Sequencing and functional expression of the malonyl-CoA-sensitive carnitine palmitoyltransferase from *Drosophila melanogaster*

Vicky N. JACKSON, Jacqueline M. CAMERON, Victor A. ZAMMIT and Nigel T. PRICE¹

Cellular Biochemistry Group, Hannah Research Institute, Ayr, KA6 5HL, U.K.

Using expressed sequence tag data, we obtained a cDNA for a carnitine palmitoyltransferase I (CPT I)-like molecule from *Drosophila melanogaster*. The cDNA encodes a 782-residue protein that shows 49 % and 48 % sequence identity with the rat liver and skeletal-muscle isoforms of CPT I respectively. The sequence has two predicted membrane-spanning regions, suggesting that it adopts the same topology as its mammalian counterparts. The sequence contains all the residues that have been shown to be characteristic of carnitine acetyltransferases. Expression in the yeast *Pichia pastoris* confirmed that the cDNA does encode a CPT enzyme. The activity was found to be associated with a mitochondria-enriched fraction. Kinetic analysis revealed a K_m for carnitine of 406 μ M and a K_m for palmitoyl-CoA of 105 μ M. The CPT activity was very sensitive to inhibition

by malonyl-CoA, with an IC_{50} of 0.74 μ M when the activity was assayed with 35 μ M palmitoyl-CoA and 1 % (w/v) albumin at pH 7.0. A histidine residue at position 140 in rat liver CPT I has been indicated to be important for inhibition by malonyl-CoA. The equivalent residue (position 136) in *Drosophila* CPT I is arginine, implying that any basic residue might be compatible with such sensitivity. Evidence is presented that, unlike in mammals, *Drosophila* has only a single CPT I gene. Sequences suggesting the existence of a splice variant in the 5' untranslated region were found; this was consistent with the existence of two promoters for the CPT I gene.

Key words: acyltransferase, mitochondria, *Pichia pastoris*.

INTRODUCTION

One of the most powerful tools for investigating structure–function relationships within an enzyme is the comparison of the protein sequences of isoforms from different species. We are interested in carnitine palmitoyltransferase I (CPT I), which is the rate-limiting enzyme for the oxidation of long-chain fatty acids by mitochondria. CPT I is a polytopic integral membrane protein of the mitochondrial outer membrane. It is a member of a large family of acyltransferases; however, no three-dimensional structures are currently available for any of the members. CPT I is closely related to CPT II, which catalyses an identical reaction but in the reverse direction, and is present on the matrix side of the inner mitochondrial membrane. The two CPTs act in concert with a carnitine acylcarnitine translocase to transfer long-chain fatty acids into mitochondria via an acylcarnitine intermediate. CPT I is distinct from CPT II and the other known acyltransferase family members in that it possesses a large N-terminal extension of approx. 140 residues. This region contains two transmembrane (TM) domains that anchor the molecule within the outer mitochondrial membrane. Its topology results in the exposure of both the short (47-residue) N-terminal domain and the much larger (650-residue) C-terminal domain to the cytosolic side of the membrane. The N-terminus is essential for the expression of sensitivity of CPT I to inhibition by malonyl-CoA, a regulatory characteristic that gives CPT I a central role in nutrient partitioning within the cell.

In mammals CPT I exists in at least two isoforms: L-CPT I and M-CPT I (also referred to as the A or α , and B or β , isoforms), characteristic of the forms found in the liver and skeletal muscle respectively. (There are also two additional splice variants for M-CPT I [1].) The amino acid sequences of L- and

M-CPT I are highly similar (62 % identity). Although the two isoforms catalyse identical reactions, they have very important differences in their kinetic properties. M-CPT I is more sensitive (lower IC_{50}) to inhibition by malonyl-CoA and has a higher K_m for carnitine.

For CPT I, protein sequence comparisons are limited by the present availability of sequences from human, rat and mouse only. These sequences are highly similar, so very little can be deduced about essential conserved residues. In particular, the L- and M- isoforms show 86 % identity with the rat and human counterparts. Such comparisons ideally require sequences of the same enzyme from more distantly related species. The availability of the complete *Saccharomyces cerevisiae* genome is normally one useful source of sequences. However, this unicellular yeast does not possess mitochondrial CPT activity. Indeed, this fact is exploited in the use of the yeasts *S. cerevisiae* [2,3] and *Pichia pastoris* [4], and present study) as suitable hosts for the heterologous expression of CPT I without the presence of interfering endogenous activity.

We therefore searched for CPT I-like sequences within the expressed sequence tag (EST) and genomic sequence data available for a number of model organisms, including the recently completed nematode *Caenorhabditis elegans* genome. As expected, no CPT I-like sequences could be identified in the *S. cerevisiae* genome. However, we were able to identify possible CPT I isoforms from *C. elegans*, zebrafish and the fruit fly *Drosophila melanogaster*.

Here we describe the use of the partial sequence data available to obtain the full sequence of *Drosophila* CPT I. Functional expression in the yeast *P. pastoris* demonstrates that the cDNA does indeed encode a mitochondrially targeted malonyl-CoA-sensitive CPT (overt CPT in the nomenclature of [5]).

Abbreviations used: CPT, carnitine palmitoyltransferase; EST, expressed sequence tag; IMS, intermembrane space; L-CPT and M-CPT, liver and muscle isoforms of CPT I respectively; TM, transmembrane.

¹ To whom correspondence should be addressed (e-mail pricen@hri.sari.ac.uk).

The nucleotide sequence data reported will appear in DDBJ, EMBL and GenBank Nucleotide Sequence Databases under the accession number AJ010150.

EXPERIMENTAL

Materials

Materials were obtained from Sigma (Poole, Dorset, U.K.) except where stated otherwise. The *P. pastoris* expression system and zeocin antibiotic were from InVitrogen (Groningen, The Netherlands). The TNT coupled transcription/translation reticulocyte lysate system and the Prime-a-Gene random primer labelling kit were supplied by Promega (Southampton, Hants., U.K.). L-[³⁵S]Methionine (*in vitro* translation grade), [α -³²P]dCTP and zymolase 100T were from ICN (Thame, Oxon., U.K.). Molecular biology enzymes were from Promega or New England Biolabs (Hitchin, Herts., U.K.). Palmitoyl-CoA and L-[methyl-³H]carnitine were obtained from Amersham Pharmacia Biotech (Amersham, Bucks., U.K.).

Database searching

The Genbank and dbEST databases were searched by using the BLASTN algorithm (nucleotide versus nucleotide) or the TBLASTN algorithm (protein sequence against nucleotide sequence translated in all six reading frames) [6].

Translation and sequencing of *Drosophila* CPT I cDNA *in vitro*

The EST clone containing the longest cDNA (BDcln15843, accession number AA441641) was obtained from the distributors Genome Systems (St. Louis, MO, U.S.A.). The putative *Drosophila* CPT I cDNA had been cloned into the *Eco*RI and *Xho*I restriction sites of pBluescript (SK). Analysis of the size of the cDNA after excision from the cloning vector suggested that the molecule was large enough to encode at least 770 residues (the size predicted from mammalian CPT I). Coupled transcription/translation *in vitro* in the presence of [³⁵S]methionine was performed to determine the size of the protein produced. Plasmid DNA [pBluescript (SK) containing the *Drosophila* cDNA, or rat L-CPT I cDNA] (0.5 μ g) was used in conjunction with T3 RNA polymerase as recommended by the manufacturer. The translation products were analysed by SDS/PAGE followed by autoradiography. The *Drosophila* cDNA was found to encode a protein of very similar mobility to rat L-CPT I (results not shown). The sequence of the cDNA insert was determined on both strands by using a vector/custom primer walking strategy.

DNA manipulations

The insert was excised from pBluescript with *Eco*RV (which cuts at nt 206 within the cDNA itself, leaving only 20 nt of 5' untranslated region) and *Xho*I, and inserted into the *P. pastoris* expression vector pGAPZ B at the *Pml*I and *Xho*I sites.

Functional expression in *P. pastoris*

All procedures involving *P. pastoris* were performed as described by the distributor of the expression system (InVitrogen). The expression vector used was pGAPZ, which contains the constitutively active promoter from the glyceraldehyde-3-phosphate dehydrogenase gene and a zeocin resistance gene. Recombinant plasmid was linearized with *Avr*II and transformed into *P. pastoris* wild-type strain X-33 by electroporation. Stable integration into the yeast genome was selected by using 0.1 mg/ml zeocin. Assays were performed on extracts prepared from yeast grown for 24 h at 30 °C in YPD medium [1% (w/v) yeast extract/2% (w/v) peptone/2% (w/v) dextrose] with shaking at 300 rev./min in baffled flasks. Extracts were made either by

breaking cells with glass beads or by the preparation of mitochondrial fractions from larger-scale cultures.

Preparation of extracts

Yeast cells were harvested and washed three times in one (culture) volume of distilled water. Extracts were prepared by vortex-mixing with acid-washed glass beads in 0.1 vol. of buffer [50 mM sodium phosphate/1 mM EDTA/5% (v/v) glycerol (pH 7.4)] containing protease inhibitors (complete protease inhibitor tablets; Boehringer Mannheim, Lewes, Sussex, U.K.), as described in the InVitrogen manual. After centrifugation (5 min at 2000 g), supernatants were either snap-frozen and stored at -70 °C or assayed immediately. Extract (25–50 μ l) was used in a 1 ml assay for CPT activity. No difference in maximal activity or kinetic characteristics was found between fresh and frozen extracts (results not shown). In some cases, mitochondria-enriched fractions were prepared from approx. 10 g wet weight of cells resulting from 24 h growth in YPD medium. Cells were broken after treatment with zymolase and mitochondria were prepared essentially as described [7]. Mitochondrial fractions were resuspended in 2.5 ml of buffer [250 mM sucrose/10 mM Hepes/1 mM EGTA (pH 7.0)]. Protein concentrations were determined with bicinchoninic acid [8].

Assay of CPT activity

CPT activity was assayed radiochemically with L-[methyl-³H]-carnitine hydrochloride (770 d.p.m./nmol) in the direction of palmitoylcarnitine formation. The standard assay buffer contained 150 mM KCl, 1 mM EGTA, 20 mM Hepes, 5 mM MgATP, 1% (w/v) dialysed defatted BSA, 1 mM dithiothreitol, 4 μ g/ml rotenone and 2 μ g/ml antimycin A, pH 7.0. Reactions were performed in duplicate in a final volume of 1 ml at 30 °C, with a 2 min preincubation of mitochondria/cell lysates before the addition of L-[³H]carnitine. Reactions were incubated for 2–10 min, ensuring that reaction rates remained linear. The assay was terminated by the addition of 0.3 ml of 6 M HCl and transfer to ice. After extraction into butan-1-ol at low pH as described previously [9], [³H]palmitoylcarnitine was determined by liquid-scintillation counting.

For determination of the K_m for carnitine, palmitoyl-CoA was included at 135 μ M and the final concentrations of [³H]carnitine ranged from 10 to 1000 μ M. For determination of the K_m for palmitoyl-CoA and IC_{50} values for malonyl-CoA, [³H]carnitine was used at 0.5 mM. For palmitoyl-CoA assays, the concentration of palmitoyl-CoA ranged from 10 to 300 μ M. To measure inhibition by malonyl-CoA, the palmitoyl-CoA concentration was fixed at 35 μ M and the malonyl-CoA concentration was varied between 0.5 and 150 μ M (*Drosophila* and rat M-CPT I) or 300 μ M (rat L-CPT I).

Data were fitted to Michaelis–Menten and simple competitive inhibition curves with SIGMAPLOT software (SSPS, Chicago, IL, U.S.A.).

Probing of Southern blot

D. melanogaster (OregonR strain) genomic DNA was a gift from Kathleen Rothwell (Institute of Cell and Molecular Biology, University of Edinburgh, Edinburgh, U.K.). The DNA was digested with four different restriction enzymes: *Not*I, which has an 8 bp recognition sequence, and *Afl*II, *Xba*I and *Xho*I (6 bp recognition sequences). The digested DNA was subjected to electrophoresis on a 0.8% (w/v) agarose gel and transferred to Hybond-N membrane (Amersham). After depurination, denaturation and neutralization as recommended by the membrane

manufacturer, the DNA was cross-linked to the membrane by UV irradiation. After prehybridization, the membrane was hybridized with a ^{32}P -labelled probe prepared by random priming of DNA corresponding to the first 852 nt of cDNA1 (see Figure 1). After hybridization, the blot was washed with $2 \times \text{SSC}$ [$1 \times \text{SSC}$ is 0.15 M NaCl/0.015 M sodium citrate (pH 7.0)] containing 0.05 % SDS at room temperature, followed by $1 \times \text{SSC}$ containing 0.1 % SDS for 15 min at 65 °C. The hybridized probe was detected with a PhosphorImager (Molecular Dynamics).

RESULTS AND DISCUSSION

Cloning of *Drosophila* CPT 1

To identify possible isoforms of mammalian CPT I from other species, only the N-terminal 150 residues of rat L- and M-CPT I (which distinguish CPT I from other members of the acyltransferase family) were used for initial database searching. Nucleotide sequences from the 5' ends of several ESTs encoding a candidate *Drosophila* CPT I were identified. These assembled sequences were all found to derive from the same species of cDNA. Further overlapping shorter EST clones (which were not detected by the original search as this was limited to the N-terminus of mammalian CPT I) were found by using the assembled consensus nucleotide sequence. In all, 5' sequences from 16 overlapping clones were identified, covering approx. 1037 nt (which later analysis showed to constitute approx. 35 % of the total number of nucleotides in the cDNA). Fourteen of the clones were from a 0–24 h mixed-stage embryo (LD) library, with the remaining two clones from an adult head (HL) library, both libraries resulting from the Berkeley *Drosophila* Genome Project [D. Harvey, L. Hong, M. Evans-Holm, J. Pendleton, C. Su, P. Brokstein, S. Lewis and G. M. Rubin (1997), unpublished GenBank annotation].

To determine the full sequence of the *Drosophila* CPT I cDNA, the EST clone containing the longest cDNA (as judged from the 5'-end sequence) was obtained. The sequence of clone BDcln15843 (accession number AA441641) was determined on both strands.

Analysis of sequence

The nucleotide sequence and deduced protein sequence are shown in Figure 1. The cDNA has 2882 nt encoding a predicted 782-residue protein. There are three upstream in-frame stop codons, suggesting that the indicated AUG codon, which lies within a good consensus according to [10], is the real start codon. This is also supported by the fact that this predicted open reading frame shows strong similarity to the mammalian CPT I sequence. The cDNA does not contain the consensus polyadenylation site AATAAA, although at its 3' end there are 17 adenine residues [distinct from those derived from the oligo(dT) primer used in cDNA synthesis] as evidence of a poly(A) tail.

The *Drosophila* protein sequence has all the motifs that are used as characteristic of acyltransferase family members in the Prosite database, except that Cys-458 within signature 2 (Prosite accession number PS00440 [11]) is at a position that is occupied by Leu, Ile, Val, Met, Phe or Tyr in all other family members. This residue lies close in the primary sequence to a conserved histidine (His-471 in *Drosophila* CPT I), which is thought to be at the active site [12]. As *Drosophila* CPT I is clearly a member of the acyltransferase family, as evidenced by its functional expression (see below), this signature needs to be revised to take account of this discrepancy.

Several residues that are important in acyltransferases have been identified by Cronin. The motifs STS [13] and $\text{G}(\text{X})_3\text{D}(\text{X})_5\text{L}$

[14,15] were found to be conserved between carnitine acetyltransferase and choline acetyltransferase. All the carnitine acetyltransferases have a conserved arginine residue within the latter motif, i.e. $\text{G}(\text{X})_3\text{DR}(\text{X})_5\text{L}$. Other conservative differences were found to distinguish choline acetyltransferase from carnitine acetyltransferases [14]: the sequences $\text{VD}(\text{N}/\text{C})$ and $\text{T}(\text{E}/\text{D})\text{T}$ respectively are likely to be involved in catalytic discrimination. In *Drosophila* CPT I the equivalent sequences are TET (residues 599–601), and the conserved Arg residue within the $\text{G}(\text{X})_3\text{D}(\text{X})_5\text{L}$ motif is also present (residues 647–657; GrgiDRhlfcl), which is consistent with its role as an acyltransferase using carnitine as an acceptor molecule. The STS and $\text{YE}(\text{X})_{10}\text{R}$ motifs are also conserved (residues 681–683 and 586–598 respectively).

The predicted *Drosophila* CPT I protein is very similar in size (782 residues) to mammalian CPT I (772 or 773 residues). The additional residues in the *Drosophila* enzyme are largely accounted for by a 12- or 13-residue C-terminal extension, as shown by the protein sequence alignment in Figure 2. The protein has two predicted TM helices with a connecting loop of approximately the same length as in mammalian isoforms of CPT I. The topology determinants discussed in [16] are also present, suggesting that the protein has the same topology as mammalian CPT I. Arg-51 and Arg-119 might define the cytosolic proximal ends of the TM domains. However, unlike in mammalian CPT I forms, no obvious charged residues can be postulated to define the intermembrane space (IMS) ends.

The *Drosophila* CPT I protein sequence shows 49 % and 48 % identity with rat L- and M-CPT I respectively (compared with 62 % identity between rat CPT I forms). The extreme N-terminus is highly conserved: 10 of the first 12 residues (which are entirely conserved in both isoforms of mammalian CPT I) are identical. As expected from the lack of conservation between CPT I from different mammalian species and between isoforms, the extreme C-terminus, the IMS loop region and the IMS proximal ends of the predicted TM domains are poorly conserved. Only one of the two proline residues present in TM1 of all mammalian CPT I forms is present in the *Drosophila* sequence. The C-domain of *Drosophila* CPT I (the portion C-terminal to TM2), which is thought to contain the active-site residues essential for catalytic activity, does not seem to resemble more closely either of the two mammalian isoforms. Therefore it is not possible to predict the kinetic behaviour of the molecule with respect to its substrates from an inspection of the primary sequence.

Throughout the sequences of mammalian CPT I forms there are 15 invariant histidine residues. Four of these are non-conserved in *Drosophila* CPT I (Lys-129, Arg-136, Asn-501 and Tyr-575). These four positions are also non-conserved in mammalian CPT II (results not shown), suggesting that they are not essential for catalytic activity. However, the first two of these (residues 129 and 136) lie in a location thought to be important for malonyl-CoA sensitivity, as discussed below.

Expression of CPT activity

To verify that the sequence did indeed code for a carnitine palmitoyltransferase, the cDNA was expressed in the methylotrophic yeast *P. pastoris*. Initial assays with cell extracts prepared with glass beads from small-scale cultures revealed readily detectable CPT activity. No such activity could be detected in the non-transformed host, or in host transformed with empty vector (results not shown).

Targeting to mitochondria

To investigate whether *Drosophila* CPT was targeted to mitochondria, larger-scale cultures were used to prepare mito-

```

1      gcaacgcttacacgcttaacattgcgcgcgctctttgeattttct
47     tgtttgcgttcgttcgcttcgcttcctcccggtgtgtcttattgtttttcccgctttgtgcgctctttattgtttccgcttttata
137    gcatttgcgatttgaagtgaggactgcaattgtttaccatttcaagaacacCTTTTCTGTACGTGGGATATCATTAGCAACCTACAGGT

227    ATGCCCGAAGCCCATGCCGCGTGGCATTCTCGTTTGGCATCACCCATGAGGCGCTTCGACATCAACTATGACCACGAGGTGCTCAATTG
      M A E A H A A V A F S F A I T H E G F D I N Y D H E V L N L 30

317    GTGTGGAAGTCAAGAGTTCTGTTCTGAAAAAGCATTGGCTCGCGCAAGGAACGGGTACGCAATGGCGTCTATCCGGCTCACATCCAG
      V W N S G V R S W K K R L A R A R N G V R N G V Y P A H I Q 60

407    AGCCTCTGGTIGATCTCAGCAATAGCGCTGGGTCTACACTTTCCGGGCTACAGGCTCCATTCAATCTGACCAACCGTATACTGGTGCAC
      S L W L I S A I A L G L H F A G Y Q A P F N L T N R I L V H 90

497    CTGCGTTCCAACACGATAACTGGCAGGTTACCGCTGCTTTTGGCGGCTCTGGTCTGGCTATCGATCTGTTTCCACCATGAGATAC
      L P S N T I N W Q V T A C F L A A L V V W L S I C F T M R Y 120

587    ACTCTCAAGTGTCTGCTCATGTACAAGGGCTGGATGTACGAGTCCAGAGCTCCCGGTAGTCTGTCTCTGCCCACCATGTTGTGGGTG
      T L K L L L M Y K G W M Y E S R A P G S R V S L P T M L W V 150

677    CGACTGGTGGGGTCTATCTAGTGGAAACACCGGGCTCTGACAGCTTCCAGGGTCTCTGCCAGGCTGCCACTGCCCTCCGTAA
      A V V R V L S S W N K P G L Y S F Q G S L P R L P L P S V A 180

767    GACACGATGACAGGTAAGTTCGCGAGTGTTCGACCGCTGCTTGATGATGAGAACTACACGCGCATGGAGCGCTGCCAAAGAATCGAA
      D T M T R Y L R S V R P L L D D E N Y T R M E R L A K E F E 210

857    CAGACCATCGGCAAGAGTCTCAGTGGTACCTCATCCTCAAGAGCTGGTGTCCACCAACTACGTGTCGGATTGGTGGGAGGAGTACGTC
      Q T I G K K L Q W Y L I L K S W W S T N Y V S D W W E E Y 240

947    TACCTCTGTTGGTTCGAGTCCCTCTGCGTGAACGCAACTTCTACGGCACGGATCGCATCTTCATGAACCTACCGACAAACAGCGGCG
      Y L R G R S P L C V N S N F Y G T D A I F M N L T D K Q A A 270

1037   CGTGGCGGCAATGTGATTCCCTGCTCTCAACTTCGCTGAGCTATCGAGCACCAGGAAGTGCAGCTATCATGTTTCAGGCGATGATT
      R A A N V I S L L L N F R R L I E H Q E L Q P I M V Q G M I 300

1127   CCTCTGCTGCTTGGCAGTATGAACGTACCTTCAACACAGCAGCGGTGCCAGGCTGGAAACCGATCGTATCATCCACTACAGGACTCC
      P L C S W Q Y E R T F N T A R V P G L E T D R I I H Y R D S 330

1217   AATCACATTGTGGTCTCCACAAGGGCTGCTACTACAAGATGCTCATCTACTACAAGGGTGCATCTCGGCCATCGCAATTGCAAGTG
      N H I V V L H K G C Y Y K M L I Y Y K G R I L R P C E L Q V 360

1307   CAAATTGAGGAATCCTGAAGGGCAAGCAACTCCCGTGGAGGTGAGGAGCACTTGGCTGCCCTGACCGCTGGAATCGCTCCAAGTGG
      Q I E E I L K G K A T P V E G E E H L A A L T A W N R S K W 390

1397   GCGGAAGCTCGCAACACATTCTTCTGTTGGGAGTCAATCAAACTCTCTGCGCACCATCGAGTGGCGGCTTTTGTCTCTCCCTGGAT
      A E A R N T F F S W G V N Q T S L R T I E S A A F V L S L D 420

1487   GATGAGCCCTTCGAGTTCGATTGGCGGACCGAATTGCTGGCAACTTTGGCAAGAGCTGTCGACGGCAATGGCTACAACCGCTGG
      D G P F E F D L A R P E L L D N F G K K L L H G N G Y N R W 450

1577   TTTGACAAGTGTTCCTGCTGTGTGGGACCAATGGTGGCTTCAATGCCGAACACCTGGTCCGACGCTGCCATCGCGTCA
      F D K C F T V C V G T N G R V G F N A E H T W S D A A I A S 480

1667   CACATGTGGGAGAACCTGATCGTCTGATCTTGTATCGACGATATGACGAACGGGTATATCAAGGGCACCCGGCTTCAGCCA
      H M W E N L I V D D L V S D G Y D E T G N T K G T P A F Q P 510

1757   CCCACGCCCTAGGCTCACTTGGGATCTTAAACCATGCCCTGGCGAGATCGAAGAGGCCACATCGATGTGACGAAGCTTATCAACGAG
      P T P T R L T W D L K P C L A Q I E E A T I D V T K L I N E 540

1847   GTCATCTGCGCATCTGTTGTCACCGAGGATAGCGAAGGGCTTTATGAAGAGTGGCGCATCTCACCAGATGCCATACAAATGGCC
      V N L R I L V H Q D Y G K G F M K K C R I S P D A Y I Q M A 570

1937   CTGCACTTGGCTTACTACCGAGATCCCGCTCGCTTCTGTTGACATATGAGGCTTCAATGACCCGCTGTTCGCGAAGGAAGAACTGAG
      L Q L A Y Y R D A G R F S L T Y E A S M T R L F R E G R T E 600

2027   ACAGTGGTCCGTGCAACATTGAGTCAATCCGCTGGGTCAAGGCAATGCAAGCCAAATACAAACACGATGAGCGTGTGAAGATGATG
      T V R P C T I E S S A W V K A M Q N P N T T N D E R V K M M 630

2117   CAGSGCTGCTGCGATCGTACCGAGTTGGGCTACAGGATGCCATGTGCGGTCTGGCATCGACAGGCATCTGTCTGCTGTATGTGGTC
      Q A A C D R H Q L G Y Q D A M C G R G I D R H L F C L Y V V 660

2207   TCCAAGTACCTGGAGTGGATTACCTTCTCAACGAGGTGCTCAGCAACCTGGCGCTCTCTACTAGTCAGACTCCGCGACGGCCAG
      S K Y L E V D S P F L N E V L S E P W R L S T S Q T P H G Q 690

2297   ACGCCCAAGATGGACCTGAAGAAGCATCCGAACATGCATCAGCGCTGGAGGCGGCTTTGGGCGGTCGGCGATGATGGTTACGGTGTGTCG
      T P K M D L K K H P N C I S A G G G F G P V A D D G Y G V S 720

2387   TACATTATTGCCGAGAGAACCTGATATTCTTCATATCTCAGCGAAGACTACGTGCCAGCAACGGATGTGCACCGCTTCGCTCAGAAC
      Y I I A G E N L I F F H I S A K T T C Q Q T D V H R F A Q N 750

2477   ATATCGCAGGCTCTGTCAGACATTCGAGCATGTTGAGCAGCAGCATGAAGGACCATCCGAAGCCCGCAATCACTACAAACGGCAAA
      I S Q A L S D I R S M F E Q H M K D H P K P A K S L T N G K 780

2567   TCCACATAATCCCGTTTAGTCCATGTAGTCGAGATTCTTGTAGCAITTTGTTTACGCCCCAGTTGGTCTTTAAACACATCTGATA
      S T * 783

2657   CATCCGCTAGTTCGAAATGAAGTGCAAAAGTCGTCGTGAGAGTTATACATAACAATGCAATTTTATAAACAATTGTGACTTGGCTG
2747   GTTTTAAGATTAGTCTTTAGATACCTTTTGCAATATTTGATGTTTCTCAATCAATTAGTCTGTAATCGAAGAAATCGCGTAACCTTAA
2837   GTGCTATAAATATATTTCATTTTAAGCGTAAAAAATAAAAAA 2882

```

Figure 1 Complete nucleotide sequence and deduced protein sequence of *D. melanogaster* CPT I

The nucleotide and protein sequences are numbered at the left and at the right respectively. The upstream in-frame termination codons are shown in bold italics. Nucleotides 1–185, shown in lower case, are unique to this form of the cDNA (cDNA 1) and distinguish it from cDNA 2 (see Figure 5 and the text). The boxed residues indicate the possible locations of TM regions (residues 52–77 and 100–118). The non-conserved histidine residues (Lys-129 and Arg-136) discussed in the text are shown underlined and in bold. Some of the other residues mentioned in the text are also underlined.

Figure 2 Alignment of the sequences of CPT I from *Drosophila* and rat

The protein sequences of *D. melanogaster* (D.m.) and rat CPT I (L and M isoforms) were aligned by using the Clustal W algorithm [24]. Residues that are identical in all three sequences or in the rat isoforms are highlighted. Letters above the *Drosophila* sequence indicate where the *Drosophila* sequence is identical with that of rat CPT I L and distinct from the muscle isoform (L) or vice versa (M). The proposed approximate positions of membrane-spanning domains are also indicated beneath the sequences. Positions of other residues discussed in the text are indicated below the sequences (X).

chondria-enriched fractions. Because no antibody against *Drosophila* CPT I is currently available, and as none of our anti-(rat CPT I) antibodies detected the expressed protein on Western blots, we were unable to use immunoblots to examine the distribution of the enzyme. An assay of these fractions showed that the activity was associated with mitochondria; it therefore seems that *Drosophila* CPT I is targeted to mitochondria when expressed in *Pichia*, as found for mammalian CPT I ([4]; V. N. Jackson, J. M. Cameron, V. A. Zammit and N. T. Price, unpublished work).

Kinetic characterization

The kinetics of the *Drosophila* CPT activity with respect to its substrates carnitine and palmitoyl-CoA was determined with the use of whole extracts or mitochondrially enriched fractions. No difference was found between CPT I assayed in the two different preparations (results not shown). The enzyme has K_m values of 406.4 ± 29.8 and 104.8 ± 8.2 μM for carnitine and palmitoyl-CoA respectively, determined with mitochondria preparations (Table 1). Representative data are shown in Figures 3(A) and 3(B). The K_m for carnitine is intermediate between those of the rat L- and M-CPT I isoforms, whereas that for palmitoyl-CoA is higher than for both rat isoforms (Table 1).

Sensitivity to malonyl-CoA

The malonyl-CoA sensitivity of *Drosophila* CPT I expressed in *P. pastoris* was assayed on whole-cell lysates and mitochondrial fractions. The rapidly prepared crude extracts were also analysed because it was possible that during the isolation of mitochondria, proteolysis of the N-terminus of the expressed CPT I protein

Table 1 Kinetic properties of CPT I from *Drosophila*, and L- and M-CPT I from rat, expressed in *P. pastoris*

Assays were performed on mitochondria-enriched fractions, except for IC_{50} values for malonyl-CoA for which crude cell extracts were used for reasons discussed in the text. Assay conditions were as given in the Materials and methods section. Results are means $\pm$ S.E.M. for the number of individual experiments indicated in parentheses.

Enzyme	$K_m^{\text{carnitine}}$ (μM)	$K_m^{\text{palmitoyl-CoA}}$ (μM)	IC_{50} for malonyl-CoA (μM)
<i>Drosophila</i> CPT I	406 ± 30 (4)	104 ± 8 (4)	0.74 ± 0.10 (5)
Rat L-CPT I	120 ± 9.8 (3)	65 ± 4 (3)	54 ± 9.6 (3)
Rat M-CPT I	744 ± 81 (3)	34 ± 4 (3)	3.5 ± 0.4 (3)

could have generated malonyl-CoA-insensitive enzyme (see [16]). As can be seen from Table 1 and Figure 3(C), *Drosophila* CPT I is inhibited by malonyl-CoA. A greater proportion of CPT I activity is inhibitable by malonyl-CoA in crude extracts, suggesting that some proteolysis had occurred during the isolation of mitochondria (results not shown). Therefore whole extracts were used for a detailed analysis of malonyl-CoA effects.

The IC_{50} for malonyl-CoA was 0.74 ± 0.10 μM [in assays with 35 μM palmitoyl-CoA, 0.5 mM carnitine and 1 % (w/v) albumin at pH 7.0]. Therefore *Drosophila* CPT I is much more sensitive to inhibition by malonyl-CoA than either L-CPT I (IC_{50} 50-fold higher) or M-CPT I (IC_{50} almost 5-fold higher), when all the enzymes were expressed in *P. pastoris* (Figure 3C and Table 1). Further studies are under way to determine the residues of *Drosophila* CPT I that are important for its high sensitivity to inhibition by malonyl-CoA. The values determined here for L- and M-CPT I expressed in *P. pastoris* are an order of magnitude

Figure 3 Kinetic analysis of *Drosophila* CPT I expressed in *P. pastoris*

Representative data for the activity of *Drosophila* CPT I with respect to its substrates carnitine (A) and palmitoyl-CoA (B), and the inhibitor malonyl-CoA (C). (C) Inhibition of *Drosophila* CPT I activity by malonyl-CoA (●), with curves for rat M-CPT I (○) and rat L-CPT I (▼) (also expressed in *P. pastoris*) shown for comparison.

different from those observed for isolated rat mitochondria under identical assay conditions (V. N. Jackson, J. M. Cameron, V. A. Zammit and N. T. Price, unpublished work). This phenomenon is also seen for sheep L-CPT I expressed in *P. pastoris* compared with that in sheep liver mitochondria (V. N. Jackson, J. M. Cameron, V. A. Zammit and N. T. Price, unpublished work). This presumably reflects differences between the CPT I environments in *P. pastoris* and mammalian mitochondria. For example, yeast mitochondrial outer membranes

Figure 4 Probing of a Southern blot of *Drosophila* genomic DNA

D. melanogaster genomic DNA was digested with Afl II, Not I, Xba I and Xho I (lanes 1–4 respectively), subjected to agarose-gel electrophoresis and transferred to a nylon membrane as described in the text. The membrane was probed with ^{32}P -labelled probe derived from nt 1–852 of the *Drosophila* cDNA. After being washed, the hybridized probe was detected with a PhosphorImager. The positions of migration of DNA size markers are indicated (in kb).

have a lower sterol content [17], which would be expected to result in a more fluid lipid environment. Increased membrane fluidity is known to result in decreased malonyl-CoA sensitivity for rat L-CPT I (see [18]). Differences between our IC_{50} values for malonyl-CoA for CPT I expressed in *P. pastoris* and those reported elsewhere [4] are presumed to result from differences in assay conditions.

Sensitivity to inhibition by malonyl-CoA is a feature of mammalian CPT I that is central to its role in various cellular functions [19]. Previous studies have suggested that the extreme N-terminus of the molecule is essential for such sensitivity. Proteolytic removal of a small number of N-terminal residues [16] or expression of CPT I with the first 18 residues removed [20] results in the loss of the normal malonyl-CoA response. Mutation of His-5 (which is conserved in all mammalian CPT I forms) to alanine [21] also leads to a lowered malonyl-CoA sensitivity. Histidine residues 133 and 140 of rat L-CPT I are also conserved in all mammalian sequences. Mutation of His-133 to alanine had no effect but the same mutation of His-140 yielded CPT I with decreased sensitivity to malonyl-CoA [21].

The sequence alignment (Figure 2) shows that the extreme N-terminus of *Drosophila* CPT I is highly conserved: 11 of the first 15 residues are identical, including His-5. However, residues equivalent to His-133 and more importantly His-140 are not conserved in the *Drosophila* sequence (Lys-129 and Arg-136 respectively). The structures of mammalian and *Drosophila* CPT I clearly differ slightly in this region, because the sequence alignment (Figure 2) shows a relative insertion of three residues in the *Drosophila* sequence at position 137. If Arg-136 lies in an equivalent location within the structure of CPT I to His-140 in

tccaacgcgataagtcgtttaaaggaaggggatccgaagcatccccagtggaatttctcgtgtgcatctg
gcctaggatttgcgaattgtcatctacattccctctccctacaacgtttaaataaccacgctcacacattgc
ggtgcaattaaacacacccggaaacagtttacaccgaattgctttaattcgactagaagCACTTTCTGTACGT
GGGATATCATTAGCAACCTACAGGTATG

Figure 5 Sequence of the proposed alternative 5' untranslated region of *Drosophila* CPT I cDNA2

The three upstream in-frame stop codons are indicated in bold italics. Nucleotides 212–255, shown in capitals, are identical to nt 186–229 in cDNA1 (see Figure 1); nt 253–255 comprise the initiator methionine codon.

rat L-CPT I, it can be concluded that probably any basic residue can function in this position to confer sensitivity to inhibition by malonyl-CoA.

Evidence for the existence of only a single CPT I gene in *Drosophila*

As mammals have at least two closely related CPT I forms derived from different genes, we decided to examine the possibility that multiple isoforms are also present in *Drosophila*. A fragment of the cDNA corresponding to nucleotides encoding residues 1–209 of the protein sequence was used as a probe. The blot was washed at low stringency to detect any related sequences. Only a single fragment was detected with DNA cut with four different enzymes, indicating that it is highly unlikely that there is more than one gene for CPT I in *D. melanogaster* (Figure 4). Additionally, no evidence for another related gene product was found within the available *Drosophila* genomic or EST sequence data.

Because the nucleotide and protein sequences do not more closely resemble either of the two mammalian isoforms, it is not possible to infer any evolutionary relationship that might suggest that one mammalian isoform resembles an ancestral form from which the other diverged after gene duplication.

Evidence for *Drosophila* CPT I splice variants

After our initial database searches, 5' sequences for three further clones were identified. Their sequences matched one another and nt 186 onwards in the cDNA sequence shown in Figure 1 (cDNA 1) but they have a different 5' untranslated region. Figure 5 shows the 5' end of this variant cDNA (here termed cDNA 2). cDNA 1 and cDNA 2 would encode the same protein as the AUG codon and upstream in-frame stop codons are common to both (Figures 1 and 5).

As sequences for cDNA 2 are present in three different clones from two different cDNA libraries (GH, adult head; LP, various stages of larvae and early pupae) it is unlikely that cDNA 2 is an artifact. The point of divergence between the two cDNA species does not contain a consensus splice acceptor site, which would have been indicative of incompletely processed mRNA still containing an intron. The sequences of these clones have not been verified but there are no conflicts within the consensus sequence shown in Figure 5. The sequences of cDNA 1 and cDNA 2 are identical in the region of similarity (except for minor differences in sequence data beyond that described as high quality), suggesting that the two are derived from the same rather than from closely related genes, as supported by the detection of only a single gene by probing of a Southern blot (see above).

Further evidence is clearly required; however, the existence of

cDNA 1 and cDNA 2 is consistent with the presence of two different promoters, each giving rise to the alternative first exons. As a possibly less likely alternative, there might be a single CPT I promoter and the two cDNA species might represent different splice variants, with none of the cDNA sequences being long enough at the 5' end to reveal any common first exon sequence. The human M-CPT I gene also has two alternative promoters giving rise to two alternative 5' non-coding exons [22]. The rat L-CPT I gene has a single promoter [23].

The existence of a CPT I in *Drosophila* mitochondria suggests that the fruit fly uses a carnitine shuttle to transport long-chain fatty acids into mitochondria, as in mammals. Indeed sequences exist within the EST database that seem to correspond to *Drosophila* CPT II and the carnitine acylcarnitine carrier (results not shown). In addition, the very high sensitivity of the expressed protein to malonyl-CoA suggests that the inhibitor might also be an important regulator of CPT I activity in this organism, suggesting a widespread distribution of this mechanism of control of fatty acid oxidation and other cellular processes [19]. In view of the relatively low similarity in the sequence of *Drosophila* CPT I to those of mammalian CPT I, it is suggested that further study of the *Drosophila* protein will be a useful addition to approaches available for the elucidation of structure–function relationships in this important enzyme.

This work was supported by funding from the British Diabetic Association (V. N. J.) and the Scottish Office Agriculture and Fisheries Department.

REFERENCES

- 1 Yu, G. S., Lu, Y. C. and Gulick, T. (1998) *Biochem. J.* **334**, 225–231
- 2 Brown, N. F., Esser, V., Foster, D. W. and McGarry, J. D. (1994) *J. Biol. Chem.* **269**, 26438–26442
- 3 Prip-Buus, C., Cohen, I., Kohl, C., Esser, V., McGarry, J. D. and Girard, J. (1998) *FEBS Lett.* **429**, 173–178
- 4 de Vries, Y., Arvidson, D. N., Waterham, H. R., Cregg, J. M. and Woldegiorgis, G. (1997) *Biochemistry* **36**, 5285–5292
- 5 Zammit, V. A. (1999) *Progr. Lipid Res.*, in the press
- 6 Altschul, S. F., Madden, T. L., Schäffer, A. A., Zhang, J., Zhang, Z., Miller, W. and Lipman, D. J. (1997) *Nucleic Acids Res.* **25**, 3389–3402
- 7 Daum, G., Böhm, P. C. and Schatz, G. (1982) *J. Biol. Chem.* **257**, 13028–13033
- 8 Walker, J. M. (1994) *Methods Mol. Biol.* **32**, 5–8
- 9 Kolodziej, M. P. and Zammit, V. A. (1990) *Biochem. J.* **272**, 421–425
- 10 Kozak, M. (1986) *Cell* **44**, 283–292
- 11 Hofmann, K., Bucher, P., Falquet, L. and Bairoch, A. (1999) *Nucleic Acids Res.* **27**, 215–219
- 12 Brown, N. F., Anderson, R. C., Caplan, S. L., Foster, D. W. and McGarry, J. D. (1994) *J. Biol. Chem.* **269**, 19157–19162
- 13 Cronin, C. N. (1997) *Biochem. Biophys. Res. Commun.* **238**, 784–789
- 14 Cronin, C. N. (1997) *Eur. J. Biochem.* **247**, 1029–1037
- 15 Cronin, C. N. (1998) *J. Biol. Chem.* **273**, 24465–24469
- 16 Fraser, F., Corstorphine, C. G. and Zammit, V. A. (1997) *Biochem. J.* **323**, 711–718
- 17 Daum, G. (1985) *Biochim. Biophys. Acta* **822**, 1–42
- 18 Zammit, V. A., Corstorphine, C. G., Kolodziej, M. P. and Fraser, F. (1998) *Lipids* **33**, 371–376
- 19 McGarry, J. D. and Brown, N. F. (1997) *Eur. J. Biochem.* **244**, 1–14
- 20 Shi, J., Zhu, H., Arvidson, D. N., Cregg, J. M. and Woldegiorgis, G. (1998) *Biochemistry* **37**, 11033–11038
- 21 Swanson, S. T., Foster, D. W., McGarry, J. D. and Brown, N. F. (1998) *Biochem. J.* **335**, 513–519
- 22 Yu, G. S., Lu, Y. C. and Gulick, T. (1998) *J. Biol. Chem.* **273**, 32901–32909
- 23 Park, E. A., Steffen, M. L., Song, S., Park, V. M. and Cook, G. A. (1998) *Biochem. J.* **330**, 217–224
- 24 Thompson, J. D., Higgins, D. G. and Gibson, T. J. (1994) *Nucleic Acids Res.* **22**, 4673–4680