

## The Structural Determination of an Insect Sterol Carrier Protein-2 with a Ligand-bound C16 Fatty Acid at 1.35-Å Resolution\*

Received for publication, June 12, 2003, and in revised form, July 9, 2003  
Published, JBC Papers in Press, July 10, 2003, DOI 10.1074/jbc.M306214200

David H. Dyer‡, Scott Lovell‡, James B. Thoden‡, Hazel M. Holden‡, Ivan Rayment‡§, and Que Lan¶||

From the Departments of ‡Biochemistry and ¶Entomology, University of Wisconsin, Madison, Wisconsin 53706

**Yellow fever mosquito sterol carrier protein (SCP-2) is known to bind to cholesterol. We report here the three-dimensional structure of the complex of SCP-2 from *Aedes aegypti* with a C16 fatty acid to 1.35-Å resolution. The protein fold is exceedingly similar to the human and rabbit proteins, which consist of a five-stranded  $\beta$ -sheet that exhibits strand order 3-2-1-4-5 with an accompanying layer of four  $\alpha$ -helices that cover the  $\beta$ -sheet. A large cavity exists at the interface of the layer  $\alpha$ -helices and the  $\beta$ -sheet, which serves as the fatty acid binding site. The carboxylate moiety of the fatty acid is coordinated by a short loop that connects the first  $\alpha$ -helix to the first  $\beta$ -strand, whereas the acyl chain extends deep into the interior of the protein. Interestingly, the orientation of the fatty acid is opposite to the observed orientation for Triton X-100 in the SCP-2-like domain from the peroxisomal multifunctional enzyme (Haapalainen, A. M., van Aalten, D. M., Merilainen, G., Jalonen, J. E., Pirila, P., Wierenga, R. K., Hiltunen, J. K., and Glumoff, T. (2001) *J. Mol. Biol.* 313, 1127–1138). The present study suggests that the binding pocket in the SCP-2 family of proteins may exhibit conformational flexibility to allow coordination of a variety of lipids.**

Sterol carrier protein-2 (SCP-2),<sup>1</sup> or nonspecific lipid transfer protein, was first isolated as a cholesterol transporter involved in cholesterol and lipid intracellular trafficking in vertebrates (1, 2). It is a small protein, with a molecular mass of about 13 kDa, whose function has been the source of considerable debate. SCP-2 belongs to a family of proteins, known as the SCP-2 gene family, that all contain a domain homologous to the sterol carrier protein-2. This gene family includes SCP-2, SCP-X, 17 $\beta$ -hydroxysteroid dehydrogenase type IV, metallo- $\beta$ -lacta-

mase, and stomatin (2).<sup>2</sup> In vertebrates, SCP-2 and SCP-X share exactly the same nucleotide sequences of SCP-2 domains, because they are transcribed from the same gene (the SCP-2/SCP-X gene). In all cases where the SCP-2 domain is part of a larger protein, the domain is located at the C terminus. Furthermore, the sequence identity among gene family members is very high, suggesting a conserved functional role for this protein (2). Interestingly, those enzymes that contain this domain all modify hydrophobic substrates.

SCP-2 has been shown to bind to cholesterol, fatty acid, fatty acyl-CoA, and acidic phospholipids (4, 5) where the affinity for different ligands is in the following order: cholesterol  $\gg$  straight chain fatty acid  $>$  kinked chain fatty acid (6). Given this broad specificity and wide spread tissue distribution, it has proved difficult to identify the primary role of this protein (2). However, SCP-2 has been shown to accelerate translocation of cholesterol between intracellular membrane structures (7–10). Likewise, the biological importance of SCP-2 in cholesterol trafficking was deduced from the knockout mice that show moderately decreased level of cholesterol absorption by intestine and severally reduced bile salt formation (11). SCP-2 is also involved in fatty acid metabolism showing symptoms related to peroxisomal deficiency in SCP-2 knockout mice (12). In addition, SCP-2 may participate in the lipid transport of phosphatidylinositol between intracellular sites and the plasma membrane (13). Although there is strong biological evidence for the multifunctional binding of lipids to SCP-2, this has not been demonstrated at the molecular level. For example, SCP-2 extracted from liver does not contain associated cholesterol or any lipid (2, 14). Thus, how SCP-2 binds to cholesterol or fatty acids *in vivo* and how it functions as a transporter continues to be a mystery.

Homologs of SCP have been found across the animal kingdom including in insects (Fig. 1). The latter are particularly interesting, because insects have lost a number of key enzymes in cholesterol biosynthesis pathways (15), which results in complete dependence on exogenous sources of cholesterol for biosynthesis of its steroid derivatives (16–20). This therefore demands that insects must have a mechanism for uptaking, transporting, and storing the cholesterol that is necessary throughout their life cycle. Indeed, insects have the tendency to accumulate cholesterol in the body during feeding stages when their diet is richer in lipids (21). During that life stage, at least in *Manduca sexta*, there is a very low rate of cholesterol catabolism (22).

Intracellular transportation of cholesterol in insects must meet two important biological needs: first the necessity to absorb free cholesterol for the construction of cellular mem-

\* This work was supported in part by United States Department of Agriculture CSREES Hatch Project 04487 (to Q. L.) and National Institutes of Health Grant AR-35186 (to I. R.). The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

The atomic coordinates and structure factors (code 1PZ4) have been deposited in the Protein Data Bank, Research Collaboratory for Structural Bioinformatics, Rutgers University, New Brunswick, NJ (<http://www.rcsb.org/>).

§ To whom correspondence may be addressed: Dept. of Biochemistry, 433 Babcock Dr., Madison, WI 53706. Tel.: 608-262-0437; Fax: 608-262-1319; E-mail: Ivan\_Rayment@biochem.wisc.edu.

¶ To whom correspondence may be addressed: Dept. of Entomology, University of Wisconsin, Madison, WI 53706. Tel.: 608-263-7924; Fax: 608-262-3322; E-mail: qlan@entomology.wisc.edu.

<sup>1</sup> The abbreviations used are: SCP-2, sterol carrier protein-2; AeSCP-2, *A. aegypti* SCP-2; rAeSCP-2, recombinant AeSCP-2; GST, glutathione S-transferase.

<sup>2</sup> NCBI conserved domain search (2002), available on the World Wide Web at [www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi](http://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi).

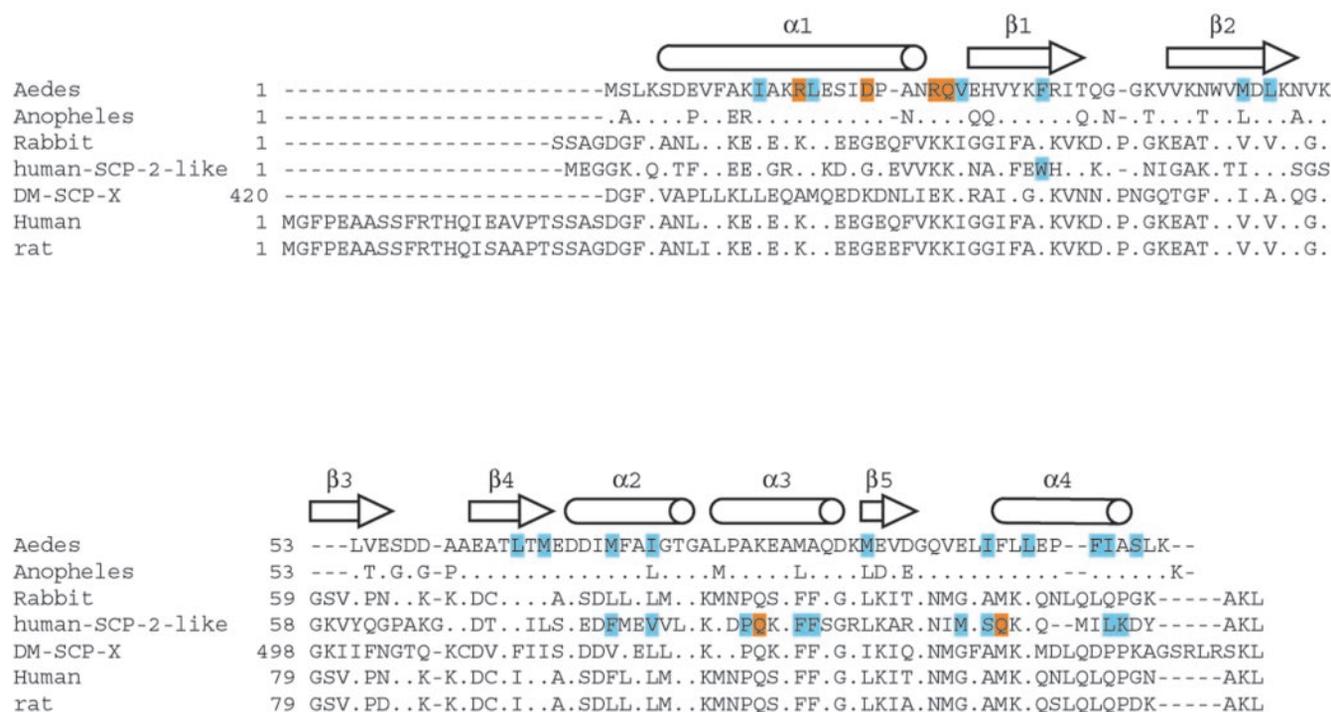


FIG. 1. **Amino acid sequence alignment of members of the sterol carrier protein family of proteins.** This list includes the following: *Aedes*, yellow fever mosquito SCP-2 *A. aegypti*, AAO34708; *Anopheles*, *Anopheles gambiae* SCP-2 (EAA08376); *Rabbit*, rabbit SCP-2 AAC15422; *Human-SCP-2-like*, human multifunctional enzyme type 2-SCP-2 domain 1IKTA; *DM-SCP-X*, fruit fly C-terminal portion of SCP-X X97685; *Human*, human SCP-2 (NM\_002979); *Rat*, rat SCP-2 (M57454). The alignment was prepared with the program ClustalW version 1.8 (41). The secondary structural elements correspond to the structure of yellow fever mosquito SCP-1, where the residues in contact with the hydrophobic acyl chain are highlighted in cyan, and those that coordinate the carboxyl moiety are depicted in orange. The residues that contact Triton X-100 in the structure of the human multifunctional enzyme type 2-SCP-2 domain (1IKTA) are depicted with a similar color scheme.

branes, and second to provide cholesterol as a precursor for the steroid biosynthesis. These two pathways most likely utilize the same intracellular transport protein(s) to mobilize cholesterol. At this time, SCP-2 appears to be a good candidate as a participant in this task.

SCP-X has been reported to have high levels of expression in the midgut of *Drosophila* embryos; however, only a 1.6-kb mRNA transcript arises from this transcript (23). This differs from vertebrates, where the SCP-X/SCP-2 gene combination produces multiple transcripts. In yellow fever mosquito (*Aedes aegypti*), an independent gene has been identified that is similar to vertebrate SCP-2 (AeSCP-2). This protein also has high levels of expression in the midgut of larvae and high binding affinity for cholesterol (24).

The *in vivo* function of insect SCP-2 is unknown. However, because insects do not synthesize cholesterol (25), it is hypothesized SCP-2 may be involved in shuttling cholesterol and dietary sterols from lysosomes from which exogenous sterol enters the cell to the endoplasmic reticulum and mitochondria. After conversion of dietary sterols to cholesterol or cholesterol to 7-dehydrocholesterol in endoplasmic reticulum, SCP-2 may also be involved in transfer of the cholesterol to the mitochondria for steroid biosynthesis. The lack of peroxisomal localization sequence in the C terminus of AeSCP-2 (24) indicates that AeSCP-2 may only be involved in absorption and trafficking of cholesterol. This simplicity offers a unique opportunity for studying the sterol/lipid transporting function of the SCP-2 class of proteins.

The three-dimensional structures of SCP-2 domain from rabbits and humans have been determined (26, 27) together with the SCP-2-like domain from the human peroxisomal multifunctional enzyme (28). These structures contain a unique fold that consists of one layer of antiparallel  $\beta$ -sheet juxtaposed against a layer of  $\alpha$ -helices. A large hydrophobic cavity exists between

these two layers of secondary structural elements, where it is presumed that cholesterol and fatty acyl derivatives are bound. The size of the cavity is appropriate for either a sterol or fatty acyl derivative. This fold is different from that found in most fatty acid-binding proteins that belong to the lipocalin superfamily and contain a  $\beta$ -clam motif (29, 30).

Although the affinity for lipids is very high, there has only been one detailed structural report of a ligand bound to the SCP-2-like domain of the human multifunctional enzyme (28). In this structure, Triton X-100, which bears some analogy to the protein's normal acyl-Co-A substrates, was observed to bind in a hydrophobic tunnel. Apart from this study, earlier observations by NMR revealed that the hydrophobic cleft is the likely binding site for fatty acyl ligands where the lipid interacts with residues of the first and third helices in SCP-2 (27). However, the precise molecular interaction between SCP-2 and any of its natural ligand(s) is unknown.

In an effort to understand of structure and function of class of protein we report here, the structure of SCP-2 from the yellow fever mosquito complexed with a C16 fatty acid to 1.35-Å resolution. This study was initiated in part to expand our understanding of invertebrate members of the SCP-2 superfamily (24) but also because changing the source of a protein is often the best route for obtaining crystals of a ligand complex. As such, this structure provides insight into the manner in which fatty acids bind to this class of proteins. The results also show that there is a high degree of conservation in the three-dimensional structures of invertebrate and vertebrate SCP-2, which implies that the function of SCP-2 is conserved across the animal kingdom.

#### MATERIALS AND METHODS

**Purification of Recombinant AeSCP-2 Protein**—To produce recombinant AeSCP-2 (rAeSCP-2), the PCR product of the entire coding region

TABLE I  
X-ray data collection statistics

Parameter	Value
No. of crystals	1
Resolution (Å)	30.0–1.35 (1.41–1.35) <sup>a</sup>
Measurements	78,412 (5542)
Average <i>I</i> / $\sigma$	21.8 (3.3)
Unique reflections	21,157 (2383)
Redundancy	3.7 (2.3)
Completeness (%)	97.0 (91.5)
<i>R</i> <sub>merge</sub> <sup>b</sup>	3.8 (25.2)

<sup>a</sup> The values in parentheses are for the highest resolution shell between 1.41 and 1.35 Å.

<sup>b</sup>  $R_{\text{merge}} = (\sum |I_{hkl} - \bar{I}| / (\sum I_{hkl})) \times 100$ , where the average intensity *I* is taken over all symmetry equivalent measurements and *I*<sub>hkl</sub> is the measured intensity for a given reflection.

TABLE II  
Refinement statistics

Parameter	Value
Resolution limits (Å)	30.0–1.35
R-factor <sup>a</sup> (overall) % (No. of reflections)	18.7% (21,157)
R-factor (working) % (No. of reflections)	18.4% (19,050)
R-factor (free) % (No. of reflections)	22.6% (2107)
No. of protein atoms <sup>b</sup>	890
No. of heteroatoms <sup>c</sup>	107
Average B values (Å <sup>2</sup> )	
Main chain atoms	16.8
Side chain atoms	23.8
All protein atoms	20.3
Water molecules	35.8
Fatty acid	25.8
Weighted root mean square deviations from ideality	
Bond lengths (Å)	0.012
Bond angles (degrees)	2.234
Trigonal planes (Å)	0.007
General planes (Å)	0.012
Torsional angles (degrees) <sup>d</sup>	16.6

<sup>a</sup> R-factor =  $(\sum |F_o - F_c| / \sum |F_o|) \times 100$ , where *F*<sub>o</sub> is the observed structure-factor amplitude and *F*<sub>c</sub> is the calculated structure-factor amplitude.

<sup>b</sup> These include multiple conformations for Ser<sup>18</sup> and Asn<sup>43</sup>.

<sup>c</sup> These include 88 water molecules.

<sup>d</sup> The torsional angles were not restrained during the refinement.

of the AeSCP-2 gene was cloned into the pGEX-4T-2 GST tag vector (Amersham Biosciences). PCR primers were 5'-gtgaattcgaATGTCTCT-GAAGTCCG-3' (capital letters represent coding sequence; boldface letters represent the start codon; an *Eco*RI site was incorporated for cloning) and 5'-tactcgagTTACTTCAGCGAGG-3' (capital letters represent the antisense of coding sequence; boldface letters represent the antisense of the stop codon; a *Xho*I site was incorporated for cloning). The expression vector was transferred into *Escherichia coli* strain BL21 under 100 µg/ml ampicillin selection (Amersham Biosciences). Sequence analysis was performed to confirm that the fusion protein was in frame with GST.

The rAeSCP-2 expression bacteria were grown in 200 ml of Luria-Bertani (LB) medium with 100 µg/ml ampicillin at 37 °C overnight, and 50 ml of overnight bacterial culture was added into 500 ml of fresh LB medium with 100 µg/ml ampicillin and grown at 37 °C for about 2 h (*A*<sub>600</sub> = 0.8). Then expression of rAeSCP-2 was induced by the addition of isopropyl-β-D-thiogalactoside to a final concentration of 0.2 mM, and the culture was incubated at 18 °C overnight to prevent the formation of inclusion body-bound rAeSCP-2.

Cells from 2.5 liters were harvested and resuspended in 30 ml of phosphate-buffered saline (140 mM NaCl, 10 mM Na<sub>2</sub>HPO<sub>4</sub>, 1.8 mM KH<sub>2</sub>PO<sub>4</sub>, 2.7 mM KCl, pH 7.4) with 5 mM dithiothreitol and 2 mM EDTA. Cells were lysed with a French press at 15,000 p.s.i. at 4 °C. The cell lysate was centrifuged at 14,000 × *g* at 4 °C for 1 h to remove cellular debris. The GST/rAeSCP-2 fusion protein was purified on a GST affinity column (10-ml bed volume; Amersham Biosciences), and the GST tag was removed by digesting with 500 units of thrombin (Amersham Biosciences) in the column at 22 °C overnight. The use of GST fusion and thrombin cleavage introduces additional residues at the N terminus compared with the wild type protein. Consequently, the amino acid sequence of the protein for the N terminus is Gly-Ser-Pro-Gly-Ile-Arg-

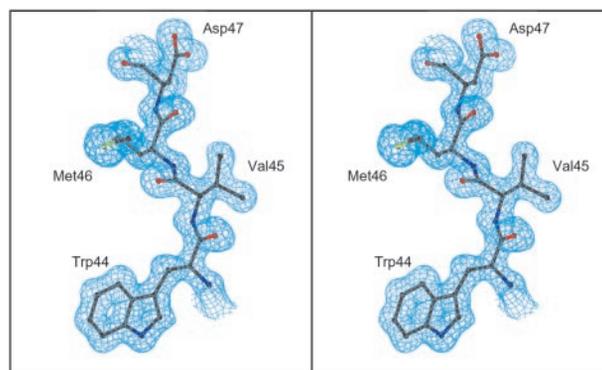


FIG. 2. Stereo view of a representative region of electron density from the final model for the mosquito SCP-2. The map was calculated with coefficients of the form  $2F_o - F_c$  map and was contoured at 1  $\sigma$ . The region shown extends from Trp<sup>44</sup> to Asp<sup>47</sup>. Figs. 2–6 were prepared with the programs Molscript and Bobscript (39, 42) and rendered with Raster3d (3).

Met . . . , where the methionine indicates the start of the native sequence. The predicted molecular weight of the expression construct (rAeSCP-2) after cleavage is 12,839.90, which was confirmed by mass spectroscopy to be 12,841 (data not shown). Thrombin was removed from eluted rAeSCP-2 by passing through a benzamidine column (Amersham Biosciences). Purified rAeSCP-2 was concentrated in a Centricon YM-10 device (Amicon) to 10 mg/ml in phosphate-buffered saline, pH 7.4, stored in phosphate-buffered saline at –80 °C. A typical purification procedure yielded 100 mg of protein from 2.5 liters of cell culture. Prior to crystallization trials, the protein was dialyzed in a Slide-A-Lyzer dialysis cassette (Pierce) for 20 h against 25 mM KCl, 50 mM Tris-HCl, pH 8.5, at 4 °C.

**Crystallization of rAeSCP-2**—Crystals of ligand-bound rAeSCP-2 were grown by hanging drop vapor diffusion at room temperature. Typically, 3 µl of rAeSCP-2 protein at 10 mg/ml was combined with an equal volume of 2 M sodium malonate, 100 mM Tris-HCl, pH 7.5. Crystals grew as hexagonal plates to a size of 0.6 × 0.6 × 1.5 mm within 5 days. The crystals were very stable at room temperature and retained strong diffraction a year after they were initially prepared. The crystals belonged to space group P2<sub>1</sub>2<sub>1</sub>2<sub>1</sub> and had cell dimensions of *a* = 29.3 Å, *b* = 46.1 Å, *c* = 70.4 Å.

**Structural Analysis of rAeSCP-2**—The x-ray data were collected to 1.35-Å resolution with CuK $\alpha$  radiation at 4 °C with a Bruker HSTAR area detector equipped with Super long mirrors. The x-ray source was a Rigaku RU200B x-ray generator operated at 50 kV and 90 mA. The x-ray data were processed with SAINT and internally scaled with XSCALIBRE.<sup>3</sup> X-ray data collection statistics are presented in Table I.

The structure was determined by molecular replacement starting from the structure of rabbit SCP-2 (RSCB accession number 1C44 (26)) utilizing the program EPMR (31). Iterative cycles of least-squares refinement and manual model building with the programs TNT and Turbo (32, 33) reduced the R-factor to 18.7% for all measured x-ray data from 30.0- to 1.35-Å resolution. The R-free was 22.6% for 10% of the data that were excluded from the refinement. Least-squares refinement statistics are presented in Table II. Analysis of the coordinates with the program PROCHECK (34) revealed that 92.2% of the residues lie in the most favored regions of the Ramachandran plot, whereas the remaining 6.8% of the residues lie in additionally allowed areas. One residue, Val<sup>51</sup>, lies in the generously allowed region, but its electron density is unequivocal. No residues are located in the disallowed regions. A section of representative electron density is shown in Fig. 2.

## RESULTS AND DISCUSSION

**Structural Description**—The structure of recombinant SCP-2 from the yellow fever mosquito (rAeSCP-2) was determined to 1.35-Å resolution in complex with a 16-carbon fatty acid. The electron density for the polypeptide chain extends almost continuously from Gly<sup>-3</sup> to Lys<sup>110</sup>, where Gly<sup>37</sup> and Gly<sup>38</sup> are less well ordered than other sections of the polypeptide chain (Fig. 2). The first three visible amino acids, Gly<sup>-3</sup>, Ile<sup>-2</sup>, and Arg<sup>-1</sup>, are a residual from the thrombin cleavage introduced by the

<sup>3</sup> I. Rayment and G. Wesenberg, unpublished results.

FIG. 3. Stereo ribbon representations of mosquito SCP-2. The  $\alpha$ -helices and  $\beta$ -sheets are numbered according to their occurrence in the structure.

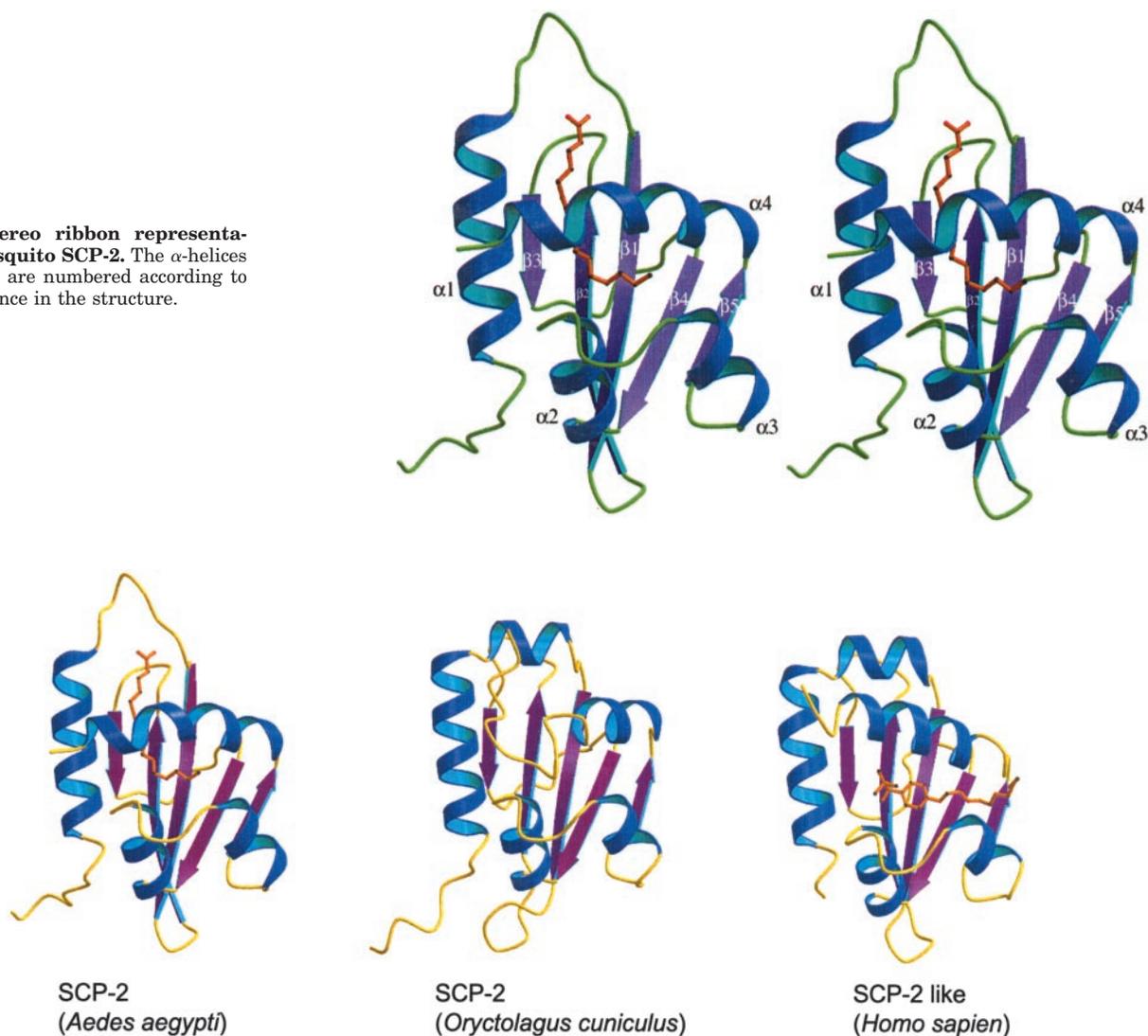


FIG. 4. Comparison of the tertiary structures of sterol carrier protein from mosquitoes, rabbits, and humans.

GST fusion construct. The first three amino acids in this cloning artifact (Gly<sup>-6</sup>, Ser<sup>-5</sup>, and Pro<sup>-4</sup>) are not well ordered but are present in the crystal lattice on the basis of mass spectroscopic measurements.

As shown in Fig. 3, SCP-2 belongs to the  $\alpha + \beta$  tertiary structural classification of proteins where the fold is characteristic of the sterol carrier protein superfamily as defined by structural classification of proteins (35). The order of secondary structural elements in the protein is as follows:  $\alpha_1$ - $\beta_1$ - $\beta_2$ - $\beta_3$ - $\beta_4$ - $\alpha_2$ - $\alpha_3$ - $\beta_5$ - $\alpha_4$ . This fold is dominated by a five-stranded  $\beta$ -sheet that exhibits strand order 3-2-1-4-5 with an layer of four  $\alpha$ -helices that cover the  $\beta$ -sheet. A large cavity exists at the interface of the layers of  $\alpha$ -helices and the  $\beta$ -sheet, which serves as the lipid binding site discussed below.

The tertiary structure of mosquito SCP-2 is very similar to that of both human and rabbit SCP-2 (Fig. 4). Indeed, superposition of the coordinates of human protein on the mosquito carrier protein with the program Align (36) reveals that the root mean square difference between 97 structurally equivalent amino acids is only 1.15 Å. Likewise for the protein from rabbit, the root mean square difference is similarly 1.1 Å for 97 equivalent residues. This indicates a remarkable structural similarity between the mammalian and insect proteins in light of their limited sequence identity of approximately ~30% (Fig. 1). The major structural difference between mosquito SCP-2

and rabbit and human SCP-2 is the loop that coordinates the carboxylate group of the fatty acid in mosquito SCP, which is replaced by an  $\alpha$ -helix in the mammalian proteins. The orientation of the final  $\alpha$ -helix differs somewhat between these proteins as does the loops that connect. There are also a few small insertions and deletions in the loops that connect  $\beta$ -strands 2 and 3, and  $\beta$ -strands 3 and 4.

**Lipid Binding Site**—As noted above, the arrangement of secondary structural elements creates a large cavity that serves as the lipid binding site. In the structure of rAeSCP-2 described here at 1.35-Å resolution, this cavity contains a saturated 16-carbon fatty acid, which, on the basis of the unequivocal electron density, corresponds to palmitic acid (Fig. 5a). Although the bend in the fatty acid might suggest an unsaturated fatty acid, the separation and position of the centers of mass in the electron density are only consistent with a saturated fatty acid. Furthermore, the distribution of methylene groups is inconsistent with a *cis*-linkage between any atoms. Finally, the bend occurs after C-10, which precludes oleic acid.

The carboxylate group of the fatty acid lies toward the outer opening of the hydrophobic pocket and is sequestered primarily by the loop that connects  $\alpha$ -helix-1 to  $\beta$ -strand-1, which is composed of residues Ser<sup>18</sup>-His<sup>28</sup> (Fig. 5b). One oxygen of the carboxylate moiety is coordinated by the side chain of Arg<sup>24</sup> and main chain amide hydrogens of Gln<sup>25</sup> and Val<sup>26</sup>. This

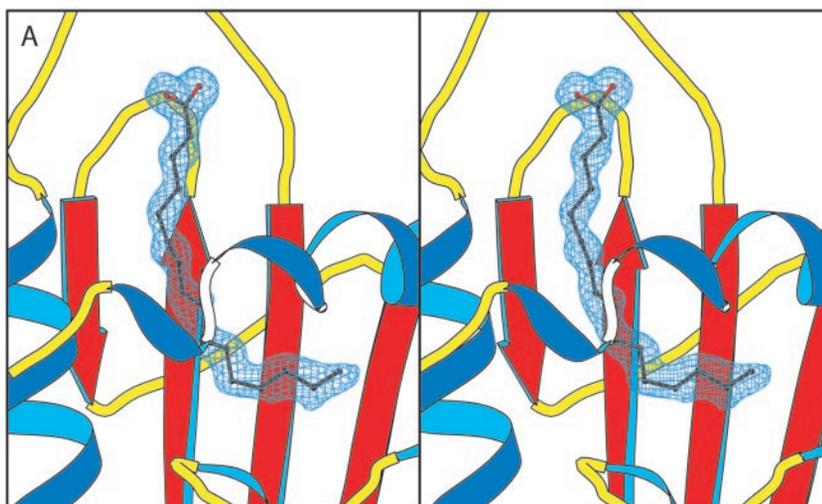
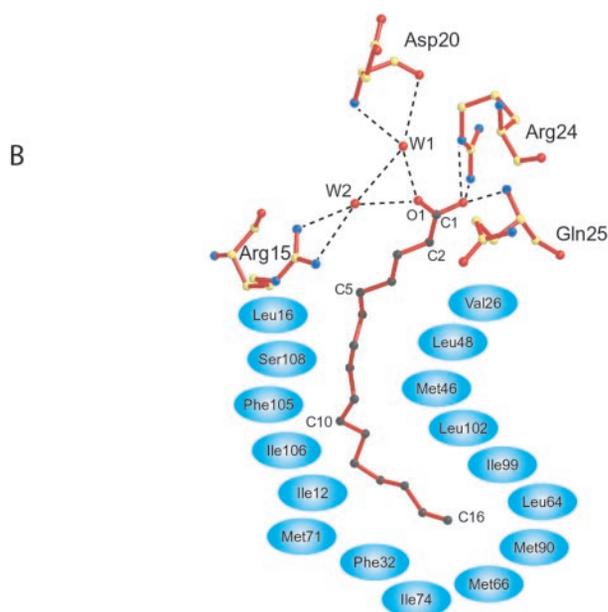


FIG. 5. Stereo view of the difference electron density for the lipid bound to mosquito SCP-2 (A) and an interaction diagram depicting the major interactions between the lipid and the protein (B). The map was calculated with coefficients of the form  $2F_o - F_c$  map and was contoured at  $1 \sigma$ .



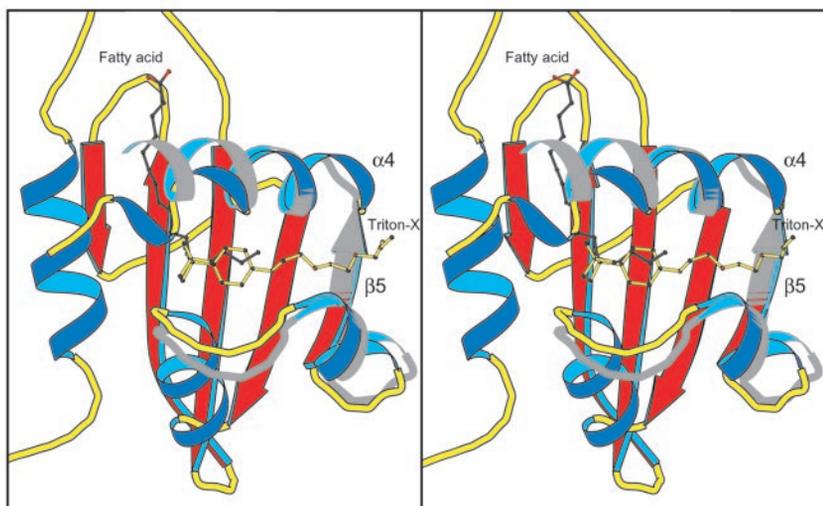
series of interactions serves to orient the carboxyl group in the binding pocket. The second oxygen is more solvent-exposed but is still hydrogen-bonded to two well defined water molecules. One of these is coordinated to the side chain of Arg<sup>15</sup>, whereas the second is bound to the main chain oxygens of Asn<sup>23</sup> and Asp<sup>20</sup>, together with the main chain amide hydrogen of Asp<sup>20</sup>. All of the interactions with the fatty acid carboxylate moiety, both direct and water-mediated, are provided by residues in the first helix and the loop that forms the connection to the first  $\beta$ -strand. It is of interest that the use of arginine residues to coordinate the carboxylate of a fatty acid is a feature common to other fatty acid-binding proteins that exhibit completely different protein folds (37).

The acyl chain of the fatty acid extends from the surface of protein deep into the binding pocket. In the first instance, the path of the acyl chain runs approximately parallel to the strands of the  $\beta$ -sheet roughly adjacent to strands 2 and 3. At around atom C-11 the acyl chain bends so that the final atoms of the chain run across the strands of the  $\beta$ -sheet. Coordination of the carboxyl moiety at the outer edge of the binding pocket increases the length of the acyl chain that can be accommodated by this protein.

The entire length of the acyl chain is surrounded by hydrophobic residues. Those in direct Van der Waals contact with the lipid include Met<sup>71</sup>, Ala<sup>81</sup>, Leu<sup>102</sup>, Phe<sup>105</sup>, and Ile<sup>99</sup> together with the hydrophobic components of side chains of Arg<sup>15</sup> and Arg<sup>24</sup>. This list represents only a subset of the hydrophobic residues that line the lipid binding cavity as indicated in Fig. 5b. This suggests that the hydrophobic cavity might accommodate a larger ligand such as cholesterol.

*Structural Comparison with Mammalian SCPs*—As noted above, the structure of yellow fever mosquito SCP-2 is exceedingly similar to that of the mammalian structures described earlier. This is perhaps not surprising, since they share the ability to bind and transport sterols. Even so, the percentage of residues that are structurally equivalent is much higher than expected for most homologous proteins that exhibit 30% sequence identity (Fig. 1). Indeed, for a typical group of homologous proteins, only 60% of the residues would be expected to reside in the common core, whereas for the SCP-2 family, ~90% of the residues lie within 3 Å of their structurally related counterparts (38). This suggests that the necessity of maintaining a cavity for the transport of lipids has placed restraints on the evolution of this protein fold. Indeed, comparison of the

**FIG. 6. Stereo superposition of the mosquito SCP-2 and Triton X from the structure of human SCP-2-like domain of the multifunctional enzyme.** The entire polypeptide chain is shown for the mosquito protein in a *ribbon representation*, whereas only the final  $\beta$ -strand and its associated  $\alpha$ -helices are shown for the human protein. For the latter, the secondary structural elements are depicted in *gray*, whereas the ordered component of the Triton-X 100 moiety is depicted with *yellow bonds*. The coordinates for the human SCP-2-like domain were obtained from the RSCB with accession number 1IKT (28) and superimposed with the program Align (36).



sequences of vertebrate and insect SCP-2s shows that although the exact residue varies on the interior of the cavity, the hydrophobicity is highly conserved (Fig. 1).

All of the structures of SCP-2 reported thus far have either been of the apo forms or in the presence of a nonnatural ligand (26–28). The structural determination reported here is the first to show how a fatty acid might bind in the lipid binding pocket. Interestingly, the secondary structure associated with the coordination of the carboxyl moiety in yellow fever mosquito SCP-2 is different from that in human or rabbit homologs. In the mosquito SCP-2, the coordination site is formed by a loop that connects the first helix with the first  $\beta$ -strand, whereas in both rabbits and humans, this loop is replaced by a short  $\alpha$ -helix. The presence of an  $\alpha$ -helix at this location demands that the mammalian proteins must coordinate the carboxyl moiety of fatty acids in a different manner from the insect proteins, since this secondary structural element eliminates the binding loop. The only way in which the binding could be maintained in the same manner as the insect protein would be if the  $\alpha$ -helix were to unfold to create a binding loop. Although this is unlikely, the number of residues associated with the loop in the insect protein and the helix in the mammalian proteins is similar. Interestingly, the sequence alignment shows that residues appropriate for forming a carboxyl binding pocket are found in each sequence, although they fulfill different structural roles in the mammalian sterol carrier proteins (Fig. 1).

The structure of the human SCP-2-like domain from the peroxisomal multifunctional enzyme complexed with the detergent Triton X has been reported (28). In this complex, the tetramethylbutyl-phenyl moiety of the triton molecule is deeply buried within the binding pocket. When the insect and human proteins are superimposed, the phenyl moiety of the Triton-X overlaps with C-11–C-16 of the palmitic acid (Fig. 6). As a consequence, the polyoxyethylene tail of the detergent extends in the opposite direction away from the carboxyl group of the fatty acid and exits the protein at a different portal (only three and a half oxyethylene units are observed in the model). This portal is formed by the triangulation of the fifth  $\beta$ -strand and the immediately preceding and succeeding  $\alpha$ -helices. A similar arrangement of secondary structural elements is found in mosquito SCP-2.

In the human SCP-2 domain, there is a cluster of exposed hydrophobic amino acid residues that surround the exit portal for the Triton-X polyoxyethylene tail, where the only polar component is provided by a group of glutamine residues (Gln<sup>90</sup>, Gln<sup>108</sup>, and Gln<sup>111</sup>). A similar arrangement is found in the mosquito protein so that although its hydrophobic binding

pocket does not extend to the surface at this point, it is conceivable that with minor rearrangements of the helices this portal could exist.

The manner in which cholesterol binds to SCP-2 is unknown for any member of this family, although there is unequivocal evidence that these proteins can carry a variety of lipids (2). Comparison of the human SCP-2-like domain with the insect protein suggests that the hydrophobic binding cleft may exhibit some flexibility in order to accommodate their ligands. Indeed, a larger cavity volume will certainly be required to accommodate cholesterol in the mosquito protein, since the Van der Waals volume of cholesterol is considerably larger than a C16 fatty acid (350 Å<sup>3</sup> compared with 230 Å<sup>3</sup> as calculated with the program Atvol (by M. J. Word; available on the World Wide Web at kinemage.biochem.duke.edu/software/software3.html#atvol). In the same manner, the orientation of cholesterol in the binding pocket is unknown. Although there is precedence for orientation of the polar moiety close to the carboxylate binding loop in the insect protein, the reverse orientation is suggested from the structure of Triton-X bound to the human SCP-2-like domain.

As noted earlier, although there is limited sequence similarity between mosquito SCP-2 and mammalian analogs, the structural similarity is remarkable. Given that the structure of rabbit SCP-2 was determined in its apo form and yet there is still a close structural relationship between both mosquito and the human SCP-2-like domain, which are bound to ligand, this suggests that the overall shape of the cavity does not change when lipids are sequestered. Thus, it would appear that, as in other fatty acid-binding proteins, the binding site in SCP-2 is preformed. If this is true, this would favor orientation of cholesterol with its polar moiety toward the carboxylate binding site, since the cavity is broader at that end. This would be in opposition to the human SCP-2-like domain, which utilizes a different exit portal. The overall volume of the cavity is substantial, although it is difficult to obtain a meaningful estimate of the volume due to variability between algorithms and probe parameters (40).

It is interesting that a C16 fatty acid was incorporated adventitiously into mosquito SCP-2 during its heterologous biosynthesis in *E. coli*, whereas expression of the mammalian proteins under similar conditions did not lead to incorporation of a hydrophobic ligand. This suggests that either the affinity of this protein for fatty acids is higher than that of its mammalian counterparts or that the manner in which the protein interacts with the cell membrane to load its ligand is different.

**Conclusions**—The results of this study show for the first time

how fatty acids bind to SCP-2. Since the fatty acid was incorporated during biosynthesis, this implies that, at least for insect proteins in this family, this might be a natural ligand, as has been implicated by fatty acid-binding studies (2). As a member of the SCP-2 gene family, SCP-2 is expected to function as an intracellular lipid transporter. AeSCP-2 is the first insect SCP-2 that has been studied in detail (24), and although its biological function still unknown, it is likely that SCP-2 plays an important role in lipid metabolism in insects. The three-dimensional structure of rAeSCP-2 is of interest in two aspects. First, rAeSCP-2 has high similarity to vertebrate SCP-2 structurally. Second, rAeSCP-2 is a fatty acid-binding protein and might have a function similar to that of vertebrate SCP-2.

A major finding of this study is that the fatty acid in the insect SCP binds in the opposite orientation to Triton X-100 in the human SCP-2-like domain of the multifunctional enzyme (28). In this way, the hydrophobic moieties of these lipids are oriented at opposite ends of the molecule. This suggests that these proteins may exhibit multiple binding modes for their ligands, depending on the nature of the lipid. Further experimental studies to examine this problem are in progress.

*Acknowledgments*—We thank Drs. Ann M. Fallon, Walter G. Goodman, Luis M. Rubio, David C. White, and James M. Cantu for helpful discussions on the study. We also thank Amanda Pitterle for technical assistance.

## REFERENCES

- Seedorf, U., Ellinghaus, P., and Roch Nofer, J. (2000) *Biochim. Biophys. Acta* **1486**, 45–54
- Gallegos, A. M., Atshaves, B. P., Storey, S. M., Starodub, O., Petrescu, A. D., Huang, H., McIntosh, A. L., Martin, G. G., Chao, H., Kier, A. B., and Schroeder, F. (2001) *Prog. Lipid Res.* **40**, 498–563
- Merritt, E. A., and Murphy, M. E. P. (1994) *Acta Crystallogr. Sect. D* **50**, 869–873
- Frolov, A., Cho, T. H., Billheimer, J. T., and Schroeder, F. (1996) *J. Biol. Chem.* **271**, 31878–31884
- Colles, S. M., Woodford, J. K., Moncecchi, D., Myers-Payne, S. C., McLean, L. R., Billheimer, J. T., and Schroeder, F. (1995) *Lipids* **30**, 795–803
- Schroeder, F., Frolov, A., Starodub, O., Atshaves, B. B., Russell, W., Petrescu, A., Huang, H., Gallegos, A. M., McIntosh, A., Tahotna, D., Russell, D. H., Billheimer, J. T., Baum, C. L., and Kier, A. B. (2000) *J. Biol. Chem.* **275**, 25547–25555
- van Noort, M., Rommerts, F. F., van Amerongen, A., and Wirtz, K. W. (1988) *Biochem. Biophys. Res. Commun.* **154**, 60–65
- Xu, T. S., Bowman, E. P., Glass, D. B., and Lambeth, J. D. (1991) *J. Biol. Chem.* **266**, 6801–6807
- Baum, C. L., Reschly, E. J., Gayen, A. K., Groh, M. E., and Schadick, K. (1997) *J. Biol. Chem.* **272**, 6490–6498
- Huang, H., Gallegos, A. M., Zhou, M., Ball, J. M., and Schroeder, F. (2002) *Biochemistry* **41**, 12149–12162
- Fuchs, M., Hafer, A., Munch, C., Kannenberg, F., Teichmann, S., Scheibner, J., Stange, E. F., and Seedorf, U. (2001) *J. Biol. Chem.* **276**, 48058–48065
- Seedorf, U., Raabe, M., Ellinghaus, P., Kannenberg, F., Fobker, M., Engel, T., Denis, S., Wouters, F., Wirtz, K. W., Wanders, R. J., Maeda, N., and Assmann, G. (1998) *Genes Dev.* **12**, 1189–1201
- Schroeder, F., Zhou, M., Swaggerty, C. L., Atshaves, B. P., Petrescu, A. D., Storey, S. M., Martin, G. G., Huang, H., Helmkamp, G. M., and Ball, J. M. (2003) *Biochemistry* **42**, 3189–3202
- Chanderbhan, R., Noland, B. J., Scallen, T. J., and Vahouny, G. V. (1982) *J. Biol. Chem.* **257**, 8928–8934
- Zdobnov, E. M., von Mering, C., Letunic, I., Torrents, D., Suyama, M., Copley, R. R., Christophides, G. K., Thomasova, D., Holt, R. A., Subramanian, G. M., Mueller, H. M., Dimopoulos, G., Law, J. H., Wells, M. A., Birney, E., Charlab, R., Halpern, A. L., Kokoza, E., Kraft, C. L., Lai, Z., Lewis, S., Louis, C., Barillas-Mury, C., Nusskern, D., Rubin, G. M., Salzberg, S. L., Sutton, G. G., Topalis, P., Wides, R., Wincker, P., Yandell, M., Collins, F. H., Ribeiro, J., Gelbart, W. M., Kafatos, F. C., and Bork, P. (2002) *Science* **298**, 149–159
- Clayton, R. (1964) *J. Lipid Res.* **5**, 3–19
- Noda, H., Wada, K., and Saito, T. (1979) *J. Insect Physiol.* **25**, 443–447
- Ritter, K. S., and Nes, W. R. (1981) *J. Insect Physiol.* **27**, 175–182
- Dwivedy, A. K., and Shukla, S. P. (1982) *Entomol.* **7**, 411–422
- Nes, W. D., Lopez, M., Zhou, W., Guo, D., Dowd, P. F., and Norton, R. A. (1997) *Lipids* **32**, 1317–1323
- Beydon, P., and Lafont, R. (1987) *Arch. Insect Biochem. Physiol.* **5**, 139–154
- Jouni, Z. E., Zamora, J., and Wells, M. A. (2002) *Arch. Insect Biochem. Physiol.* **49**, 167–175
- Kitamura, T., Kobayashi, S., and Okada, M. (1996) *Dev. Growth Differ.* **38**, 373–381
- Krebs, K. C., and Lan, Q. (2003) *Insect Mol. Biol.* **12**, 51–60
- Spates, G. E., DeLoach, J. R., and Chen, A. C. (1988) *J. Insect Physiol.* **34**, 1055–1061
- Choinowski, T., Hauser, H., and Piontek, K. (2000) *Biochemistry* **39**, 1897–1902
- Garcia, F. L., Szyperski, T., Dyer, J. H., Choinowski, T., Seedorf, U., Hauser, H., and Wuthrich, K. (2000) *J. Mol. Biol.* **295**, 595–603
- Haapalainen, A. M., van Aalten, D. M., Merilainen, G., Jalonen, J. E., Pirila, P., Wierenga, R. K., Hiltunen, J. K., and Glumoff, T. (2001) *J. Mol. Biol.* **313**, 1127–1138
- Flower, D. R., North, A. C., and Sansom, C. E. (2000) *Biochim. Biophys. Acta* **1482**, 9–24
- Zimmerman, A. W., and Veerkamp, J. H. (2002) *Cell Mol. Life Sci.* **59**, 1096–1116
- Kissinger, C. R., Gehlhaar, D. K., and Fogel, D. B. (1999) *Acta Crystallogr. D Biol. Crystallogr.* **55**, 484–491
- Tronrud, D. E. (1997) *Methods Enzymol.* **277**, 306–319
- Roussel, A., and Cambillau, C. (1991) *Silicon Graphics Geometry Partners Directory*, Vol. 86, Silicon Graphics, Mountain View, CA
- Laskowski, R. A., MacArthur, M. W., Moss, D. S., and Thornton, J. M. (1993) *J. Applied Crystallogr.* **26**, 283–291
- Lo Conte, L., Ailey, B., Hubbard, T. J., Brenner, S. E., Murzin, A. G., and Chothia, C. (2000) *Nucleic Acids Res.* **28**, 257–259
- Cohen, G. H. (1997) *J. Applied Crystallogr.* **30**, 1160–1161
- Thompson, J., Ory, J., Reese-Wagoner, A., and Banaszak, L. (1999) *Mol Cell Biochem.* **192**, 9–16
- Chothia, C., and Lesk, A. M. (1986) *EMBO J.* **5**, 823–826
- Esnouf, R. M. (1999) *Acta Crystallogr. Sect. D* **55**, 938–940
- Chakravarty, S., Bhingre, A., and Varadarajan, R. (2002) *J. Biol. Chem.* **277**, 31345–31353
- Thompson, J. D., Higgins, D. G., and Gibson, T. J. (1994) *Nucleic Acids Res.* **22**, 4673–4680
- Kraulis, P. J. (1991) *J. Applied Crystallogr.* **24**, 946–950