RESEARCH COMMUNICATION

Divergent sequences in the 5' region of cDNA suggest alternative splicing as a mechanism for the generation of carnitine acetyltransferases with different subcellular localizations

Olga CORTI, Stefano DiDONATO and Gaetano FINOCCHIARO*

Istituto Nazionale Neurologico 'C. Besta', Divisione di Biochimica e Genetica, 20133 Milano, Italia

We identified two partially overlapping cDNAs containing divergent 5' sequences of human carnitine acetyltransferase (CAT). cDNA λ SM-1400 extends the sequence of peroxisomal CAT, whereas cDNA λ SM-1200 creates a new open reading

frame encoding a putative mitochondrial leader peptide. An intron is located where sequences diverge, suggesting that mitochondrial, peroxisomal and possibly endoplasmic reticulum CAT mRNAs derive from alternative splicing of the CAT gene.

INTRODUCTION

Carnitine acetyltransferase (CAT; EC 2.3.1.7) is the enzyme that reversibly transfers short-chain acyl groups from acyl-coenzyme As (acyl-CoAs) to carnitine [1]. The conversion of acyl-CoAs into the corresponding acylcarnitine may protect cellular metabolism, maintaining coenzyme A availability and defending the cell from the accumulation of potentially toxic acyl-CoAs [2]; in particular, CAT-mediated synthesis of acetylcarnitine and propionylcarnitine can reverse respectively acetate- and propionate-induced inhibition of pyruvate oxidation [3,4].

CAT activity has been detected in different eukaryotes, from yeast to man, in different organs and in different subcellular compartments. In rat liver, in particular, approx. 50% of CAT activity is located in mitochondria, and the remaining 50% is present in similar amounts in peroxisomes and endoplasmic reticulum (ER) [5,6]. These ratios may vary after treatment with clofibrate [6,7] but also according to the tissue examined: in ram spermatozoa, for example, CAT is very abundant but is found exclusively in mitochondria [8]. However, the mechanisms that allow CAT to reach different subcellular targets are unknown.

In this paper we produce a first set of data that start to define these mechanisms at the molecular level and are based on the identification of two divergent sequences, overlapping with the 5' region of human CAT cDNA [9].

EXPERIMENTAL PROCEDURES

The screening of a λ GT10 cDNA library from human skeletal muscle (Clontech) was performed with a probe including the initial 0.3 kb of CAT cDNA, radiolabelled to a final specific activity of 3.5×10^8 c.p.m./ μ g of DNA by random priming (Random Primed DNA Labelling Kit, Boehringer). Nitrocellulose filters were prehybridized and hybridized at 65 °C in 5 $\times$ Denhardt's solution [0.1% (w/v) BSA, 0.1% (w/v) Ficoll, 0.1% (w/v) poly(vinylpyrrolidone)], 6 $\times$ SSC (1 $\times$ SSC is 0.15 M NaCl, 0.015 M sodium citrate), 0.1% (w/v) SDS and 100 μ g/ml of denatured salmon sperm DNA. The last wash was in 0.5 $\times$ SSC at room temperature.

Phage DNA was extracted by standard procedures [10] and

cDNA inserts were amplified by PCR with λ GT10 forward and reverse primers.

cDNA inserts were subcloned into plasmid pCRII (TA Cloning System, Invitrogen) according to the supplier's protocol or into vector pGEM7ZF(+) (Promega) after digestion with a suitable restriction enzyme. Plasmid DNA was purified by alkaline lysis [10] and sequences were determined using the T7 DNA polymerase Kit and the related protocol (Pharmacia). GC-rich regions were sequenced using 7-deaza-dGTP.

Constructs for transcription-translation experiments were obtained as follows. The SM-1200 cDNA insert was subcloned into vector pCRII and subsequently excised by *Eco*RI and *Dra*III digestion. The FB-1700 cDNA insert (corresponding to nt 817–2436 in CAT sequence [9]) was cut by *Cl*al and subcloned into the *Cl*al site of pGEM7ZF(+). Digestion of pFB1700 with *Dra*III (restriction site between nt 827 and 828 in the CAT sequence) and *Xho*I excised a cDNA fragment which was ligated to the pSM-1200 *Eco*RI/*Dra*III insert. The ligation product was amplified with primers SM-6.0 and FB-N2b (5' TGTCCGCGGACGGGACGCGAAGAT 3' and 5' GTGGCATTGGCAGGCCTGAGTCCT 3' respectively (Figure 1) and cloned into vector pCRII, creating plasmid pM-CAT. This plasmid was used as template to amplify with primers T7DSM (5' TAATACGACTCACTATAGTGAAGCCTCTGGGCTTCCTGAAG 3', Figure 1) and FB-N2b, a putative peroxisomal-specific cDNA (P-CAT) which lacks part of the region encoding the mitochondrial leader peptide.

pM-CAT was linearized by *Bam*HI and used together with P-CAT in a transcription experiment with the Riboprobe Gemini System from Promega, using the manufacturer's protocol. Capped mRNA derived from synthesized cDNA was translated *in vitro* by the rabbit reticulocyte lysate in the presence of [³⁵S]methionine, following the suggested protocol (Promega).

The full-length cDNA encoding putative mitochondrial CAT was also subsequently amplified from λ GT10 recombinant DNA extracted from the human muscle cDNA library. Primers SM-6.0 and opt-3' (5' AAGAGCTCTTGCCAGTCTCC 3', Figure 1) were used in 35 cycles of PCR, with annealing at 60 °C and extension time of 3 min. The amplification reactions contained, in a final volume of 100 μ l: 1 μ g of template DNA, 50 pmol of

Figure 1 Relationships between peroxisomal CAT cDNA and divergent sequences of clones λSM-1400 and λSM-1200

Schematic drawings of the cDNA sequence of peroxisomal CAT cDNA (a) and of the relevant part of clones λSM-1400 and λSM-1200 [left part of (b) and (c) respectively]. PCR primers are indicated by arrowheads: the three cDNAs are aligned with reference to primer FB-7c. Open reading frames (ORFs) are also indicated. Sequences and PCR primers corresponding to the drawings are also shown [right parts of (b) and (c)]. The TAG stop codon of clone λSM-1400 is in evidence. The open triangle indicates the location of the intron (see Figure 2). The EMBL accession numbers for sequence SM-1400 and SM-1200 are X79827 and X79825 respectively.

each primer, 200 μ M each dNTP, 50 mM KCl, 10 mM Tris/HCl, pH 8.3, 0.01% (w/v) gelatin, 1.5–2 mM $MgCl_2$ and 2.5 units (1 unit = 10 nmol dNTPs/30 min at 75°C) Taq DNA polymerase (Boehringer).

Genomic DNA (1 μ g) was amplified by primers SM-6.0b (5' ACAGCCATCGCCACTGCCACCTGG 3', Figure 1) and FB-7c (5'-CTTCAGGTAGTGGTCCAGGGA 3', Figure 1) using 35 cycles of denaturation (first cycle: 6 min at 95°C; subsequent cycles: 1 min at 95°C, 30 s at 62°C and 90 s at 72°C). The definition of the DNA sequence at 5' splicing junction of the intron was performed by PCR on genomic DNA using sense primer JUNCT (5' TTGGCTCTGTCTTCTCTT-TCCAGCGCCGG 3', Figures 1 and 2) and antisense primer VSJ5' (5' CCCTGTCTGTCTGCCTTGGCTGGA 3', Figure 2). The nucleotides underlined in primer JUNCT modify the CAT sequence by introducing a *NaeI* restriction site in the presence of a GC splice donor site in the intron. PCR products were subsequently precipitated, digested by *NaeI* and analysed by agarose gel electrophoresis.

RESULTS AND DISCUSSION

Identification of divergent sequences in the 5' region of CAT cDNA in clones λSM-1400 and λSM-1200

In previous studies we identified a 2436 bp cDNA encoding

human CAT and containing the peroxisomal targeting signal AKL [11] and the entire 3' untranslated region (EMBL no. X78706). This cDNA was not of full-length since CAT mRNA is approx. 2.9 kb in size [9].

In order to extend the 5' sequence we screened a cDNA library from human skeletal muscle and identified two clones, containing *EcoRI* inserts of approx. 1.4 (λSM-1400) and 1.2 kb (λSM-1200) respectively. The cDNA sequence of each matches the initial 1.1 kb of CAT cDNA [9]. Clone SM-1400, however, extends that sequence for 263 nt towards the 5' end, while clone SM-1200 shares the sequence of SM-1400 but diverges upstream of G261 (corresponding to G54 in the sequence of SM-1200, Figure 1).

Clone λSM-1400 extends the 5' end of peroxisomal CAT cDNA

The extension of the 5' end of CAT cDNA [9] due to clone λSM-1400 includes a stop codon at nt 147–149 and rules out the presence of other initiation codons (Figure 1). These data confirm that, in this reading frame, at least seven amino acids precede the N-terminus of CAT purified from human liver [9]. In agreement with this, a CAT form slightly larger than mature mitochondrial CAT, which is identical in size to purified CAT, was decorated by anti-CAT antibodies in peroxisome-enriched fractions from human liver [12].

It is likely that in rat liver peroxisomal CAT is also larger than

Figure 2 Identification of an intron in the CAT gene separating divergent and common cDNA sequences

(a) Lane 1: result of PCR amplification on genomic DNA with primers SM-6.0b and FB-7c (see also Figure 1). Lane 2: molecular-mass markers are type VI from Boehringer. (b) Left: direct sequence of the donor site. Right: reverse sequence of the acceptor site. Arrowheads indicate the exon-intron boundaries. (c, d) Detailed sequences of intron boundaries [EMBL no. X79826, (c); EMBL no. X79824, (d)]. Exon sequences are in capital letters, and primers used for the study of the non-conforming splice site are underlined.

its mitochondrial counterpart. It has been reported that the post-mitochondrial fraction obtained from liver of rats treated with the peroxisomal proliferator di(2-ethylhexyl) phthalate contains 69 and 67.5 kDa forms of CAT [13]. The larger form was considered the precursor of mitochondrial CAT but was only partially processed to its mature form [14], suggesting that part of the 69 kDa polypeptide is peroxisomal CAT.

Previous reports demonstrated that peroxisomal and microsomal CAT have "similar properties and could be the same protein" [15] and that the larger form of rat CAT "was found in the microsomal and soluble fractions, but not in the nuclear and mitochondrial fractions" [13]. One can therefore speculate that the 'peroxisomal' form of CAT is also targeted to ER, but the molecular basis for such localization is still unclear.

The retention of a number of resident proteins in the ER is dependent on the C-terminal signal KDEL [16]. This is absent in the CAT clones that we examined but a similar motif, KVEL, is present at position 471–474 in CAT [9] and in the yeast protein Sec-21, which is required for protein transport from ER to the Golgi apparatus [17].

Clone λ SM-1200 contains an alternative 5' end encoding the putative mitochondrial leader peptide of human CAT

The divergent sequence of clone λ SM-1200 contains an ATG codon (nt 27–29 in Figure 1) that defines a longer open reading frame containing an amino acid sequence with the characteristics of mitochondrial leader peptides. The sequence, in fact, contains a fairly regular distribution of positively charged residues and the isoelectric point, statistically determined, is 9.8.

It is likely that the cleavage point of this leader peptide is between F²⁸ and K²⁹. This is suggested by the sequence of peptide P3 obtained after trypsin digestion of purified CAT [9]. The P3

sequence is similar to the cDNA-derived amino acid sequence (AHQD) which overlaps the N-terminal sequence determined from purified CAT (XXQDAL...) [18]. The original sequence of the peptide was AAQD. The cDNA-encoded sequence AHQD, however, has been confirmed in different experiments and the difference could be due to either a polymorphism or an error in peptide sequencing. Peptide P3 is therefore at the extreme N-terminal part of CAT and K²⁹ could have been removed from it by trypsin digestion. If this is the case, the cleavage point that we propose (F²⁸–K²⁹) places an arginine residue at position –2, as found in 74% of matrix-processing protease cleavage sites [19], and releases a leader sequence of 28 amino acid residues, in agreement with the size defined by previous experiments on mitochondrial import of human CAT [12]. Furthermore, the molecular mass of the mitochondrial precursor would be of 70 kDa, again in agreement with the molecular mass of approx. 65 kDa deduced from immunoprecipitation studies [12].

The open reading frame defined by clone SM-1200 contains two potential translation start sites. The upstream ATG (nt 27–29 in Figure 1) is preceded by a purine at position –3, a highly conserved nucleotide in sequences surrounding strong initiator codons [20]. The sequence preceding the 'peroxisomal' ATG (nt 40–42 in Figure 1), on the contrary, only partially matches the consensus for translation initiation: C nucleotides occupy positions –5 and –4, but purine at position –3 and G at position +4 are both absent. Therefore the downstream ATG is a weak initiator codon [20].

These previsions have been confirmed by transcription-translation experiments. Plasmid pM-CAT, containing the complete coding sequence of mitochondrial CAT, and the amplification product P-CAT, lacking the initial 17 codons of mitochondrial ATG, were both transcribed using T7 RNA polymerase. Translation products are shown in Figure 3 and dem-

Figure 3 Transcription/translation experiments with 'mitochondrial' and 'peroxisomal' CAT cDNAs

Transcription/translation was performed starting from pM-CAT cDNA (mitochondrial) and P-CAT cDNA (peroxisomal) as explained in the Experimental procedures section. The relevant part of the corresponding fluorograms are shown: lane 1 pM-CAT, overnight exposure; lane 2 P-CAT, 2 day exposure. The uppermost band in lane 1 is similar in size to the 65–66 kDa precursor of mitochondrial CAT [12]. The weak uppermost band in lane 2 is, as expected, 1–2 kDa smaller than the putative mitochondrial precursor. The ~50 kDa band present in both lanes might represent a spurious translation start codon: ATG at nt 337–339 in the CAT sequence [9] could be a candidate for such a start codon.

onstrate that the downstream ATG is considerably less efficient than the upstream ATG as a translation start point.

It is worth of note that the 'mitochondrial' form of CAT contains both a mitochondrial and a peroxisomal targeting signal within the N- and C-terminal regions respectively. This has been confirmed by PCR experiments on cDNA from the skeletal muscle library with the 'mitochondrial'-specific sense primer SM-6.0 (nt 5–28 in Figure 1b) and antisense primer FB-opt3', situated in the 3' non-coding region of CAT cDNA (nt 2163–2186 of EMBL no. X78706; see also Figure 1). The expected amplification product of 2.2 kb (not shown) was further amplified with sense primer SM-com1 (nt 1036–1059 of EMBL no. X78706) and FB-opt3'. The sequence of this second amplification product demonstrated that the last three codons encode the AKL peroxisomal signal.

Therefore, in mitochondrial CAT the peroxisomal signal is present but the mitochondrial leader is prevailing. This is also the case for a mutant of human peroxisomal alanine:glyoxylate aminotransferase 1 (AGT1), a protein that shows 78% sequence identity with rat mitochondrial AGT1 but lacks the putative mitochondrial targeting signal [21]. Deficiency of human peroxisomal AGT1 results in primary hyperoxaluria type 1 which, in one third of the patients, is caused by AGT1 diversion from peroxisomes to mitochondria because of a proline-to-leucine substitution which creates a mitochondrial targeting signal in the N-terminal domain [22]. Therefore, in the case of AGT1 also, when mitochondrial and peroxisomal targeting signals co-exist it is the former that prevails.

An intron is located immediately downstream of the 'peroxisomal' sequence in the CAT gene

In order to investigate the region of the CAT gene where clones λSM-1200 and λSM-1400 diverge, we performed a PCR on genomic DNA with the 'mitochondrial' or the 'peroxisomal' forward primer SM-6.0 or SM-6.0b respectively and with the reverse primer FB-7c (Figure 1). A genomic fragment of approx. 1030 bp, 0.75 kb larger than the corresponding cDNA, was amplified only in the presence of the 'peroxisomal' forward primer (Figure 3).

The sequence of this fragment demonstrated that an intron is located in correspondence of the divergence between cDNAs (Figure 3). In the 5' splicing site of the intron a non-conforming GC replaces the highly conserved dimer GT. Non-conforming splice sites have been identified in a subset of exon-intron boundaries, and GC is the most common variant

[23]. However, in order to exclude PCR artifacts, we examined the 5' splice donor site in the genome of 14 controls. This was done by PCR with forward primer JUNCT, introducing the restriction site for *NaeI* in the presence of the non-conforming C, and reverse primer VSJ5'. The digestion of the 108 bp amplification product indicated that a *NaeI* site was created in all control DNAs (results not shown), consistent with the presence of the non-conforming GC at the 5' intron border.

Conclusion

The genomic structure of the 5' region of the CAT gene is therefore characterized by the presence of one intron which separates a putative 'peroxisomal' exon from a downstream sequence, shared by the 'mitochondrial' and the 'peroxisomal' form of CAT. Neither Southern analysis of genomic DNA (M. Rocchi, personal communication) nor *in situ* hybridization experiments [9] indicated the existence of different CAT genes. Rather, our data suggest the possibility that 'mitochondrial' and 'peroxisomal' cDNAs originate from alternative splicing of one single CAT gene, in agreement with the observation that non-conforming splice sites, such as the one we have described, may be involved in crucial gene regulation events [23]. In this context, it is likely that the 'mitochondrial' exon is located upstream of the 'peroxisomal' exon, separated by one (or more than one) intron; such an intron, however, could not be amplified by PCR, possibly because of its large size (O. Corti, unpublished work).

Our findings, together with previous reports by Markwell et al. [15] and Miyazawa et al. [13], raise the possibility that 'peroxisomal' CAT may also reach the ER. In this system, variations in the amount of these two latter forms would be regulated in parallel, which is exactly what happens in rat liver after administration of clofibrate [6,7].

CAT is therefore a promising model for the understanding of molecular mechanisms underlying the different subcellular localization of one same protein: the data we have presented provide the initial clues for such understanding, but further information is required before the picture can be completed. The significance of the results is also increased by the relevant role of CAT in cellular metabolism, so that regulation of its subcellular localization at the transcriptional level might be of relevance for orienting different metabolic pathways in the cell in response to external stimuli.

This work has been supported by grant no. 315 from TeleThon to G.F.

REFERENCES

- Colucci, W. J. and Gandour, R. D. (1988) *Bioorgan. Chem.* **16**, 307–334
- Chalmers, R. A., Roe, C. R., Tracey, B. M., Stacey, T. E., Hoppel, C. L. and Millington, D. S. (1983) *Biochem. Soc. Trans.* **11**, 724–725
- Uziel, G., Garavaglia, B. and DiDonato, S. (1988) *Muscle & Nerve* **11**, 720–724
- Brass, E. P., Gandour, R. D. and Griffith, O. W. (1991) *Biochim. Biophys. Acta* **1095**, 17–22
- Markwell, M. A. K., McGroarty, E. J., Bieber, L. L. and Tolbert, N. E. (1973) *J. Biol. Chem.* **248**, 3426–3432
- Kahonen, M. T. (1976) *Biochim. Biophys. Acta* **428**, 690–701
- Markwell, M. A. K., Bieber, L. L. and Tolbert, N. E. (1977) *Biochem. Pharmacol.* **26**, 1697–1702
- Day-Francesconi, M. and Casillas, E. R. (1982) *Arch. Biochem. Biophys.* **215**, 206–214
- Corti, O., Finocchiaro, G., Rossi, E., Zuffardi, O. and DiDonato, S. (1994) *Genomics*, in the press
- Sambrook J., Fritsch, E. F. and Maniatis, T. (1989) *Molecular Cloning: a Laboratory Manual*, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY
- Gould, S. J., Keller, G. A., Hosken, N., Wilkinson, J. and Subramani, S. (1989) *J. Cell Biol.* **108**, 1657–1664
- Bloisi, W., Colombo, I., Garavaglia, G., Giardini, R., Finocchiaro, G. and DiDonato, S. (1990) *Eur. J. Biochem.* **189**, 539–546

-
- 13 Miyazawa, S., Ozasa, H., Furuta, S., Osumi, T. and Hashimoto, T. (1983) *J. Biochem. (Tokyo)* **93**, 439–451
 - 14 Miyazawa, S., Ozasa, H., Furuta, S., Osumi, T., Hashimoto, T., Miura, S., Mori, M. and Tatibana, M. (1983) *J. Biochem. (Tokyo)* **93**, 453–459
 - 15 Markwell, M. A. K., Tolbert, N. E. and Bieber, L. L. (1976) *Arch. Biochem. Biophys.* **176**, 479–488
 - 16 Pelham, H. R. (1990) *Trends Biochem. Sci.* **15**, 483–486
 - 17 Hosobuchi, M. M., Kreis, T., Schekman, R. (1992) *Nature (London)* **360**, 603–605
 - 18 Finocchiaro, G., Taroni, F., Rocchi, M., Liras-Martin, A., Colombo, I., Torri Tarelli, G. and DiDonato, S. (1991) *Proc. Natl. Acad. Sci. U.S.A.* **88**, 661–665
 - 19 Hendrick, J., Hodges, P. and Rosenberg, L. (1989) *Proc. Natl. Acad. Sci. U.S.A.* **86**, 4056–4060
 - 20 Kozak, M. (1989) *J. Cell Biol.* **108**, 229–241
 - 21 Takada, Y., Kaneko, N., Esumi, H., Purdue, P. E. and Danpure, C. J. (1990) *Biochem. J.* **268**, 517–520
 - 22 Purdue, P. E., Allsop, J., Isaya, G., Rosenberg, L. E. and Danpure, C. J. (1991) *Proc. Natl. Acad. Sci. U.S.A.* **88**, 10900–10904
 - 23 Senapathy, P., Shapiro, M. and Harris, N. L. (1990) *Methods Enzymol.* **183**, 252–278

Received 20 May 1994/11 July 1994; accepted 14 July 1994