Crystallization, Structure Determination and Least-squares Refinement to 1.75 Å Resolution of the Fatty-acid-binding Protein Isolated from Manduca sexta L.

Matthew M. Benning¹, Alan F. Smith², Michael A. Wells²
and Hazel M. Holden¹†

¹Institute for Enzyme Research, Graduate School and Department of Biochemistry University of Wisconsin, Madison, WI 53705, U.S.A.
²Department of Biochemistry
University of Arizona, Tucson, AZ 85721, U.S.A.

(Received 6 April 1992; accepted 6 July 1992)

The molecular structure of an insect fatty-acid-binding protein isolated from Manduca sexta L. has been determined and refined to a nominal resolution of 1.75 Å. Crystals used in the investigation were grown from 1.6 M-ammonium sulfate solutions buffered at pH 4.5 with 50 mM-sodium succinate, and belonged to space group P2₁ with unit cell dimensions of a = 27.5 Å, b = 71.0 Å, c = 28.7 Å and β = 90.8°. An electron density map, phased with four heavy-atom derivatives and calculated to 2.5 Å resolution, allowed for complete tracing of the 131 amino acid residue polypeptide chain. Subsequent least-squares refinement of the model reduced the R-factor from 46.0% to 17.3% using all measured X-ray data from 30.0 Å to 1.75 Å. Approximately 92% of the amino acid residues fall into classical secondary structural elements including ten strands of anti-parallel β-pleated sheet, two α-helices, one type I turn, three type II turns, four type II' turns and one type III turn. As in other fatty-acid-binding proteins, the overall molecular architecture of the insect molecule consists of ten strands of anti-parallel β-pleated sheet forming two layers that are nearly orthogonal to one another. A helix-turn-helix motif at the N-terminal portion of the protein flanks one side of the up-and-down β-barrel. The functional group of the fatty acid is within hydrogen-bonding distance of Gln39, Tyr129, Arg127 and a sulfate molecule, while the aliphatic portion of the ligand is surrounded by hydrophobic amino acid residues lining the β-barrel. The binding of the carboxylic acid portion of the ligand is very similar to that observed in P2 myelin protein and the murine adipocyte lipid-binding protein, but the positioning of the hydrocarbon tail after approximately C6 is completely different.

Keywords: protein structure; X-ray crystallography; fatty-acid-binding proteins; lipid transport; insect proteins

1. Introduction

In recent years it has become increasingly apparent that insects can provide valuable information regarding lipid metabolism in general, and can serve as ideal model systems for lipid-transport protein structure and function studies. The advantages of working with insect systems are indeed numerous (Law & Wells, 1988). For example, many insect species can be reared in the large numbers required for biochemical studies with less laboratory maintenance than their vertebrate counterparts. In addition, their tissues tend to be less fragile, there is vast variation among these animals, and many have reasonably short life cycles. Also of importance is the fact that experiments with insects are not subject to the limiting regulations imposed on vertebrate systems.

The validity of insects as models for structural investigations of lipid-protein interactions has been demonstrated within the last five years by the X-ray crystallographic studies of insecticyanin isolated from the tobacco hornworm, Manduca sexta L. (Holden et al., 1987), bilin-binding protein obtained from Pieris brassicae (Huber et al., 1987), and apolipoporphin-III purified from the African migratory locust, Locusta migratoria (Breiter et al., 1991). In the case of insecticyanin and bilin-binding...
protein, both are involved in the transport of the \( \gamma \)-isomer of biliverdin IX and both contain eight strands of anti-parallel \( \beta \)-pleated sheet forming a barrel that is flanked on one side by an \( \alpha \)-helix of approximately 4.5 turns. While this type of three-dimensional architecture had already been observed in the X-ray models of human serum retinol-binding protein (Newcomer et al., 1984; Cowan et al., 1990) and bovine \( \beta \)-lactoglobulin (Sawyer et al., 1985; Monaco et al., 1987), it was the structure determinations of the insect proteins that suggested this fold to be more common than was once anticipated. By contrast, until the three-dimensional structure of the insect apolipoprotein-I had been determined, there had been no direct visualization of an apolipoprotein. Subsequent X-ray crystallographic analyses of a fragment of apolipoprotein E, however, revealed that the human molecule has a very similar structural motif of an up-and-down helical bundle as observed in the insect protein (Wilson et al., 1991).

Like any organism with a circulatory system, insects must possess a mechanism by which long-chain fatty acids can be safely transported throughout both intra- and extracellular compartments. Cytosolic fatty-acid-binding proteins appropriate for this purpose have recently been identified in the flight muscle of the migratory locust, Schistocerca gregaria (Haunerland & Chisholm, 1990), and from the midgut of Manduca sexta L. (Smith et al., 1992). In the midgut cytosol of Manduca sexta L., there are two abundant fatty-acid-binding proteins referred to as MFB1 and MFB2. They are 55% identical with respect to amino acid sequence and are restricted to the midgut (Smith et al., 1992). MFB1 is found predominantly in the anterior two-thirds while MFB2 is located in the posterior two-thirds of the midgut. Both have relative molecular masses of approximately 14,000 and both have been demonstrated to contain bound fatty acids in a 1:1 molar ratio. Interestingly, in the MFB2 molecule there is a lysine residue that is typically an arginine residue in other fatty-acid-binding proteins (Smith et al., 1992).

Here we describe the crystallization, structure determination and refinement to a nominal resolution of 1.75 Å of MFB2 (1 Å = 0.1 nm). At present, the three-dimensional structures of five fatty-acid-binding proteins from vertebrate sources have already been determined to various resolutions: the fatty-acid-binding protein from rat intestine (Sacchettini et al., 1988), P2 protein from the bovine peripheral nervous system (Jones et al., 1988), chicken liver basic fatty-acid-binding protein (Scapin et al., 1990), the fatty-acid-binding protein from bovine heart muscle (Müller-Fahrnow et al., 1991) and the murine adipocyte lipid-binding protein (Xu et al., 1992). All five proteins contain ten strands of anti-parallel \( \beta \)-pleated sheet and a helix-turn–helix motif. The long chain fatty acid binds in the interior of the up-and-down \( \beta \)-barrel. While the overall three-dimensional architecture of the insect molecule to be described here is similar to that observed in the aforementioned vertebrate proteins, the long chain fatty acid binds to the protein in a completely different manner. The X-ray coordinates for MFB2 have been deposited in the Brookhaven Protein Data Bank (Bernstein et al., 1977) or may be obtained immediately via: HOLDEN@VS.MAC.WISC.EDU (INTERNET) or HOLDEN@WISC.MAC (BITNET).

### 2. Materials and Methods

(a) Crystallization and preparation of heavy-atom derivatives

The fatty-acid-binding protein referred to as MFB2 was isolated from day-2 5th-instar-larval midgut tissue as previously described (Smith et al., 1992). Crystallization trials were conducted by the hanging drop method of vapor diffusion (for a review, see McPherson, 1982). Micro-crystals were first observed growing overnight at low pH in droplets equilibrated against 30 M-ammonium sulfate. Subsequently, somewhat larger crystals were grown more slowly from 1.6 M-ammonium sulfate solutions containing 50 mM-Na\(^+\)/K\(^+\) succinate, 5 mM-Na\(_2\)PO\(_4\) (pH 4.5). These crystals were still not large enough for X-ray data collection, however, and tended to grow as bundles of rods. Attempts to grow crystals at higher pH values were not successful.

To ensure a reproducible supply of single crystals, the technique of macro-seeding was subsequently employed with the sitting drop method of vapor diffusion (Thaller et al., 1985). For these experiments, 20 μl of protein at 11 mg/ml and buffered with 10 mM-N-2-hydroxyethylpiperazine-N’-2-ethanesulfonic acid (Hepes), pH 7.0, were mixed with 20 μl of a 1.4 M-ammonium sulphate solution containing 50 mM-Na\(^+\)/K\(^+\) succinate. 5 mM-Na\(_2\)PO\(_4\) (pH 4.5). These droplets were equilibrated for 5 days in glass Petri dishes against 10 ml of 1.4 M-ammonium sulfate, buffered with 50 mM-succinate at pH 4.5. Small protein crystals previously grown in hanging drops were then rapidly washed in 10 M-ammonium sulfate solutions at pH 4.5 and transferred to the equilibrated sitting drops. At the time of the seeding process, the solution in the bottom of the Petri dish was replaced with 10 ml of a 1.6 M-ammonium sulfate solution buffered with 50 mM-succinate at pH 4.5. All crystallization experiments were conducted at room temperature and were generally complete within 4 weeks.

Based on precession photographs recorded with a conventional rotating anode X-ray source and \( \mu = 13^\circ \), the crystals were assigned to space group \( P2_1 \), with unit cell dimensions of \( a = 27.5 \text{ Å}, b = 71.0 \text{ Å}, c = 287 \text{ Å} \) and \( \beta = 90^\circ \). The asymmetric unit contained 1 molecule. From still setting photographs, the maximum resolution of measurable X-ray data was estimated to be 1.7 Å resolution.

For the preparation of heavy-atom derivatives, crystals were transferred to a synthetic mother liquor containing 1.8 M-ammonium sulfate, 50 mM-succinate, 5 mM-Na\(_2\)PO\(_4\) (pH 4.5) and various heavy-metal reagents. Precession photography was used to monitor the binding of the heavy atom to the crystalline protein. Four isomorphous heavy-atom derivatives were readily prepared by soaking the protein crystals in 5 mM-K\(_2\)UO\(_4\)F\(_6\) for 7 days, 1 mM-UO\(_3\)(OCOCH\(_3\))\(_2\) for 48 h, 2.5 mM-NaAuCl\(_4\) for 20 h, and 2.5 mM-NaAuCl\(_4\) for 24 h followed by "back-soaking" in synthetic mother liquor for 1 h. The gold derivative
was of particular interest in that the soaking solution was bright yellow but the crystals turned very deep purple, thus indicating a change in the co-ordination geometry of the gold ion upon binding to the protein.

Before the search for heavy-atom derivatives was initiated, several attempts were made to solve the structure of the insect protein by molecular replacement using as a search model the X-ray co-ordinates of the P2 myelin protein generously supplied to us by Dr T. Alwyn Jones. A convincing solution to the rotation and translation function was never obtained, however, and in light of the fact that heavy-atom derivatives were easily prepared, it seemed prudent to proceed with the structure determination by multiple isomorphous replacement. In retrospect, as suggested by the reviewer of this manuscript, a composite search model derived from the X-ray co-ordinates of the 5 previously determined fatty-acid-binding proteins may have provided a more effective search model for molecular replacement.

(b) X-ray data collection and processing

Three-dimensional X-ray data sets were collected from the native and the 4 heavy-atom derivative crystals to 1.75 Å and 2.5 Å resolution, respectively, at 4°C with the Siemens X3000 area detector system. These X-ray data were subsequently processed with the data reduction software XDS (Kabsch, 1988a,b) and internally scaled according to the algorithm of Fox & Holmes (1966) as implemented by Dr Phil Evans. Friedel pairs were measured for 2 of the heavy-atom-derivative X-ray data sets, namely the double site NaAuCl₃ derivative and the KI₃O₃F₃ derivative. Crystals used for X-ray data collection were typically 0.5 mm × 0.5 mm × 0.2 mm in size. Only 1 crystal was required per X-ray data set. The X-ray source was nickel-filtered copper Kα radiation from a Rigaku RU200 X-ray generator operated at 50 kV and 50 mA. Each heavy-atom-derivative X-ray data set was placed on the same scale as the native X-ray data set by a "local" scaling procedure developed at the Enzyme Institute, University of Wisconsin, by Drs G. Wesenberg, W. Rypniewski and I. Rayment. With this method, the scale for a particular reflection was computed from the neighboring reflections in a volume defined by a sphere or a rectangular prism. The relative contribution of a reflection to a scale factor was weighted according to the distance of the reflection from that which was to be scaled. Relevant X-ray data collection and scaling statistics may be found in Table 1. The native X-ray data set contained 96% of the total theoretical number of observations to 1.75 Å resolution.

(c) Computational methods

The binding positions of the heavy atoms to the crystalline protein were determined by inspection of appropriate difference Patterson maps calculated with X-ray data from 30 Å to 5.0 Å and were placed on a common origin by difference Fourier maps. Each uranyl derivative X-ray data set was placed on the same scale as the native X-ray data set by a "local" scaling procedure developed at the Enzyme Institute, University of Wisconsin, by Drs G. Wesenberg, W. Rypniewski and I. Rayment. With this method, the scale for a particular reflection was computed from the neighboring reflections in a volume defined by a sphere or a rectangular prism. The relative contribution of a reflection to a scale factor was weighted according to the distance of the reflection from that which was to be scaled. Relevant X-ray data collection and scaling statistics may be found in Table 1. The native X-ray data set contained 96% of the total theoretical number of observations to 1.75 Å resolution.

Table 2

<table>
<thead>
<tr>
<th>Derivative</th>
<th>Site no.</th>
<th>Relative occupancy</th>
<th>x</th>
<th>y</th>
<th>z</th>
<th>Location</th>
</tr>
</thead>
<tbody>
<tr>
<td>UO₂(OCOCH₃)₂</td>
<td>1</td>
<td>147300</td>
<td>0.2417</td>
<td>0.0000</td>
<td>0.1069</td>
<td>Between turn 45-48 and symmetry-related Lys20</td>
</tr>
<tr>
<td>NaAuCl₄</td>
<td>1</td>
<td>89082</td>
<td>0.4441</td>
<td>-0.0260</td>
<td>0.9499</td>
<td>Side-chain His82</td>
</tr>
<tr>
<td>NaAuCl₄</td>
<td>1</td>
<td>93943</td>
<td>0.4346</td>
<td>-0.1141</td>
<td>0.8860</td>
<td>Side-chain His82</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>123340</td>
<td>0.4860</td>
<td>-0.2031</td>
<td>0.9492</td>
<td>Between Met41, Lys105 and Tyr107</td>
</tr>
<tr>
<td>K₁₃O₃F₃</td>
<td>1</td>
<td>57741</td>
<td>0.8722</td>
<td>-0.4149</td>
<td>0.7586</td>
<td>Between Asp46 and symmetry-related Asn14 and Asp16</td>
</tr>
</tbody>
</table>

x, y, z are the fractional atomic co-ordinates.
were used for determining the correct hand of the heavy-atom constellation. Protein phases were calculated with the program HEAVY (Terwilliger & Eisenberg, 1983) and relevant phase calculation statistics are given in Table 3. The phasing included the anomalous scattering information from 2 of the heavy-atom derivatives.

A polyalanine peptide chain, based on the P2 myelin protein, was globally rotated into an electron density map calculated with X-ray data from 30 Å to 2.5 Å using an Evans and Sutherland PS390 graphics system and the molecular modeling program FRODO (Jones, 1985). For the most part, the electron density was very well defined except at several surface loops connecting the α-strands. The "globally" fitted polyalanine model was then manually adjusted to the electron density. Subsequently, protein phases based on the model were calculated with the software package TNP (Tronrud et al., 1987) and combined with the phases determined from the heavy-atom derivatives according to the algorithm of Dr Randy J. Read (1986) who kindly supplied us with the software. Calculation of another electron density map to 2.5 Å resolution with these "combined" phases allowed for the positioning of most of the amino acid side-chains. Least-squares refinement of the model was then initiated at 2.5 Å resolution with the package TNT. The starting R-factor was 46%. When the R-factor had dropped to 30%, the refinement was extended to 1.75 Å resolution. Once the R-value dropped to 25% for all measured X-ray data from 30 Å to 1.75 Å, solvent molecules were systematically added to the X-ray co-ordinate set. Peaks of electron density were considered to be ordered solvent molecules if they were within 3.2 Å of potential hydrogen-bonding groups, and if they appeared in both electron density maps calculated with $|F_o - F_c|$ and $|F_o - F_c|$ coefficients and contoured at 1σ and 3σ, respectively. All positions and temperature factors for the individual protein atoms and the 71 solvent molecules were refined. The occupancies of the solvent molecules were set to unity. Bond lengths and angles for the hydrocarbon chain of the fatty-acid ligand were restrained to typical values observed for carbon–carbon single bonds and for $sp^3$ hybridized carbon atoms, respectively. The carboxylic acid moiety was restrained to be planar and to maintain typical carbon–oxygen bond lengths observed for other carboxylic acid groups. Reduction of the R-factor from 46% to 17.3%, took approximately 2 weeks and required 16 cycles of refinement and manual model building. Relevant refinement statistics may be found in Table 4 and the distribution of the mean main-chain temperature factors is given in Fig. 1.

**Table 3**

<table>
<thead>
<tr>
<th>Resolution range</th>
<th>x&lt;sub&gt;-8.86&lt;/sub&gt;</th>
<th>5.64</th>
<th>4.42</th>
<th>3.76</th>
<th>3.32</th>
<th>3.01</th>
<th>2.77</th>
<th>2.58</th>
</tr>
</thead>
<tbody>
<tr>
<td>No. of reflections</td>
<td>153</td>
<td>323</td>
<td>398</td>
<td>469</td>
<td>538</td>
<td>590</td>
<td>634</td>
<td>660</td>
</tr>
<tr>
<td>Figure of merit</td>
<td>0.81</td>
<td>0.82</td>
<td>0.77</td>
<td>0.71</td>
<td>0.71</td>
<td>0.69</td>
<td>0.63</td>
<td>0.57</td>
</tr>
<tr>
<td>Phasing power (CO&lt;sub&gt;2&lt;/sub&gt;(OCOCH&lt;sub&gt;3&lt;/sub&gt;)&lt;sub&gt;2&lt;/sub&gt;)</td>
<td>1.13</td>
<td>1.60</td>
<td>1.11</td>
<td>0.95</td>
<td>0.83</td>
<td>1.13</td>
<td>1.17</td>
<td>1.18</td>
</tr>
<tr>
<td>Phasing power (NaAuCl&lt;sub&gt;4&lt;/sub&gt;)</td>
<td>1.74</td>
<td>1.64</td>
<td>1.45</td>
<td>1.08</td>
<td>1.15</td>
<td>1.33</td>
<td>1.19</td>
<td>1.49</td>
</tr>
<tr>
<td>Phasing power (S&lt;sub&gt;2&lt;/sub&gt;S&lt;sub&gt;2&lt;/sub&gt;(OCOCH&lt;sub&gt;3&lt;/sub&gt;)&lt;sub&gt;2&lt;/sub&gt;)</td>
<td>1.14</td>
<td>1.55</td>
<td>1.23</td>
<td>1.16</td>
<td>1.43</td>
<td>1.70</td>
<td>1.70</td>
<td>1.30</td>
</tr>
<tr>
<td>Phasing power (NaAuCl&lt;sub&gt;4&lt;/sub&gt;)&lt;sup&gt;+&lt;/sup&gt;</td>
<td>2.68</td>
<td>2.24</td>
<td>1.64</td>
<td>1.33</td>
<td>1.53</td>
<td>1.46</td>
<td>1.65</td>
<td>2.18</td>
</tr>
<tr>
<td>Phasing power (NaAuCl&lt;sub&gt;4&lt;/sub&gt;)&lt;sup&gt;†+&lt;/sup&gt;</td>
<td>1.02</td>
<td>1.95</td>
<td>1.25</td>
<td>1.09</td>
<td>1.13</td>
<td>1.41</td>
<td>1.26</td>
<td>1.01</td>
</tr>
<tr>
<td>Phasing power (K&lt;sub&gt;2&lt;/sub&gt;CO&lt;sub&gt;3&lt;/sub&gt;F&lt;sub&gt;4&lt;/sub&gt;)</td>
<td>1.03</td>
<td>0.39</td>
<td>0.38</td>
<td>0.43</td>
<td>0.36</td>
<td>0.50</td>
<td>0.46</td>
<td>0.33</td>
</tr>
<tr>
<td>Phasing power (K&lt;sub&gt;2&lt;/sub&gt;CO&lt;sub&gt;3&lt;/sub&gt;F&lt;sub&gt;4&lt;/sub&gt;)&lt;sup&gt;‡+&lt;/sup&gt;</td>
<td>1.09</td>
<td>0.89</td>
<td>0.46</td>
<td>0.58</td>
<td>0.55</td>
<td>0.36</td>
<td>0.56</td>
<td>0.55</td>
</tr>
</tbody>
</table>

Phasing power is the ratio of the root-mean-square heavy-atom scattering factor amplitude to the root-mean-square lack of closure error.

† Single-site derivative.

‡ Double-site derivative.

**Table 4**

<table>
<thead>
<tr>
<th>Refinement statistics</th>
</tr>
</thead>
<tbody>
<tr>
<td>Resolution limits (Å)</td>
</tr>
<tr>
<td>Final R factor (%)</td>
</tr>
<tr>
<td>No. of reflections used</td>
</tr>
<tr>
<td>No. of atoms</td>
</tr>
<tr>
<td>Weighted root-mean-square deviations from ideality</td>
</tr>
<tr>
<td>Bond length (Å)</td>
</tr>
<tr>
<td>Bond angle (deg.)</td>
</tr>
<tr>
<td>Planarity (trigonal) (Å)</td>
</tr>
<tr>
<td>Planarity (other planes) (Å)</td>
</tr>
<tr>
<td>Torsion angle (deg.)&lt;sup&gt;‡&lt;/sup&gt;</td>
</tr>
</tbody>
</table>

<sup>‡</sup> R-factor = $\frac{\sum |F_o| - |F_c|}{\sum |F_o|}$

† The torsion angles were not restrained during refinement.

**Figure 1.** Plot of the mean B-value versus amino acid residue for all main-chain atoms. Most of the amino acid residues have mean B-values well below 300 Å² except for Ser34 and Ser52, both of which adopt alternate conformations. Asn44, Gly45 and Asp46, which reside in an approximate Type II' turn, and Gly55, Gly56, Gly57 and Ala58, which are located in a surface loop. The average B-value for all polypeptide chain backbone atoms including the above-mentioned amino acid residues is 15.1 Å².
3. Results and Discussion

The three-dimensional structure of the fatty-acid-binding protein, MFB2, isolated from *Manduca sexta* L. has now been determined and refined to a nominal resolution of 1.75 Å and a crystallographic R-factor of 17.3% for all measured X-ray data. Representative portions of the electron density map calculated to 1.75 Å resolution with coefficients of the form $|2F_o - F_c|$ and contoured at 1σ are shown in Figure 2(a) and (b). For the most part, the electron density is very well ordered with only one break in the polypeptide chain backbone occurring at Gly56. MFB2 is an acetylated protein (Smith et al., 1992) and this functional group is clearly visible in the electron density map at the N-terminal serine residue. There are several amino acids whose side-chains are not well defined and these include Lys11 (CE and NZ), Lys20 and 28 (CE and NZ), Lys36 (NZ), Asn44 (CG, OD1, ND2), Asp46 (CG, OD1, OD2), Asn88 (ND2, OD1) and Lys130 (NZ). Also, there is a break in the electron density between CG and CD of Gln30 and there is no electron density for the side-chain of Asp122. All of these amino acid

Figure 2. Representative portions of the electron density map calculated to 1.75 Å resolution. Portions of electron density map shown here were contoured at 1σ and calculated with coefficients of the form $(2F_o - F_c)$, where $F_o$ is the observed structure factor amplitude and $F_c$ is the calculated structure factor amplitude. (a) The electron density is very well ordered at both the N and C termini. Shown here is the electron density near the C terminus and corresponding to Arg126, Arg127, Tyr128 and Tyr129. (b) The electron density corresponding to the fatty-acid ligand and various surrounding amino acid residues.
residues are located at the surface of the molecule. The electron densities corresponding to amino acid residues Ser34 and Ser52 appear like valine or threonine residues, thus suggesting that these side-chains adopt alternate conformations. Consequently, for refinement purposes, Ser34 and Ser52 were each modeled in two conformations. As estimated from the height of the electron density, the occupancies for the alternate conformations of Ser34 were set to 0.75 and 0.25, while the occupancies for the two conformations of Ser52 were given the equal weight of 0.5.

The three-dimensional model for MFB2 presented here differs from the published primary sequence by five amino acids (Smith et al., 1992). It was reported that position 50 was an isoleucine residue. The electron density for this side-chain, however, appears more trigonal than tetrahedral, suggesting that the residue may be an aspartate or asparagine residue. Also, this side-chain is located in a rather hydrophilic environment. Consequently, residue 50 has been built into the electron density as an asparagine. Another problem occurs in the region delineated by amino acid residues 55 to 59. While amino acid residue 55 should be an isoleucine, there is no corresponding electron density and, furthermore, it is located in a surface loop where isoleucine would not be expected. In the present model, residue 55 has been left as a glycine. Also, residues 58 and 59, which should be glutamate and arginine residues, respectively, have been modeled in as alanine and lysine side-chains. It is possible that the side-chain for residue 58 is disordered, since it is at a surface loop thereby explaining the discrepancy. Likewise, the side-chain for residue 59, while easily accommodating a lysine residue, may in fact be the predicted arginine with the guanidinium group somewhat disordered. Finally, the electron density for amino acid residue 77 is too small for a glutamate residue but nicely accommodates an aspartate residue. Consequently, in the present model, this residue has been built in as an aspartate. Other than these minor changes, the amino acid sequence based on the cDNA agrees well with the electron density and none of these changes occurs in positions that will effect the overall conclusions described below.

Crystals of MFB2 are densely packed with a solvent content of approximately 39%. A packing diagram of the protein within the monoclinic cell is shown in Figure 3. There are potentially 13 hydrogen bonds between symmetry-related molecules in the crystalline lattice within a cut-off limit of 3.2 Å. Seven of these electrostatic contacts are between backbone amide nitrogen or carbonyl oxygen atoms and side-chain atoms; the other six are formed by side-chain–side-chain interactions. Of particular interest is the side-chain–side-chain interaction between Asp27 and Asp110 where the OD1 atoms for each are within 2.8 Å. This close interaction suggests that at least one or both of the side-chains are protonated and may partially explain why it was not possible to grow crystals from ammonium sulfate solutions at pH values greater than 4.5. Of the 70 water molecules built into the electron density, eight are directly involved in bridging one protein molecule to another within the unit cell.

A φ,ψ plot of all non-glycinyl main-chain dihedral angles is given in Figure 4 and a ribbon drawing of the molecule in Figure 5. Approximately 92% of the amino acid residues in MFB2 adopt standard secondary structural conformations and the polypeptide chain backbone dihedral angles are all within the theoretically allowed regions. The three-dimensional positions of all polypeptide chain backbone atoms are displayed in Figure 6. As can be seen, MFB2 consists of ten strands of anti-parallel β-pleated sheet forming an up-and-down β-barrel which is flanked on one side by a helix–turn–helix motif. A list of the amino acid residues involved in secondary structural elements may be found in Table 5 and a summary of the φ,ψ angles for the reverse turns is given in Table 6. The strands of β-pleated sheet form two layers that are nearly orthogonal to one another with one layer containing four and the other six β-strands. As listed in Table 5, there are technically 11 strands of β-pleated sheet, but the two β-strands delineated by amino acid residues 7 to 9 and 11 to 13 are generally
considered as one and it is this stretch of amino acid residues that contributes to the formation of both layers of sheet. The first $\alpha$-helix, delineated by amino acid residues 15 to 22, is decidedly amphipathic with an average hydrophobicity of 0.26 and a mean hydrophobic moment of 0.39 as calculated with the program MOMENT (Eisenberg et al., 1989). Accordingly, this stretch of amino acids falls into the category of surface-seeking peptides and as can be seen from Figure 6, is located near the aliphatic portion of the fatty acid ligand. The second $\alpha$-helix, composed of amino acid residues 26 to 34, falls into the range for normal globular proteins with an average hydrophobicity of −0.02 and a mean hydrophobic moment of 0.36.

In the refined model of MFB2 presented here, there are 70 ordered water molecules and one sulfate ion. The sulfate ion is located near the carboxylic acid moiety of the fatty acid ligand as described below. Temperature factors for the water molecules range from 8.4 to 48.4 Å$^2$ with 52 of them having $B$-values below 35 Å$^2$. Most of the ordered solvent molecules are located at the surface of the protein. There are, however, 13 water molecules located within the interior of MFB2, as shown in Figure 7. With the exception of the one water molecule

Figure 4. A Ramachandran plot of all non-glycyl main-chain dihedral angles for the MFB2 model. Fully allowed $\phi, \psi$, values are enclosed by continuous lines; those only partially allowed are enclosed by broken lines.

Figure 5. Ribbon drawing of the MFB2 molecule. This Fig. was generated with software kindly provided by Dr J. P. Priestle (Priestle, 1988). For this type of molecular structure representation, $\beta$-pleated sheets are represented as arrows and $\alpha$-helices as coils. The insect protein has overall dimensions of 36 Å x 40 Å x 30 Å and contains 10 strands of anti-parallel $\beta$-pleated sheet and 2 rather short $\alpha$-helices. All subsequent Figs presented in this paper are in the same orientation as this.

Figure 6. Stereo view of the MFB2 molecule. This Fig. was generated with the plotting software package PLUTO, originally written by Dr Sam Motherwell and modified for proteins by Eleanor Dodson and Phil Evans. For this representation, all polypeptide chain backbone atoms are shown and drawn with filled bonds. The fatty-acid ligand is displayed using open bonds.
Structure of an Insect Fatty-acid-binding Protein

Table 5
List of secondary structural elements

<table>
<thead>
<tr>
<th>Amino acid residues</th>
<th>Type of structure</th>
</tr>
</thead>
<tbody>
<tr>
<td>2-5</td>
<td>Type II turn</td>
</tr>
<tr>
<td>7-9 (A)</td>
<td>β-Sheet</td>
</tr>
<tr>
<td>11-13 (A)</td>
<td>β-Sheet</td>
</tr>
<tr>
<td>15-22</td>
<td>γ-Helix</td>
</tr>
<tr>
<td>26-34</td>
<td>β-Sheet</td>
</tr>
<tr>
<td>37-43 (B)</td>
<td>β-Sheet</td>
</tr>
<tr>
<td>44-47</td>
<td>Approximate type II' turn</td>
</tr>
<tr>
<td>48-53 (C)</td>
<td>β-Sheet</td>
</tr>
<tr>
<td>54-57</td>
<td>Type II turn</td>
</tr>
<tr>
<td>58-63 (D)</td>
<td>β-Sheet</td>
</tr>
<tr>
<td>64-67</td>
<td>Type II turn</td>
</tr>
<tr>
<td>68-72 (E)</td>
<td>β-Sheet</td>
</tr>
<tr>
<td>73-76</td>
<td>Approximate type II' turn</td>
</tr>
<tr>
<td>77-85 (F)</td>
<td>β-Sheet</td>
</tr>
<tr>
<td>86-95 (G)</td>
<td>Type II turn</td>
</tr>
<tr>
<td>96-99</td>
<td>Type III turn</td>
</tr>
<tr>
<td>100-107 (H)</td>
<td>β-Sheet</td>
</tr>
<tr>
<td>108-111</td>
<td>Type II turn</td>
</tr>
<tr>
<td>112-117 (I)</td>
<td>β-Sheet</td>
</tr>
<tr>
<td>118-121</td>
<td>Type I turn</td>
</tr>
<tr>
<td>124-131 (J)</td>
<td>β-Sheet</td>
</tr>
</tbody>
</table>

While it is not well understood how the fatty acid ligand enters and leaves the protein, it is obvious that the binding pocket is accessible to molecules larger than water such as sulfate, phosphate, and tetrachloroaurate(III) ions.

For the most part, the interior of MFB2 is lined with hydrophobic and uncharged amino acid residues, with the exceptions being Asp35, Asp71, His92, Lys105 and Arg127. A close-up view of the binding pocket is shown in Figure 9. The fatty acid ligand was modeled as palmitic acid and only those amino acid residues within 4.0 Å of atoms of the ligand are displayed. The hydrocarbon portion of the ligand is surrounded in the immediate vicinity by various hydrophobic amino acid residues including Phe15, Phe18, Phe103, Ile73 and Ile116, Leu19 and Val79. There are three solvent molecules located within the binding pocket. Two are presumably water and are not directly involved in the binding of the fatty acid while the electron density for the third is large enough to accommodate either a sulfate or phosphate anion. For the model presented here, this third solvent has been fitted into the electron density as a sulfate anion. The carboxylic acid moiety of the fatty acid is linked to the protein via hydrogen bonds as indicated by the broken lines in Figure 10. One of the carboxylic acid oxygen atoms is 2.6 Å from the hydroxyl group of Tyr129 and 2.9 Å from NH1 of Arg127 while the

Table 6
List of dihedral angles for reverse turns

<table>
<thead>
<tr>
<th>Amino acid residues</th>
<th>Type</th>
<th>$\phi_2$</th>
<th>$\psi_2$</th>
<th>$\phi_3$</th>
<th>$\psi_3$</th>
</tr>
</thead>
<tbody>
<tr>
<td>2-5</td>
<td>II</td>
<td>-560</td>
<td>132.8</td>
<td>797</td>
<td>-53</td>
</tr>
<tr>
<td>44-47</td>
<td>II</td>
<td>81.6</td>
<td>-95.7</td>
<td>-735</td>
<td>-29.4</td>
</tr>
<tr>
<td>54-57</td>
<td>II</td>
<td>-69.4</td>
<td>133.6</td>
<td>88.6</td>
<td>5.8</td>
</tr>
<tr>
<td>64-67</td>
<td>II</td>
<td>-63.4</td>
<td>130.4</td>
<td>84.4</td>
<td>0.1</td>
</tr>
<tr>
<td>73-76</td>
<td>II</td>
<td>64.1</td>
<td>-132.3</td>
<td>-1068</td>
<td>57.3</td>
</tr>
<tr>
<td>86-89</td>
<td>II</td>
<td>67.3</td>
<td>-120.4</td>
<td>-90.8</td>
<td>2.8</td>
</tr>
<tr>
<td>96-99</td>
<td>III</td>
<td>-61.9</td>
<td>-35.3</td>
<td>-70.7</td>
<td>-23.9</td>
</tr>
<tr>
<td>108-111</td>
<td>II</td>
<td>64.5</td>
<td>-135.8</td>
<td>-78.4</td>
<td>3.8</td>
</tr>
<tr>
<td>118-121</td>
<td>I</td>
<td>-78.3</td>
<td>-56.9</td>
<td>-1074</td>
<td>50</td>
</tr>
</tbody>
</table>

While it is not well understood how the fatty acid ligand enters and leaves the protein, it is obvious that the binding pocket is accessible to molecules larger than water such as sulfate, phosphate, and tetrachloroaurate(III) ions.

For the most part, the interior of MFB2 is lined with hydrophobic and uncharged amino acid residues, with the exceptions being Asp35, Asp71, His92, Lys105 and Arg127. A close-up view of the binding pocket is shown in Figure 9. The fatty acid ligand was modeled as palmitic acid and only those amino acid residues within 4.0 Å of atoms of the ligand are displayed. The hydrocarbon portion of the ligand is surrounded in the immediate vicinity by various hydrophobic amino acid residues including Phe15, Phe18, Phe103, Ile73 and Ile116, Leu19 and Val79. There are three solvent molecules located within the binding pocket. Two are presumably water and are not directly involved in the binding of the fatty acid while the electron density for the third is large enough to accommodate either a sulfate or phosphate anion. For the model presented here, this third solvent has been fitted into the electron density as a sulfate anion. The carboxylic acid moiety of the fatty acid is linked to the protein via hydrogen bonds as indicated by the broken lines in Figure 10. One of the carboxylic acid oxygen atoms is 2.6 Å from the hydroxyl group of Tyr129 and 2.9 Å from NH1 of Arg127 while the

Figure 7. Positions of the internal solvent molecules in MFB2. The 13 internal solvent molecules are shown here as open spheres. For simplicity, only those amino acid side-chains that interact with these solvent molecules are specifically displayed. Broken lines indicate potential hydrogen bonds between solvent and amino acid side-chains. Those water molecules that interact with only main-chain atoms are displayed as open spheres with no surrounding broken lines.

Figure 8. The positions of the two gold-binding sites are shown. Both of these heavy atoms substitute in the interior of the protein with one gold ion co-ordinating to His92 and the other wedged between Met41, Lys105 and Tyr107. The gold-binding site located near His92 is occupied by the putative sulfate ion in the native holo-protein.
Figure 8. Binding sites for the potassium tetrachloroaurate(III) derivative. The positions of the gold heavy-atom-binding sites are shown as open spheres. One of the gold ions co-ordinates to His92, whereas the other is located near Met41, Lys105 and Tyr107. In the native holo-protein a sulfate or phosphate ion occupies the gold site near His92. While it is not well understood how the fatty acid enters or leaves the binding pocket, clearly the interior of the pocket is accessible to small molecules.

other carboxylic acid oxygen atom is 3.2 Å from NE2 of Gln29 and 2.7 Å from one of the oxygen atoms of the sulfate group. Clearly, for the oxygen atom of the fatty acid to be within 2.7 Å of the oxygen atom donated by the sulfate, the functional group of the fat must be protonated. Since the crystals were grown at pH 4.5 it is not surprising that the fatty acid is protonated. Whether or not the putative sulfate is important in the binding of the fatty acid at physiological pH or rather is an artifact of the crystallization conditions remains to be determined. Experiments designed to transfer the crystals to higher pH values are presently underway. As can be seen from Figure 9, if the sulfate ion was removed, it is possible that the amino group of Lys105 could move into position to form a hydrogen bond with the carboxylate group of the fat.

The hydrogen-bonding pattern around the carboxylic acid group of the fatty acid in MFR2 is different from that observed in the rat intestinal fatty-acid-binding protein. In the rat fatty-acid-binding protein, the guanidinium group of Arg106 is in an approximate planar orientation to the carboxylate group of the ligand (Sacchettini et al., 1989). This arginine amino acid residue in the rat protein corresponds to Lys105 in the insect protein. Likewise, there are no tyrosine residues in the immediate vicinity of the fatty acid functional group in the rat protein as compared to the insect molecule. Also, the exact location of the carboxylic acid moiety within the binding pocket is quite different. With respect to the P2 myelin protein and the adipocyte lipid-binding protein, however, the binding of the carboxylate portion of the ligand is much more similar to that observed in the insect protein. In P2 myelin protein the carboxylate group is within the hydrogen-bonding distance of the hydroxyl group of Tyr128 and NH2 of Arg106 (as determined from X-ray co-ordinates provided by Dr

Figure 9. Close up view of the fatty acid binding pocket. Only those amino acid residues within approximately 4.0 Å of atoms in the fatty acid are shown. Potential hydrogen bonds between the carboxylic acid moiety of the ligand and the protein side-chains are indicated by broken lines. Open spheres indicate solvent molecules.
Figure 10. Hydrogen-bonding pattern around the functional group of the ligand. Those amino acid residues that participate in hydrogen bonding with the fatty acid are shown. Only the first 3 carbon atoms of the ligand are displayed. Both the fatty acid carboxylic acid group and the putative sulfate ion are located at the left.

T. Alwyn Jones) and in the adipocyte lipid-binding protein it is within hydrogen-bonding distance of the hydroxyl group of Tyr128, NE of Arg126 and two solvent molecules (Xu et al., 1992).

The type of three-dimensional fold seen in MFB2 has been previously observed in a variety of X-ray investigations including those of the rat intestinal fatty-acid-binding protein, both holo and apo-forms (Sacchettini et al., 1989; Scapin et al., 1992), bovine P2 myelin protein (Jones et al., 1988), chicken liver basic fatty-acid-binding protein (Scapin et al., 1990), and the fatty-acid-binding protein from bovine heart muscle (Müller-Fährnow et al., 1991) and murine adipocyte lipid-binding protein (Xu et al., 1992).

A superposition of the alpha carbon atom positions for the insect and the murine adipocyte lipid-binding proteins was made according to the algorithm of Rossmann & Argos (1975) and is shown in Figure 11. These two molecules are approximately 67% identical and superimpose with a root-mean-square value of 0.9 Å using 88 structurally equivalent alpha carbon atom positions.

Based on the amino acid sequence alignment of eight vertebrate proteins known to belong to this family of lipid binding proteins, it was originally concluded that only six amino acid residues are strictly conserved among all of them (Jones et al., 1988). Quite strikingly, while many of these presumed conserved amino acid residues are retained in the insect protein, there are a few that are not. For example, three of the conserved amino acid residues are Gly6, Gly46 and Gly67 in the P2 myelin protein. In the insect molecule, however, while both Gly46 and Gly67 (Gly45 and Gly66 in the insect numbering) are retained, Gly6 is replaced by a lysine residue. Likewise, according to the vertebrate protein sequences aligned thus far, position 42 (P2 myelin numbering), is always an isoleucine residue but in the insect protein it is replaced with a methionine residue (M41). By far the most significant difference between the insect and the vertebrate proteins is at position 33 which for the vertebrate seems is either a glycine or an alanine residue. It was suggested by Jones et al. (1988) that any larger side-chain in this position would cause a reduction in the volume that could be occupied by the fatty acid. In the insect protein, however, this position is occupied by a leucine residue (Leu32) and indeed, because this side-chain is more bulky, the hydrocarbon chain of the ligand binds to the insect protein in a very different manner from that observed previously with the vertebrate proteins, as can be seen in Figure 10. The direct consequence of this substitution is that the hydrocarbon chain of

Figure 11. Superposition of the insect and the murine lipid-binding proteins. The insect protein is drawn as filled bonds while the murine adipocyte lipid binding protein is displayed as open bonds. While both proteins bind the functional group of the fat in similar regions, the orientation of the hydrocarbon tails after C6 is quite different. X-ray co-ordinates of the murine adipocyte lipid-binding protein used for this Fig. were graciously supplied by Dr Leonard J. Banaszak.
the fatty acid ligand at position C6 and beyond, binds to the insect molecule more towards the interior of the β-barrel rather than towards the two conserved helices as observed in the vertebrate systems.

In summary, the molecular structure of an insect fatty-acid-binding protein, MFB2, has now been determined and refined to high resolution. As expected, the overall three-dimensional fold of the molecule is very similar to other lipid-binding proteins such as the rat intestinal fatty-acid-binding protein, P2 myelin protein, and adipocyte lipid-binding protein. The carboxylic acid moiety of the fatty acid ligand binds in the same general vicinity within the β-barrel as seen in P2 myelin and adipocyte lipid-binding proteins, but in a rather different position as compared to the rat fatty-acid-binding protein. The detailed hydrogen bonding patterns around the ligand functional group are different, however, in all four of these proteins. Also, due to the substitution of a leucine residue at position 32 in the insect molecule, the hydrocarbon chain of the fatty acid will bind in a manner more similar to that of MFB2 with the lipid. It will be of interest to determine whether the fatty acid will bind in a manner more similar to that observed in the vertebrate proteins. This research is underway.

We would like to express our appreciation to Drs T. Alwyn Jones and Leonard J. Banaszak for supplying us with the refined X-ray co-ordinates of the P2 myelin protein and the adipocyte lipid-binding protein, respectively. We also thank Dr Ivan Rayment for helpful discussions and advice throughout the course of the investigation. This research was supported in part by grants from the NIH (HL42322 to H.M.H. HL39116 to M.A.W. and GM13656 to A.F.S.). H.M.H. is an Established Investigator of the American Heart Association.

References


Scapin, G., Spadon, P., Mammi, M., Zanotti, G. & Monaco, H. L. (1990). Crystal structure of chicken...


*Edited by B. W. Matthews*