chicken proteins (Table 2). Only the residue pairs 54/34 and 24/27 displayed changes in the amino acid type that correlated with cooperativity. The two pairs, residues 54/27 and 99/101, were present only in the human protein, with only the 99/101 pair displaying changes in amino acid type that correlated with cooperativity.
COMPARED WITH A LOSS-OF-FUNCTION EXPERIMENTAL SCHEME. WE THEREFORE SET OUT TO VERIFY THE ROLE OF THE ABOVE IDENTIFIED AMINO ACID PAIR 99/101 IN ESTABLISHING EFFICIENT BINDING COOPERATIVITY, WITH THE OPTIMAL INTERACTION REPRESENTED BY THE COUPLE Q/S. FOR THIS PURPOSE, TWO MUTANTS WERE DESIGNED AND ENGINEERED: THE SINGLE MUTANT A101S cI-BABP (SAME AMINO ACIDS AS IN RABBIT I-BABP) AND THE DOUBLE MUTANT H99Q/A101S cI-BABP (SAME AMINO ACIDS AS IN hI-BABP) TO REPRODUCE AN INTERMEDIATE AND HIGH COOPERATIVITY SYSTEM, RESPECTIVELY. THE STABILITY AND FOLD OF THE PROTEIN MUTANTS PRODUCED WERE VERIFIED (SUPPLEMENTAL FIG. S4). THE PROTEINS WERE TITRATED WITH INCREASING AMOUNTS OF [15N]GCDA, AND THE RECORDED 1H,15N HSQC SPECTRA WERE EMPLOYED TO MONITOR THE BINDING EVENTS. HERE WE EXPLOITED THE ALREADY ASSESSED POWER OF NMR TO GAIN SITE-SPECIFIC BINDING INFORMATION AND DETECT POSITIVE COOPERATIVITY IN THESE SYSTEMS (11, 15, 19). TITRATION EXPERIMENTS PERFORMED ON THE WILD TYPE PROTEIN SHOW THE APPEARANCE OF THE SIGNAL OF GCDA-II ($\delta^{14}N$) = 7.18 ppm, $\delta^{15}N$ = 112.0 ppm) ONLY AT L/P > 2, INDICATIVE OF LOW COOPERATIVITY (FIG. 5). FOR THE A101S MUTANT THE CORRESPONDING LIGAND SIGNAL IS ALSO NOT VISIBLE BEFORE L/P = 4, SUGGESTING THAT OTHER CONCURRING AMINO ACIDS MAY BE INVOLVED IN ESTABLISHING THE INTERMEDIATE ENERGETIC COUPLING VALUE REPORTED FOR hI-BABP. HOWEVER, IN THE CASE OF THE DOUBLE MUTANT H99Q/A101S, THE RESONANCES CORRESPONDING TO BOTH BOUND LIGANDS ARE ALREADY PRESENT AT LOW LIGAND CONCENTRATIONS (L/P \approx 1), AN INDICATION OF ACQUIRED POSITIVE BINDING COOPERATIVITY. THE ALTERED BINDING SITE OCCUPANCY IN THE MUTANT PROTEINS IS SHOWN QUANTITATIVELY IN Fig. 5D, REPORTING THE INTEGRATED VOLUMES OF THE PEAK OF GCDA-II ALONG THE TITRATIONS. THE STEPWISE OCCUPANCY OF BINDING SITE 1 COULD NOT BE MONITORED BECAUSE THE RESONANCE OF GCDA-I IS OVERLAPPED WITH THAT OF THE FREE LIGAND.

**Molecular Dynamics Simulations**—To gain a deeper understanding of the structural and dynamic basis of cooperativity in the systems under study, two 50-ns MD simulation runs were collected for the wild type and double mutant complexes. The calculated NMR structure provided the starting coordinates for wild type cI-BABP as well as for the mutant protein, with the simple substitution of the mutated amino acids. The analysis of the conformations sampled by the systems was performed by calculating the time evolution of the atom-positional r.m.s.d. from the initial structures for each protein. The r.m.s.d. values stabilized after 10 ns at around 0.15 nm from the initial structure in the case of the wild type protein and after 15 ns at around 0.2 nm in the case of the mutant (SUPPLEMENTAL FIG. S5), showing the substantial stability of the complexes under the simulation conditions. The overall conservation of structural properties was also apparent in the secondary structure analysis. No major variation could be detected, indicating the absence of large conformational changes or folding-unfolding events. The position of the ligands inside the binding cavity was checked continuously during the MD simulations as a further validation of the entire procedure. Indeed, the spatial restraints between the protein and the ligands, observed in the NMR experiments, were maintained (not violated) during the simulations.

A detailed inspection of hydrogen-bonding patterns along the trajectories highlighted interesting, distinctive features between the wild type and mutant proteins (SUPPLEMENTAL FIGS. S6 AND S7). In the former, a “lower region” H-bond network was established and maintained throughout the simulation involv-
ing Y97-H\textsubscript{1}, H99-N\textsubscript{1}, H99-He2, and E110-Oe1,e2, and E110-Oe1,e2 and R121-He. In this protein region only Tyr-97 is connected to GCDA-I, whereas no residue binds GCDA-II through the H-bonds (Fig. 6A). A very different pattern is found in the double mutant (Fig. 6B). During the simulation the hydroxyl group of Ser-101 forms hydrophilic contacts with Glu-110-Oe1,e2, Gln-99-He2, and OH-3 of GCDA-II. The presence of Ser-101 also decreases the probability of the formation of an H-bond between Gln-99 and Glu-110. Additional H-bonds are formed between Gln-99 and the OH groups of GCDA-II and between OH-3 of GCDA-I and both Tyr-97 and Arg-121. Ser-101 and Gln-99 thus provide an anchor region for GCDA-II in the lower part of the cavity that is not present in the wild type. A comparison of Fig. 6, A

B

Chicken I-BABP (FIN077) 1 AFTGK/TTS DENP/QF/KK MGLD/KIEM IGRG/KVTEV VQ/G/DPTW 50
Human I-BABP (PS1161) 1 AFTGK/TTS EKNT/QF/KK LGIS/DVIEK ARGK/IVTEV VQ/G/DPTW 50
Zebrasfish I-BABP (QGMM5) 1 AFTGK/TTS EGAT/FP/KK MGLD/DVIEK SRK/IKVTEY VQ/G/DPTW 50
Chicken I-BABP (FIN077) 51 QHHFGRT/TT NSPT/KEAD M Huntington ATN/MEGKLV AFGQP/YET 100
Human I-BABP (PS1161) 51 QHYSGRT/TT NSPT/KEESN ICG/MEGKLV ATV/MPPYET 100
Zebrasfish I-BABP (QGMM5) 51 QYYFNNVHT NSPT/KEESD M Huntington ATN/MEGKLV ATV/MPPYET 100
Chicken I-BABP (FIN077) 101 TIEG/TKLV ISTEGWY----- YRVE/SKLA 127
Human I-BABP (PS1161) 101 TIEG/TKLV ISTEGWY----- YRVE/SKLA 127
Zebrasfish I-BABP (QGMM5) 101 TIEG/TKLV ISTEGWY----- YRVE/SKLA 127

**TABLE 2**

Pairs of nonconserved residues exhibiting distances shorter than 4 Å in chicken and human I-BABPs

| Residue pair | cI-BABP residues | zI-BABP residues | rI-BABP residues | hI-BABP residues |
|--------------|-----------------|-----------------|-----------------|-----------------|
| 53/54        | Phe/Pro         | Tyr/Pro         | Tyr/Ser         | Tyr/Ser         |
| 53/34        | Phe/Cys         | Tyr/Phe         | Tyr/Ile         | Tyr/Ile         |
| 54/34        | Pro/Cys         | Pro/Phe         | Ser/Ile         | Ser/Phe         |
| 24/27        | Pro/Lys         | Pro/Val         | Pro/Val         | Ser/Val         |
| 27/30        | Lys/Met         | Val/Lys         | Val/Lys         | Val/Lys         |
| 30/34        | Met/Cys         | Lys/Phe         | Lys/Ile         | Lys/Phe         |
| 54/27\textsuperscript{a} | Pro/Lys | Pro/Val | Ser/Val | Ser/Val |
| 99/10\textsuperscript{a} | His/Ala | Gln/Thr | His/Ser | Gln/Ser |
| $\Delta G^{\circ}$ (kJ mol$^{-1}$) | $-0.87$ | $-4.55$ | $-6.96$ | $-15.50$ |

\textsuperscript{a} Residence pair displaying interactions only in human I-BABP.

\textsuperscript{b} The $\Delta G$ values are the upper limits of the free energy coupling between the two binding events and provide a measure of cooperativity (19).
and B, clearly shows that only in the double mutant is a continuous pattern of hydrophilic interactions established throughout the protein, involving the two bound ligands in the communications pathway.

A further analysis of MD trajectories was performed through essential dynamics to identify the relevant, low frequency, concerted motions of groups of residues (49), which can be appreciated by residue-based root mean square fluctuations (supplemental Fig. S8, RMSF) projected along the directions of the eigenvectors. A significant difference between both holosystems was observed; in the wild type protein a concerted fluctuation is apparent only for the CD loop and, to a lesser extent, for the H9251I-H9251II and EF loops, both being in contact with the carboxyl tails of the ligands in the NMR structure. On the contrary, for the double mutant, correlated motions are transmitted throughout the protein segments defining the protein open end, namely the H9251I-H9251II, CD, GH, and IJ loops. Additional concerted motions are present at the closed end of the protein at the level of the BC loop.

**DISCUSSION**

**Structural Comparison with Related Protein Complexes**—High resolution structures of BABP-bile salt complexes are still very few. The only ternary complex of an I-BABP available for comparative analysis of the determined structure herein is that of zI-BABP bound to two cholate molecules (Protein Data Bank...
code: 3EM0 (13). Superposition of the structures of zl-BABP and cl-BABP yielded r.m.s.d. values of 1.72 Å and indicated substantial differences localized to helix-II, loops EF and IJ, all defining the protein open end. The binding sites appeared to be almost coincident, with the internal ligand essentially superimposable after the protein coordinates were matched, whereas the ligands in site I were twisted reciprocally by ~90° around the longest molecular axis.

The binding cavities of the two proteins were analyzed with LIGPLOT, a software that generates two-dimensional schematic diagrams of protein-ligand interactions, highlighting both hydrophilic and hydrophobic contacts. A diagram representing the interactions of cl-BABP with two GCDA molecules is shown in Fig. 7. In the zebrafish holoprotein the bound ligands are cholate molecules (CA), presenting shorter side chains compared with GCDA as well as an additional hydroxyl group in position 12. In the two proteins, the internal ligand establishes interactions via its carboxylate group with the EF loop (residues Thr-73 and Arg-77 in chicken and residue Val-74 in zebrafish protein). Additional hydrophilic interactions are established in the zebrafish protein between OH-7 and Glu-110 and OH-12 and Tyr-97. The hydrophobic interactions of the internal ligand are similar in the two proteins, although a more extended hydrophobic environment is present in cl-BABP, attendant with the increased hydrophobicity of GCDA compared with CA.

The more external ligand, GCDA-I, in chicken protein, makes two H-bonds between its C24 carbonyl group and the side chain of Lys-27, its terminal carboxylate and the backbone amide of Arg-57. In zebrafish, H-bonds are established between OH-7 and Tyr-53 and OH-12 and Tyr-14. Therefore, the main difference between the two proteins is that the ligand hydro-
philic anchoring moiety is constituted by the sterol rings in zl-BABP and by the linear chain in cl-BABP. Hydrophobic contacts with the external cholate in zl-BABP appear more limited compared with cl-BABP.

In summary, the binding site geometries of the two bile salt complexes appear relatively similar. The differences, however, relate to the patterns of hydrophilic interactions involving the more externally bound ligand as well as to an increased number of hydrophobic contacts observed for the chicken protein.

**Structure-Cooperativity Relationship and Mechanistic Implications**—In I-BABPs, residues 99 and 101 are buried within the protein cavity. The pair 99/101, with the exception of the chicken protein, is always constituted by polar amino acids: Gln/Thr (zl-BABP), His/Ser (rl-BABP), and Gln/Ser (hl-BABP). It is interesting to note that in the human protein, five residues (Tyr-97, Gln-99, Ser-101, Glu-110, and Arg-121) constitute a connected platform where H-bonds are formed between the pairs 97/99, 99/101, and 110/121. It is possible that residues Ser-101 and Tyr-97 keep Gln-99 in a favorable position to produce a polar cavity capable of anchoring one ligand. In the chicken protein, which lacks a polar residue at position 101, His-99 establishes a very stable H-bond with Tyr-97 and Glu-110, which in turn is connected with Arg-121 through an H-bond. Thus a very rigid H-bond pattern is formed, possibly disfavoring the conformational flexibility needed for an efficient coupling between the binding subcavities. The nature and orientation of residue 99 may thus have an effect on the binding mechanism. Indeed, calorimetry data coupled to mutation experiments reported for the human protein indicate that the mutant Q99A exhibits a strong decrease in cooperativity for binding to GCDA (9).

The functional relevance of the amino acid couple 99/101 is supported by both the structural and dynamic information derived from MD simulations. A comparison between representative structures of the computed trajectories of the wild type and H99Q/A101S proteins highlights the fact that in the double mutant a rearranged conformational state is stabilized (supplemental Fig. S9) displaying a closure of the open end that can be ascribed to a positional change of the CD, GH, and IJ loops. Importantly, in this structure the mutated residues allow an extension of the H-bond hyper-network, increasing the communication between binding sites (Fig. 6). This notion supports a long-range effect on ligand binding instead of a local affinity change. In addition to these structural (mainly enthalpic) contributions to binding affinity, differential protein flexibility may also play a role in ligand recognition (entropic contribution). To investigate this possibility, essential dynamics was carried out on the MD trajectories of the complexes. This analysis indicated that in the wild type protein the CD loop undergoes the largest fluctuation, with correlated motions substantially confined to loops that are in contact with the carboxyl tails of the two ligands (supplemental Fig. S8). The double mutant complex instead shows fluctuations of minor amplitude, however, spanning a larger protein domain, extending beyond residues in contact with the ligands, and defined by loops in the so-called portal area. It is interesting to note that these loops connect strands involved in the major H-bond network, whereas the EF loop, connecting strands E and F, is excluded from the main correlated motions. A functional consequence of the described equilibrium fluctuations could be the stabilization of a "closed" structural architecture characterized by shorter average distances between the interhelix and CD loops.

Altogether, our observations establish the involvement of specific amino acids in the reorganization of the protein structure that occurs upon ligand binding. In our previous work (19) we described cl-BABP as allosterically activated by the first ligand binding event. Indeed, NMR chemical shifts of the unbound and fully complexed proteins revealed sensitive indicators of atom positional changes, which are presumably small but distributed over the entire protein, suggestive of a global structural rearrangement. It was proposed previously that, in BABPs, allostery derives from conformational selection (50–52), and it was further hypothesized that positive cooperativity in hl-BABP may be related to a slow conformational change of the protein, occurring after the second binding step (10). The acquisition observed herein of an extended H-bond network in the double mutant cl-BABP represents a final binding event that could be described according to an "induces fit" model.

In summary, the cooperative binding mechanism in ileal BABPs requires the presence of a small number of latch residues located in the inner protein cavity that "click in" to their final conformation upon ligand binding and stabilize an extended communication network necessary for establishing energetic coupling between binding sites. The available data fit very well with the proposed extended conformational selection model (21), described as an initial conformational selection where ligands bind one of the several fluctuating protein conformers and induce a subsequent conformational rearrangement of the protein. Binding cooperativity can be particularly important for the ileal BABP family as it can provide ability to function as an on-off switch by responding to changes in bile acids concentration and protecting the cell against their toxic effects (53, 54). Interestingly, I-BABPs have been shown to display different degrees of cooperativity, which may correlate with species-specific differences in the bile acid pool size. We have demonstrated that key interactions established by amino acids specific to the human protein can be introduced artificially in cl-BABP to gain energetic coupling between binding sites. The determined correlates of cooperativity provide a structural perspective on the functional differences among BABPs and may be of value for the rational design of synthetic drugs aimed at controlling bile acid circulation.

**Acknowledgments**—We thank Simona Tomaselli for help with structure calculations. We thank the European Community (EU-NMR Contract RII3-026145) for access to the HWB-NMR facility at Birmingham, United Kingdom; the Consorzio Interuniversitario di Risonanze Magnetiche di Metalloproteine Paramagnetiche (CIRMMP) for access to high field spectrometers; the e-NMR project (a European FP7 e-Infrastructure grant, Contract 213010), supported by the national GRID Initiatives of Italy, Germany, and the Dutch BiG Grid project (Netherlands Organization for Scientific Research) for the use of Web server facilities; and the University of Verona and the Cariverona Foundation for financial support for the acquisition of an NMR spectrometer and cryoprobe.
