The Structure of Crystalline *Escherichia coli*-derived Rat Intestinal Fatty Acid-binding Protein at 2.5-Å Resolution*

(Received for publication, October 15, 1987)

James C. Sacchettini‡‡, Jeffrey I. Gordon‡‡‡, and Leonard J. Banaszak‡

From the Departments of Biological Chemistry and Medicine, Washington University School of Medicine, St. Louis, Missouri 63110

Rat intestinal fatty acid-binding protein (I-FABP) is an abundant cytoplasmic protein which is synthesized in the small intestinal lining cell where it is thought to participate in the absorption and intracellular metabolism of fatty acids. Each mole of this 132-residue polypeptide binds 1 mol of long chain fatty acid in a noncovalent fashion. Because of its small size and single ligand-binding site, I-FABP represents an attractive model for defining the molecular details of long chain fatty acid-protein interactions. The structure of *Escherichia coli*-derived rat I-FABP has now been solved to 2.5 Å resolution using three isomorphous heavy atom derivatives. The protein consists of 10 anti-parallel β-strands present as two orthogonal β-sheets. Together a "clam shell-like" structure is formed with an opening located between two β-strands and an interior that is lined with the side chains of nonpolar amino acids. The bound fatty acid ligand is located in the interior of the protein and has a bent conformation, possibly reflecting the presence of several gauche bonds in the hydrocarbon tail. Our present interpretation of the electron density map suggests that the fatty acid is oriented with its carboxylate group facing the guanidinium group of Arg127, whereas the end of its hydrocarbon tail is in close proximity to Val108. The indole side chain of Trp106 forms the molecular framework around which the principal bend of the hydrocarbon chain occurs.

The small intestinal lining cell or enterocyte contains a 15-kDa polypeptide which binds long chain fatty acids. In the adult rat, intestinal fatty acid-binding protein (I-FABP) represents 2-3% of the enterocyte's cytoplasmic protein mass (1). The primary structure of this 132-residue polypeptide has been deduced from an analysis of cloned cDNAs (2, 3). I-FABP is one of three homologous FABPs which have been isolated and characterized to date; the others are known as liver and heart FABPs (4, 5). Their distinct structures and patterns of tissue-specific and developmental expression suggest that they may fulfill different functions in their cells of origin (6). Proposed physiologic roles for the FABPs include (i) facilitating the uptake of long chain fatty acids into cells and their transport to organelles or to specific metabolic pathways and (ii) protecting cell membranes and enzymes from the potentially toxic, detergent effects of high concentrations of fatty acids and their acyl-CoA derivatives (reviewed in Ref. 7). The FABPs are members of a protein family presently consisting of nine similar-sized proteins which have evolved to bind other hydrophobic ligands including retinoids, e.g. cellular retinol-binding protein (8), cellular retinol-binding protein II (9), and cellular retinoic acid-binding protein (10). Genes encoding family members are organized in a similar fashion: 4 exons divided by 3 comparably placed introns (11-14).

The mechanism by which proteins bind long chain fatty acids has yet to be defined at an atomic level. For example, the tertiary structure of serum albumin, the major extracellular long chain fatty acid-binding protein, has not been described. There are several reasons why the FABPs in general, and I-FABP in particular, represent an attractive model system for examining long chain fatty acid-protein interactions. The FABPs are small proteins of 127-132 amino acids. I-FABP contains a single ligand binding site that binds long chain fatty acids (C16-20) saturated and unsaturated fatty acids with *Kₐ* values of 1-4 μM (3). Rat I-FABP has been expressed in *Escherichia coli*, purified to homogeneity and crystallized (15). Crystalline I-FABP contains ~1 mol of ligand/mol protein (3, 15). Fatty acid analysis of the *E. coli*-derived protein prior to crystallization indicated (3) that only saturated fatty acids were bound, and that the major species was palmitate (C16:0). We have now determined the tertiary structure of *E. coli*-derived rat I-FABP to 2.5 Å resolution. Our model permits us to define several details of long chain fatty acid-protein interaction.

**EXPERIMENTAL PROCEDURES**

**Purification and Crystallization of Rat I-FABP from *E. coli***—Procedures used for the purification and crystallization of *E. coli*-derived rat I-FABP have been previously described (15). Two major modifications have improved the quality of the protein crystals: the crystallization buffer was changed to 0.1 M PIPES, pH 7.3, and the free interface diffusion method was used instead of the hanging drop vapor diffusion method. Crystals were routinely obtained in excess of 0.3 X 0.3 X 1.5 mm (space group P2₁, cell dimensions a = 36.9 Å, b = 56.8 Å, c = 31.8 Å, β = 114.0°).

**Preparation and Analysis of Heavy Atom Derivatives**—Heavy atom searches were conducted in 0.1 M PIPES, pH 7.3. Native crystalline data and potential heavy atom derivative data were collected to 2.5 Å resolution on an Enraf Nonius CAD4 diffractometer. Crystals were soaked in solutions of ethylmercurithiosalicylic acid (2 mM), K₃PtCl₆ (2 mM), trimethyl lead acetate (10 mM), and p-hydroxymercuribenzoic acid (10 mM). Diffraction data from these crystals indicated that only ethylmercurithiosalicylic acid failed to yield an inter-
The tertiary structure of rat intestinal fatty acid-binding protein (I-FABP) has now been determined to a resolution of 2.5 Å using the methods described under "Experimental Procedures." Fig. 1A presents the structure in schematic form. The protein is composed of 10 anti-parallel β-strands which are labeled β A–J. All 10 strands are in the "up-and-down" motif in a +1, +1, +1, +1, +1, +1, +1 topology (19). Most connecting turns between neighboring β-strands contain glycine and/or asparagine residues. This is typical for sharp turns (19). Two α helices, labeled α-I and α-II in Fig. 1A, are present near the NH₂ terminus of crystalline I-FABP and are also interconnected by a sharp turn so that their helical axes are nearly parallel. They include residues 14–23 and 26–34.

Fig. 1B presents a cartoon diagram of rat I-FABP. The β-strands of I-FABP form a "clam shell-like" structure. The molecule contains two β-sheets arranged such that their strands are oriented nearly orthogonal to one another. The first β-sheet or half-sheet is composed of strands A–F and the second sheet, strands F–J. With the exception of strands D and E, all strands are separated by distances suggesting a pattern of hydrogen bonding typically seen in antiparallel β-sheets. Whereas strands D and E form hydrogen bonds with their adjacent strands C and F, there is a ~3 Å gap between them. This is visible both in Fig. 1B and in the stereo-diagram contained in Fig. 2. The first few residues of strand E are within hydrogen-bonding distance to strand F on one side and are exposed to solvent on the other. The last few residues of strand F are within hydrogen-bonding distance to strand G on one side and face the sharp turns between strands B and C and strands D and E on the other side. Strand F contains a dramatic bend when compared to the other β-strands in the protein and appears to form a "seam" between the two component β-sheets. A portion of strand A also contributes to this seam between the two β-sheets.
demnstrates how helix α-I appears to be an extension of strand A, following the overall direction of this β-strand. A similar continuity is found between α-II and β-strand B.

The gap between strands D and E forms an opening in the polypeptide shell of the molecule. This opening is defined at 30 Å. The location of the 3 introns of the I-FABP gene are indicated in this structure. The 24 amino acid differences between the rat and human proteins are also shown.

Fig. 1. Schematic drawing of E. coli-derived rat I-FABP. Panel A illustrates the principal secondary structural features of the protein. The 10 anti-parallel β-strands are labeled A through J, whereas the two α-helices are indicated as I and II. Note the relatively wide spacing between strands D and E. Residues are numbered according to the primary translation product of I-FABP mRNA. The initiator methionine residue is removed by an aminopeptidase activity present in E. coli and therefore, the crystalline protein contains residues 2-132 (3). Panel B is a ribbon diagram of crystalline I-FABP defined at 2.5 Å. The location of the 3 introns of the I-FABP gene are indicated in this structure. The 24 amino acid differences between the rat and human proteins are also shown.

introduced between carbons 7 and 8 (g*), 9 and 10 (g*), and 13 and 14 (g*) of palmitate (see Fig. 2 and Fig. 3A).

Energy barriers for rotation from a trans to a gauche* or from a trans to a gauche+ conformation are only of the order of 3.5 kcal/mol for unbound fatty acids. As a result, at room temperature thermal energy is sufficient to result in rapid rotation from one staggered conformation to another (20). The bent conformation of the bound saturated fatty acid is not unfavorable from a thermodynamic standpoint and appears to be the result of interactions between this ligand and the amino acid side chains described above.

The side chains of a number of residues are in close proximity to the electron density which we are ascribing to the fatty acid. These include: Trp83, Phe103, Tyr157,171,172,172, and Val118. Several of these fitted side chains, Phe103, Val118, are visible in A of Fig. 3. This figure shows that the indole side chain of Trp83 is partially surrounded by the principal bend of the hydrocarbon chain belonging to the bound fatty acid. Panel B of Fig. 3 shows the relationship between this indole side chain and the bend in the fatty acid from a different perspective. Last of all, the guanidinium group of Arg127 is within ~4 Å of the electron density assigned to one end of the bound fatty acid, whereas Val118 is near the other end. These two observations aided in determining the orientation of the fatty acid. Arg127 appears to be the only basic amino acid whose side chain extends into the predominately hydrophobic core of I-FABP near this bound ligand and is not easily accessible to solvent. When the primary structures of the eight members of the FABP “family” are aligned using the ALIGN algorithm (13), it is apparent that an arginine residue is present at this comparable position in all family members except the cellular retinol-binding proteins which have a Gln residue at this location.

In view of these observations, a number of conclusions regarding fatty acid binding to I-FABP can be made. First, it appears likely that the fatty acid is bound in an ionized form with the carboxylate group facing the guanidinium group of Arg127. Electrostatic charge interactions between these two groups probably contribute to the binding energy. Second, the side chains of the aromatic and hydrophobic amino acids which form the fatty acid-binding site probably stabilize the hydrocarbon tail in a bent conformation. Third, we have previously shown that E. coli-derived I-FABP binds both long chain saturated and unsaturated fatty acids with equal affinity in vitro (3). This appears to be possible because the binding site accommodates a bent conformation which is normally associated with unsaturated fatty acids. Fourth, the β-sheet structure contains an opening or interruption in the main chain hydrogen bonding pattern. This separation between strands D and E may be the location for entry of the fatty acid into the binding site.

It is unknown what role the fatty acid ligand plays in defining the conformation of rat I-FABP. Furthermore, higher resolution studies and refinement will be necessary to define the exact orientation and conformation of the bound fatty acid and how the side chains of amino acids bordering the ligand interact with it. If these studies support the present model of ligand binding, it will be interesting to determine if, or how, this protein accommodates other fatty acids with less flexible structures.

Sequence Comparisons—The amino acid sequences of rat and human I-FABP are highly conserved: 108 of 132 residues are identical (14). There have been no insertions or deletions in these homologous proteins (21). Inspection of the model of E. coli-derived rat I-FABP presented in Fig. 1B indicates that these 24 substitutions should be easily accommodated. Fur—
Tertiary Structure of Rat Intestinal Fatty Acid-binding Protein

FIG. 2. Stereo view of the α-carbon model of I-FABP with a model fatty acid ligand. The 9 by 30 Å opening, bounded by β-strands D and E, and helix α-II, is highlighted by the bolder lines. Note that the 10 antiparallel β-strands are organized as two orthogonal β-sheets. At 2.5 Å resolution, the best fit to the electron density attributed to the fatty acid ligand was obtained when a C16:0 fatty acid with gauche+ bonds between carbons 7 and 8, 9 and 10, and 13 and 14 was used. The rationale for orienting the fatty acid with its carboxylate group facing towards Arg'* and its bent hydrocarbon tail towards the interior of the molecule is discussed in the text. The α-carbon coordinates of I-FABP have been deposited in the Brookhaven Protein Data Bank.

FIG. 3. Interactions between I-FABP and its ligand. A illustrates the bent conformation of the electron density attributed to the fatty acid ligand. Note that the magnitude of the electron density associated with the fatty acid is similar to that of the side chains of several closely situated aromatic and hydrophobic amino acids (e.g. TrpW). A model of palmitate with g+ bonds between carbons 7-8, 9-10, and 13-14 has been superimposed upon this density. B shows another view of the tubular electron density at the major bend in the fatty acid chain and its proximity to the indole group of TrpW (indicated by the arrow).

thermore, none of the 24 amino acids listed in Fig. 1B appear critical for fatty acid binding.

The human and rat I-FABP genes have 4 exons which encode residues 1–22, 23–80, 81–116, and 117–132, respectively (14). As shown in Fig. 1B, Exon 1 specifies β-strand A and helix I. Exon 2 encodes helix II and β-strands B, C, D, E, and the first half of strand F. Exon 3 encodes the last half of strand F through the first few residues of strand I. Exon 4 gives rise to the remainder of strand I and strand J. Intron 2 occurs just after the codon specifying Thr80. Thr80 occurs in the middle of strand F, the strand that is shared by the two β-sheets of I-FABP. This location is compatible with the hypothesis that the present mammalian I-FABP gene represents the results of duplication of an ancestral sequence (22). Comparisons of the primary structures of other members of the FABP family have indicated that their NH₂-terminal halves are relatively more conserved than their COOH-terminal halves (10, 11, 15). Our analysis of the tertiary structure of E. coli-derived I-FABP indicates that the fatty acid ligand forms its key contacts with the side chains of residues encoded by exons 3 and 4. This is consistent with the notion that the sequence divergence in exons 3 and 4 may be responsible for the different hydrophobic ligand-binding characteristics of family members (11, 15). Given the high degree of primary sequence similarity between I-FABP and the other members of this protein family, it is likely that they will have tertiary structures which resemble I-FABP. Detailed comparison of the topologic features of this group of proteins, e.g. I-FABP and cellular retinol binding protein II (9), should provide insights about the structural basis of their different ligand binding characteristics.

The tertiary structures of several other hydrophobic ligand-binding proteins have recently been defined. These include serum retinol-binding protein (RBP), a 182-residue polypeptide which binds all trans-retinol (23), insecticyanin, or bilin-binding protein, a 189-amino acid-long protein which binds the open-chain tetrapyrrole biliverdin (24, 25) and bovine milk β-lactoglobulin (26). Whereas optimal alignments of the
primary structures of these three proteins failed to disclose an impressive degree of sequence similarity, crystallographic studies have shown that they have a remarkable degree of conformational homology or topologic equivalence (23-27). All possess a characteristic β-barrel composed of eight antiparallel β-strands. As in I-FABP, these strands are contained in two nearly orthogonally oriented β-sheets which surround their hydrophobic ligands. Holden et al. (25) noted that biliverdin is located towards the “mouth” of the β-barrel of tobacco hornworm insecticyanin, whereas its propionate side chains are oriented towards solvent, the net result being that this ligand assumes a somewhat bent conformation. The β-barrel of RBP is closed to solvent at one end and open at the other. Newcomer et al. (23) found that all trans-retinol sits along the axis of RBP’s barrel with its isoprene tail close to solvent and its β-ionone ring in the interior. Similarities in the topology of RBP, bilin-binding protein, and β-lactoglobulin were noted after superimposition of their α-carbon coordinates (24, 26, 27), suggesting that this structural motif has been adapted by a variety of proteins which bind and/or transport small hydrophobic ligands (27). Neither an evolutionary nor a structural relationship between RBP and members of the I-FABP family was fully established (13) when their primary structures were compared using the program RELATE and a mutation data matrix scoring system (28, 29). Such a relationship now becomes apparent when viewed in the context of tertiary structures. 1 I-FABP, bilin-binding protein, and RBP can thus be included in an enlarging group of related hydrophobic ligand-binding proteins (30, 31) that exhibit approximately the same conformation.

Acknowledgments—We thank Dianne Stockhausen and Thomas Meininger for their expert technical assistance. We are indebted to Hazel Holmen (Department of Biochemistry, University of Arizona) and Steven Roderick (University of Oregon) for their numerous suggestions concerning potential heavy derivatives and other unpublished results plus Edwin Westbrook and Mary Deacon of the Midwest Area Detector Facility at Argonne National Laboratory for their help in data collection and processing. We also thank C. Cambillau (Centre National de la Recherche Scientifique, Marseille, France) for the use of his version of the model building program, TOM.

REFERENCES


*Huber et al. (24) and Horden et al. (25) have independently compared the tertiary structures of insecticyanin and serum retinol-binding protein by superimposing their α-carbon coordinates and noted their similar conformations. We have superimposed the α-carbon coordinates of E. coli-derived rat I-FABP with those of insecticyanin (kindly provided by Hazel Holmen) to establish their topological similarities.