Characterization of a sterol carrier protein 2/3-oxoacyl-CoA thiolase from the cotton leafworm (*Spodoptera littoralis*): a lepidopteran mechanism closer to that in mammals than that in dipterans

Hajime TAKEUCHI*,1, Jian-Hua CHEN*2, John R. JENKINS†, Masanori BUN-YA‡, Philip C. TURNER* and Huw H. REES*

*Cell Regulation and Signalling Division, School of Biological Sciences, University of Liverpool, Biosciences Building, Crown Street, Liverpool, L69 7ZB, U.K., †Department of Medicine, University of Liverpool, The Henry Wellcome Laboratory, Crown Street, Liverpool, L69 3BX, U.K., ‡Faculty of Integrated Arts and Sciences, Hiroshima University, Kagamiyama 1-7-1, Higashi-Hiroshima, 739-8521, Japan

Numerous invertebrate species belonging to several phyla cannot synthesize sterols de novo and rely on a dietary source of the compound. SCPx (sterol carrier protein 2/3-oxoacyl-CoA thiolase) is a protein involved in the trafficking of sterols and oxidation of branched-chain fatty acids. We have isolated SCPs from *Spodoptera littoralis* (cotton leafworm) and have subjected it to limited amino acid sequencing. A reverse-transcription PCR-based approach has been used to clone the cDNA (1.9 kb), which encodes a 57 kDa protein. Northern blotting detected two mRNA transcripts, one of 1.9 kb, encoding SCPx, and one of 0.95 kb, presumably encoding SCP2 (sterol carrier protein 2). The former mRNA was highly expressed in midgut and Malpighian tubules during the last larval instar. Furthermore, constitutive expression of the gene was detected in the prothoracic glands, which are the main tissue producing the insect molting hormone. There was no significant change in the 1.9 kb mRNA in midgut throughout development, but slightly higher expression in the early stages. Conceptual translation of the cDNA and a database search revealed that the gene includes the SCP2 sequence and a putative peroxisomal targeting signal in the C-terminal region. Also a cysteine residue at the putative active site for the 3-oxoacyl-CoA thiolase is conserved. Southern blotting showed that SCPx is likely to be encoded by a single-copy gene. The mRNA expression pattern and the gene structure suggest that SCPx from *S. littoralis* (a lepidopteran) is evolutionarily closer to that of mammals than to that of dipterans.

Key words: cholesterol, ecdysone, lipid metabolism, prothoracic glands, steroid.

**INTRODUCTION**

Inability to synthesize sterols is apparently universal within the Arthropoda. Within this phylum the most detailed work on dietary sterol requirements and metabolism has been undertaken in insects. The majority of insect species investigated require some cholesterol, or a sterol which is convertible into cholesterol, for satisfactory growth, development and reproduction [1].

SCP2 (sterol carrier protein 2), which is encoded by a part of a fusion gene encoding SCPx (sterol carrier protein 2/3-oxoacyl-CoA thiolase) in many species, has multiple proposed functions in lipid metabolism and intracellular movement of cholesterol and other lipids in mammals [2,3]. SCP2 protein is expressed in various tissues, but is particularly abundant in organs involved in cholesterol absorption, transport or metabolism.

In vertebrate peroxisomes there are three 3-oxoacyl-CoA thiolases. While all three thiolases are involved in fatty acid oxidation, SCPx plays an exclusive role in branched-chain fatty acid oxidation and in the oxidation of the branched side chain of cholesterol to form bile acids. This is because the two other thiolases are inactive with branched-chain fatty-acyl-CoAs as substrates, but SCPx can use them [2,4]. Bile acids are essential to biliary cholesterol secretion, which involves bile acids, cholesterol and phospholipid. Bile acid formation involves β-oxidation of the branched alkyl side chain of 24-oxo-3α,7α,12α-trihydroxy-5β-cholestanoyl-CoA to cleave off propionyl-CoA. The SCPx exhibits 3β-oxothiolase activity to convert 24-oxo-3α,7α,12α-trihydroxy-5β-cholestanoyl-CoA into choleyl-CoA, which is subsequently converted into cholic acid and glycinated to glycocholic acid. Also, the SCPx shows 3β-oxothiolase activity to convert 24-oxo-3α,7α,12α-trihydroxy-5β-cholestanoyl-CoA into chenodeoxycholyl-CoA, which is the precursor of chenodeoxycholic acid [4–6]. Mice in which the gene was knocked out exhibited inhibition of β-oxidation in bile acid synthesis, based on the accumulation of 3α,7α,12α-trihydroxy-27-nor-5β-cholestone-24-one, which is a known bile alcohol derivative of the cholic acid synthesis intermediate 3α,7α,12α-trihydroxy-24-oxo-cholestanoyl-CoA, in bile and serum [7].

SCPx belongs to the SCP2 gene family, which at present includes four proteins: SCP2, SCPx, 17β-hydroxysteroid dehydrogenase IV and UNC-24 [8]. Apart from SCP2, the other homologues have SCP2 domains towards their C-termini. A single structural gene encodes both SCP2 and SCPx in humans [9], mice [10], rats [11] and chickens [12]. Alternative transcription initiation gives rise to two mRNAs encoding SCPx or SCP2 in these species. In humans, the SCPx mRNA encodes 547 amino acid residues representing a fusion protein of 3-oxoacyl-CoA thiolase (amino acid residues 1–404) and the SCP2 domain (amino acid...
residues 425–547). The product is post-translationally cleaved to the 46 kDa 3-oxoacyl-CoA thiolase and the 13 kDa SCP2. As mentioned above, SCP2 is also transcribed as a short mRNA which encodes 143 amino acid residues identical with the C-terminal 143 amino acids of SCPx. On the basis of a database search, the fused SCPx gene can be traced back to the fruitfly Drosophila melanogaster, whereas in yeast and the nematode worm Caenorhabditis elegans, the thiolase and SCP2 are not fused, but encoded by two separate genes [13,14].

Here we report the molecular cloning and characterization of the cDNA encoding SCPx from Spodoptera littoralis (cotton leafworm). The results suggest that SCPx may be involved in sterol absorption and synthesis of insect moulting hormones (ecdysteroids).

EXPERIMENTAL

Protein sequence
SCPx was isolated during the purification of the enzymes involved in the 3-epimerization of ecdysteroids as described previously [15]. The protein co-migrated closely with the second form of 3-dehydroecdysone 3′-reductase throughout the chromatographic purification steps. The purified protein was subjected to SDS/10%-(w/v)-PAGE, electrotransferred to ProBlott™ membrane (Applied Biosystems, Warrington, Cheshire, U.K.) and visualized by Coomassie Brilliant Blue staining. A single band was observed, which was excised and sequenced by an automated pulsed liquid-phase sequencer (Applied Biosystems 471A), giving the N-terminal amino acid sequence PRKVFVVGVGMNTFI.

cDNA cloning and sequencing
A PCR-based cloning strategy was used to isolate a cDNA fragment encoding this protein. Two degenerate sense primers were synthesized. Primers based on adjacent parts of the N-terminal amino acid sequence (primer 1: 5′-CCN MGI AAR GTI TTY GTN GTN GG, where N represents A/T/C/G, M is A/C, I is inosine, R is A/G, Y is C/T) were designed to incorporate the T7 promoter into the 5′-end of the sense strand. The product was excised and sequenced by an automated pulsed liquid-phase sequencer (Applied Biosystems 471A), giving the N-terminal amino acid sequence PRKVFVVGVGMNTFI.

In vitro translation of SCPx
The cDNA containing the entire open reading frame was amplified by PCR with the following gene specific primers: IVORF-F (5′-GGG AAC AGC CAC CAT GCC TAG AAA AGT GTT CGT TGT) and IVORF-R (5′-CCG CCT CTA GAG CGT ATG TTG GAG CGG ATG ATT GTG T). PCR was conducted as follows: one cycle of 94°C for 3 min, 15 cycles of 94°C for 1 min, 53°C for 30 s, 72°C for 2 min, and one cycle of 72°C for 5 min. The resulting PCR product was used as the template for the nested PCR, which was carried out as follows: 1 cycle of 94°C for 3 min, 30 cycles of 94°C for 1 min, 53°C for 30 sec, 72°C for 2 min, and 1 cycle of 72°C for 5 min. The primers for the nested PCR were as follows: T7IV primer (5′-GGT TTA TAT CAA CGC AGA GTG GCC ATT ATG GCC) and IVORF-R (5′-CCG CCT CTA GAG CGT ATG TTG GAG CGG ATG ATT GTG T). The second round of PCR yielded a product of approx. 300 bp, which was cloned into pGEM®-T Easy Vector, and the nucleotide sequences of several clones determined.

Northern- and Southern-blot analysis
Total RNA from various tissues was isolated using TRIzol® reagent (Invitrogen, Paisley, Renfrewshire, Scotland, U.K.). A 10 µg portion of total RNA was fractionated on a formaldehyde/agarose gel, transferred to Electran® nylon membrane (Merck) and hybridized with a probe corresponding to the open reading frame of the SCPx cDNA. The probe was radiolabelled with [α-32P]dCTP using a Ready-To-Go™ DNA labelling kit (Amersham Pharmacia Biotech) and loading was normalized by probing with a mouse 18S ribosomal RNA probe. Prehybridization and hybridization were carried out using QuickHyb® (Stratagene) under the conditions recommended by the manufacturer. The blots were washed at high stringency [0.1 x SSC
(1 × SSC is 0.15 M NaCl/0.015 M sodium citrate)/0.1% SDS, at 60 °C] and labelled bands visualized by autoradiography.

Genomic DNA was prepared from the fat body and Malpighian tubules of last larval instar *S. littoralis* using a modified phenol extraction method as described by Sambrook et al. [18]. Aliquots (10 µg) of DNA were digested with BamHI, BglII, EcoRI or XhoI, fractionated on a 1%-agarose gel, transferred to a nylon membrane, and hybridized using radiolabelled probes corresponding to different regions of SCPx cDNA. Hybridization was carried out using QuickHyb® (Stratagene) under the conditions recommended by the manufacturer. The blot was washed at high stringency (0.1 × SSC/0.1% SDS at 60 °C), and labelled bands were visualized by autoradiography.

Recombinant SCPx protein expression

The cDNA containing the entire open reading frame was amplified by PCR with the following gene-specific primers: ORF-F (5'-CGG ATC CCC TAG AAA AGT GTT CGT T) and ORF-R (5'-CGA ATT CTC ACA GTT TGG AGC GGA TTG T). These primers were designed to incorporate BamHI and EcoRI sites into the 5'-end and 3'-end respectively of the sense strand. PCR was conducted as follows: one cycle of 94 °C for 3 min, 30 cycles of 94 °C for 1 min, 55 °C for 1 min, 72 °C for 2 min, and one cycle of 72 °C for 10 min. The resulting PCR product was gel-purified and cloned into pGEM®-T Easy vector. The product was cut out with BamHI and EcoRI from the plasmid and transferred into pGEX-2T (Amersham Biosciences). The recombinant protein was expressed using the conditions recommended by the manufacturer. The whole-cell extract from the induced cells was used for Western-blot analysis.

Preparation of anti-SCP2 antibodies

To produce the recombinant SCP2 domain of SCPx protein, the corresponding cDNA region (positions 1381–1671; Figure 1A below) was amplified by PCR with the following gene-specific primers: SCP2F (5'-GGA ATT CGT GAT GGT GAT ATG ATG CAG AAC AGG AA) and SCP2R (5'-GGA ATT CGT GAT GGT GAT GAT GAT GAT CCA GCA TTG AGC GGA TTG T). The resulting PCR product was gel-purified and cloned into pGEX-2T Easy vector. The product was cut out with BamHI and EcoRI from the plasmid and transferred into pGEX-2T (Amersham Biosciences). The recombinant protein was expressed using the conditions recommended by the manufacturer. The GST (glutathione S-transferase)–SCP2 fusion protein was purified with glutathione–agarose (Sigma). The purified protein was digested with thrombin (Cloning of cotton-leafworm sterol carrier protein 2/3-oxoacyl-CoA thiolase; see Figure 2 below) and the SCP2 domain for 2 h. After washing with TBST solution, the membranes were incubated with horseradish-peroxidase-conjugated goat anti-rabbit IgG (Bio-Rad, 1:10000 dilution) for 1 h. The membranes were washed and developed with ECL Plus® Western-blotting detection reagents (Amersham Biosciences).

Semi-quantitative real-time PCR

Quantification of SCPx expression in prothoracic glands derived from different times within the sixth instar was performed by an ABI Prism 7700 sequence detector (Applied Biosystems), a real-time PCR machine, using a TaqMan® probe for target DNA quantification. For this, batches of seven prothoracic glands were homogenized using a QiA shredder homogenizer (Qiagen), and RNAs were prepared using the RNeasy Mini kit (Qiagen). First-strand cDNA synthesis was performed using the Reverse-itTM first-strand synthesis kit (ABgene, Epsom, Surrey, U.K.). The SCPx-specific primers used were 5'-AGG CAT CTA CGG ATT CAA AGT GAA (forward), 5'-CGC TGC TAT TGT ATG TGA CTT TGC (reverse) and 5'-FAM-CCC CAA CGG CCG GGA AGG-TAMRA (TaqMan® probe). RP49 (ribosomal protein 49)-specific primers were used as an internal control during quantification. RP49-specific primers used were 5'-ATC CTG ATG ATG CAG AAC AGG AA (forward), 5'-GAT GGT CTT GCG CTT CTT TGA G (reverse) and 5'-FAM-CCG AGA GAT GCC CGA AGG-TAMRA (TaqMan® probe). RP49 was co-amplified with the nuclear 18S rRNA fragment, and the relative expression changes were normalized comparing target mRNA levels to the control. The quantitative PCR was performed twice, with three replicates for each.

RESULTS

Cloning of the cDNA encoding SCPx

Using a combination of column-chromatographic approaches as described previously [15], we purified SCPx from *Spodoptera* midgut. SDS/PAGE analysis revealed that the apparent molecular mass of the purified SCPx was approx. 46 kDa and provided the material to obtain the N-terminal amino acid sequence.

A PCR-based cloning strategy, as detailed in the Experimental section, allowed us to obtain a cDNA fragment corresponding to the sequences between positions 70 and 1942 (Figure 1A). Gene-specific primers derived from this sequence were synthesized and used for 5'-RACE to obtain the 5'-end of the cDNA. 5'-RACE produced a cDNA clone of 285 bp, which contained a putative translation start site at position 67. Taken together, all overlapping cDNA span a total of 1422 bp. The polyadenylation signal is located at position 1907. As shown in Figure 1(A), using the first ATG as the start codon, the full-length SCPx cDNA encodes a protein of 535 amino acids with a predicted molecular mass of 57 450.13. The result of *in vitro* translation confirmed this prediction (Figure 1B). Comparison of this value with that of the
H. Takeuchi and others

Figure 1  Nucleotide and deduced amino acid sequence of SCPx

(A) The cDNA sequence is indicated on the top line, and the deduced amino acid sequence is on the second line. The putative polyadenylation signal is double-underlined, and the amino acid sequence obtained from the purified SCPx is underlined. An XbaI restriction site is shown in a box. Numbers on the right refer to the last amino acid residue on each line of the respective protein sequences. Numbers on the left refer to the first residue on each line of the respective nucleotide sequences.

(B) In vitro translation of SCPx. Lane 1, no template; lane 2, DNA. The fragment containing the open reading frame of SCPx was used as template.

purified protein (approx. 46 kDa), suggests that post-translational processing of the protein may occur. The protein is mildly basic, with an estimated pI of 8.02.

Similarity of deduced amino acid sequence to the proteins of the SCP2 gene family

The deduced amino acid sequence of the coding region showed similarity to members of the SCP2 gene family (Figure 2) such as D. melanogaster SCPx (63%), C. elegans P-44 (55%), Oryctolagus cuniculus (rabbit) SCPx (55%), Homo sapiens (human) SCPx (54%) and Mus musculus (mouse) SCPx (55%).

Tissue distribution and developmental expression of SCPx

As demonstrated in Figure 3, a cDNA probe representing the protein coding region detected two transcripts of approximately 1.9 and 0.95 kb, although the 0.95 kb mRNA was expressed at a much lower level than the 1.9 kb transcript. Presumably, these mRNAs encode SCPx and SCP2, since the probe corresponding to the SCP2 domain (1357–1671) was also able to hybridize to both bands (results not shown). The expression level of SCPx mRNA was detected during all stages of the instar, but was higher during the feeding stage. The mRNA was readily detected at 20 h into the last larval instar and was strongly expressed until 48 h. It started to decrease from 66 h, and its expression level became very low before pupation.

Western-blot analysis showed that protein levels reflected the transcription of the SCPx gene (Figure 4B, top panel). The major band detected by the antibody against C. elegans P-44 protein is approx. 46 kDa, and corresponds well to the molecular mass of the protein originally purified and suggests that SCPx is post-translationally processed.

Genomic structure of SCPx

Genomic DNA was prepared from combined fat body and Malpighian tubules and digested with various restriction enzymes, including XbaI, which has a cleavage site in the SCPx cDNA (at position 667; Figure 1), whereas BamHI, BgIII and EcoRI do not.

© 2004 Biochemical Society
Molecular cloning of cotton-leafworm sterol carrier protein 2/3-oxoacyl-CoA thiolase

Figure 2 Alignment of the deduced amino acid sequences of SCPx and the most similar database proteins

The deduced sequence of SCPx from S. littoralis (SCPX_Spod) was compared with all sequences in the database using the BLAST program. Only the amino acid sequences of the most similar proteins in the SCP2 gene family are shown in the alignment. Gaps introduced to optimize alignment are indicated by hyphens. Identical amino acids between SCPx from S. littoralis and at least one other sequence are indicated in black boxes. Numbers on the right refer to the last amino acid residue on each line of the respective protein sequences. The accession number and the percentage similarity to SCPx of S. littoralis (SCPX_Spod) of each protein are as follows: D. melanogaster SCPx (SCPX_Dro), AE003660, 63%; C. elegans P-44 (P44_Nem), D86473, 55%; rabbit SCPx (SCPX_Rabbit), AF051897, 55%; human SCPx (SCPX_Human), M75883, 54%; mouse SCPx (SCPX_mouse), M91458, 55%. The alignment was constructed by use of the CLUSTALW program.

Figure 3 Northern-blot analysis of the tissue distribution and expression of SCPx

A 10 µg aliquot of total RNA from various tissues at different times (top) within the last larval instar was used. The blot was hybridized with a 32P-labelled SCPx cDNA probe corresponding to the whole protein-coding region. The same blot was stripped and reprobed with 18 S rRNA probe to normalize sample loading. The positions of RNA size markers are shown on the left.

The resultant DNA fragments were analysed by Southern blotting using a probe representing the coding region of the cDNA (67–1674; Figure 1). As shown in Figure 7(A), more than one band was observed in BglII-, EcoRI- and XbaI-digested samples. However, when the same blot was re-probed with a cDNA fragment corresponding to the N-terminus (67–667; Figure 1), only one band was observed in BamHI-, BgIII- and EcoRI-digested DNA (Figure 7B). This result can be most simply interpreted if SCPx is encoded by a single gene in the S. littoralis genome. However, this point would need to be confirmed by performing a quantitative Southern blot, but this would require knowledge of the genome size of S. littoralis. Multiple bands detected using the full-length coding region may be due to hybridization to other members of the SCP2 gene family, but could also be caused by polymorphism, since this genomic DNA was prepared from ten individuals. The two bands in the XbaI digests in Figure 7(B) can only be due to polymorphism of the gene.

DISCUSSION

Using a reverse-transcriptase PCR-based cloning strategy, employing degenerate primers designed on the basis of the partial amino acid sequence of the purified protein, together with 5′-RACE, the complete cDNA encoding SCPx was isolated and sequenced. The full-length SCPx cDNA encodes a protein of 535 amino acids. In fact, the in vitro expression showed that this cDNA expressed an approx. 57 kDa product (Figure 1). In humans, SCPx mRNA encodes a fusion protein, and the 58 kDa product is post-translationally processed to the 46 kDa thiolase and the 13 kDa SCP2. In contrast, the mRNA encoding P-44, which is homologous with SCPx in C. elegans, lacks the SCP2 domain near the 3′-end [14]. The situation concerning regulation of SCPx in Drosophila is intermediate between that of mammals and C. elegans. Drosophila produces two types of mRNA encoding SCPx: a fused mRNA and a SCP2-lacking
mRNA, the latter being the major transcript in the larval stage [20]. In the case of *S. littoralis*, Northern blotting using the full-length coding region of SCPx detected two transcripts of approx. 1.9 and 0.95 kb (Figure 3), but the short transcript, which presumably encodes SCP2, was not detected at a much lower level than the 1.9 kb transcript (Figure 3). The SCPx mRNA lacking the SCP2 domain was not detected using Northern blotting or 3′-RACE (results not shown). On the other hand, the antibody against the 3-oxoacyl-CoA thiolase domain could detect a product of approx. 46 kDa (Figure 4B). By contrast, the antibody against the SCP2 domain could detect a product of approx. 13 kDa not only in vivo but also in *E. coli* cells which expressed the recombinant SCPx protein (Figure 5). These data suggest that *S. littoralis* SCPx is mainly transcribed as a fusion mRNA and the translation product is post-translationally cleaved into two proteins: 3-oxoacyl-CoA thiolase and SCP2. Accordingly, the lepidopteran system may be closer to that of mammals rather than that of dipterans.

Database searching revealed that the gene product has significant similarity to products of other members of the SCP2 gene family (Figure 2). Most importantly, a cysteine residue (Cys82), which has been suggested to be part of the active site in a bacterial (*Zoogloea ramigera*) 3-oxoacyl-CoA thiolase [21], is conserved in *S. littoralis* SCPx. Also, this protein contains Ser-Lys-Leu, a putative peroxisome targeting motif, at the end of the C-terminus. This means that SCPx can be transported to peroxisomes and processed in a similar manner to other species.

Northern blotting also revealed that the 1.9 kb mRNA is highly expressed in midgut and Malpighian tubules during the feeding stage of the last larval instar. In *C. elegans*, P-44 is mainly expressed in the larval intestine [19], whereas in *Drosophila*, SCPx
is highly expressed in midgut epithelium during late embryogenesis [20]. These expression patterns are very similar to that of  

*S. littoralis*. These data suggest that SCPx may be involved in lipid transport and metabolism in the gut of invertebrate species. In particular, gut plays an important role for absorption of fatty acids and cholesterol in mammals, and most invertebrate species cannot synthesize sterols and depend on a dietary source. The high expression level in midgut is very consistent with this role. The lipid metabolism may include that of sterol/steroid, since the midgut is a major site of sterol dealkylation in insects [22] and the midgut and Malpighian tubules are also known to be tissues which strongly express steroid-metabolizing enzymes [23–26].

Warren and Gilbert [27] suggested that SCP2 may be involved in the transport of 7-dehydrocholesterol from the endoplasmic reticulum to the mitochondrial outer membrane during the early steps of ecdysteroid synthesis in the prothoracic glands. Our demonstration by semi-quantitative PCR that SCPx is constitutively expressed in prothoracic glands (Figure 6) lends support to this tenet. Furthermore, a recent report revealed that SCP2 from the yellow-fever mosquito (*Aedes aegypti*) is also highly expressed in the tissues which are involved in ecdysteroid synthesis [28]. These data suggest that this protein may have a role in sterol transport, including that associated with ecdysteroid synthesis.

The work was supported by a grant from the BBSRC (Biotechnology and Biological Sciences Research Council). We thank Mr S. G. Corrigan for culture of *S. littoralis* and Ms Andrea Davies for help with the semi-quantitative PCR experiments.

REFERENCES


© 2004 Biochemical Society


Received 30 April 2004; accepted 18 May 2004
Published as BJ Immediate Publication 18 May 2004, DOI 10.1042/BJ20040717