Nucleotide sequence of cDNA coding for rat liver pI 6.1 esterase (ES-10), a carboxylesterase located in the lumen of the endoplasmic reticulum

Mariette ROBBI,* Henri BEAUFAY and Jean-Noël OCTAVE†

Laboratoire de Chimie Physiologique, Université Catholique de Louvain and International Institute of Cellular and Molecular Pathology, Avenue Hippocrate 75, B-1200 Brussels, Belgium

A commercial rat liver cDNA library in λ gt11 was screened with a rabbit antiserum to native pI 6.1 esterase (ES-10). The inserts of the immunoreactive clones were short (0.9–1.1 kbp). One of these was used as a probe to rescreen the library, yielding 30 clones, two of which contained relatively long (approx. 1.9 kbp) and widely overlapping cDNA inserts. They did not contain the first two nucleotide residues of the initiator codon, nor the 5'-end untranslated portion of the mRNA. These were derived from a home-made rat liver cDNA library in λ gt11, screened with an oligonucleotide corresponding to the 5'-end of the already known cDNA sequence. The nucleotide sequence consists of 48 bp of 5'-end non-coding region, 1695 bp of coding region and 212 bp of 3'-end non-coding region including a 20 bp poly(A) tail. The signal peptide and the mature protein subunit are 18 and 547 residues long respectively. Tyr is confirmed as *N*-terminal residue. The predicted amino acid sequence is highly similar to those of rabbit liver esterase forms 1 (77% identity) and 2 (56% identity), determined by protein sequencing [Korza & Ozols (1988) *J. Biol. Chem.* **263**, 3486–3495; Ozols (1989) *J. Biol. Chem.* **264**, 12533–12545]. The three enzymes share the Ser and His residues presumed to be part of the active site, four Cys residues and a high proportion of charged side chains at their C-terminus. The C-terminal tetrapeptides of the three esterases (-HVEL, -HIEL and -HTEL for pI 6.1 and forms 1 and 2 esterases respectively) are reminiscent of, but not identical with, the localization signal identified in other proteins of the endoplasmic-reticulum lumen {-KDEL in animal cells [Munro & Pelham (1987) *Cell* **48**, 899–907]; -HDEL in yeast [Pelham, Hardwick & Lewis (1988) *EMBO J.* **7**, 1757–1762]}. We still lack direct evidence to decide whether or not these C-terminal tetrapeptides commit esterases to reside in the endoplasmic reticulum. In that case the antepenultimate residue (D, V, I or T) would be only weakly stringent, and some sequences primed by H instead of K would be recognized in animal as well as in yeast cells.

INTRODUCTION

Esterase activity is located in the endoplasmic reticulum (ER) of rat liver cells (Amar-Costesec *et al.*, 1974; Beaufay *et al.*, 1974). It is shared by several isoenzymes, which are either free in the lumen or loosely bound to the inner side of the membrane (Mentlein *et al.*, 1980; Robbi & Beaufay, 1983). The early steps in the biosynthesis of several microsomal esterases cannot be distinguished from those of many secretory proteins. The polypeptide chain is synthesized by bound polyribosomes, co-translationally discharged into the ER lumen and processed by proteolytic cleavage and core glycosylation (Robbi & Beaufay, 1986, 1987, 1988). However, esterases are sorted out of the traffic of secretory proteins. How they acquire their stable localization in the ER is still poorly understood. It has been shown (Munro & Pelham, 1987) that several other proteins of the ER lumen share the C-terminal sequence -KDEL, which is essential for their retention in the cell. The C-terminal sequences -EVX (X = S, N or Q) and -AVL reported by Mentlein *et al.* (1984b) for pI 6.4 and pI 6.1 esterases respectively suggest that this retention mechanism might not apply to all the proteins of the ER lumen, in particular to these enzymes.

Microsomal esterases are most probably involved in the metabolism of xenobiotics and of natural substrates (Mentlein *et al.*, 1984a; Mentlein & Heymann, 1984). They are encoded by different genes (Simon *et al.*, 1985). The mature molecules are glycoproteins of the high-mannose type and migrate in

SDS/PAGE as polypeptides of approx. 60 kDa. They are most easily resolved on the basis of their electric charges, and therefore are referred to by the pI value (Mentlein *et al.*, 1980; Robbi & Beaufay, 1983). The pI 6.1 enzyme (genetic nomenclature ES-10; Mentlein *et al.*, 1987) is a homotrimer; the other three enzymes are monomers.

We have undertaken the structural characterization of these enzymes to learn more about the molecular basis responsible for their stable association with the ER and their function. The work was initiated on the pI 6.1 enzyme, the prominent esterase in rat liver. A cDNA clone, encompassing the whole coding sequence of the precursor protein, was identified in a rat liver cDNA library constructed in λ gt11. We present the sequence of this clone and the predicted amino acid sequence of the encoded protein. Similarity to other esterases and implications on the biosynthesis and intracellular targeting are also discussed. Part of these results have been presented in abstract form (Robbi & Beaufay, 1989).

MATERIALS AND METHODS

Antibody screening of cDNA library

A cDNA library in the λ gt11 expression vector, generated from rat liver polyadenylated RNA (Sprague–Dawley strain) was purchased from Clontech Laboratories (Palo Alto, CA, U.S.A.). It was screened, according to the procedure of Young &

Abbreviation used: ER, endoplasmic reticulum.

* To whom correspondence should be addressed.

† Present address: Laboratoire de Neurochimie, Université Catholique de Louvain, UCL 1352, Avenue Hippocrate 10, B-1200 Brussels, Belgium. The nucleotide sequence data reported will appear in the EMBL, GenBank and DDBJ Nucleotide Sequence Databases under the accession number X51974.

Davis (1980), with *Escherichia coli* Y1090 as the host bacterium. The filters were probed overnight with a rabbit antiserum specifically directed against the native pI 6.1 esterase (Robbi & Beaufay, 1987), at a dilution sufficient to detect 1 ng of antigen. The bound antibodies were treated with pig anti-(rabbit IgG) antibody followed by peroxidase–rabbit anti-peroxidase complex (Dakopatts, Copenhagen, Denmark). The immunoreactive plaques were located with 4-chloro-1-naphthol and H_2O_2 . The positive clones were picked and further purified through two or three additional rounds of screening until all plaques gave a positive signal, and amplified as plate stocks.

Nucleotide screening of cDNA libraries

The commercial library was rescreened with *E. coli* Y1088 as host bacterium and a 1.1 kbp immunologically detected cDNA insert as probe. Bacteriophage plaques were lifted to nylon sheets (Hybond N; Amersham International, Amersham, Bucks., U.K.). Fixation of DNA to the membrane, prehybridization and hybridization were performed essentially as recommended by the manufacturer. The probe was labelled with [α - ^{32}P]dCTP (3000 Ci/mmol) and the multiprime labelling kit of Amersham International.

To search further for a full-length cDNA, another rat liver cDNA library in λ gt11 (Wistar rats) was constructed and screened with an oligonucleotide (24-mer) corresponding to the 5'-end of the already known structure. This probe was radiolabelled with [γ - ^{32}P]ATP (> 1000 Ci/mmol) and T4 polynucleotide kinase.

Preparation of antibodies specific for pI 6.1 esterase fusion proteins

Nitrocellulose filters derived from confluent plates of each of the four immunoreactive recombinants and λ gt11 were first blocked with 3% (w/v) gelatin in Tris-buffered saline (150 mM-NaCl/50 mM-Tris/HCl buffer, pH 8.0) and then incubated for 6 h with anti-(pI 6.1 esterase) serum (1:100) in 1% gelatin in Tris-buffered saline. After extensive washing of the filters with Tris-buffered saline, the bound antibodies were eluted by four 1 min rinses in 0.15 M-NaCl/0.5% Triton X-100/1% gelatin/5 mM-glycine buffer, pH 2.3 (Weinberger *et al.*, 1985). The eluents were promptly neutralized by mixing with 40 ml 1% gelatin in Tris-buffered saline, pH 8.3, and used without further dilution for immuno-blotting.

PAGE and immuno-blotting

Proteins were resolved on a discontinuous 10% polyacrylamide gel as described by Laemmli (1970). Esterases were detected by histochemical staining with α -naphthyl acetate as substrate (Ljungquist & Augustinsson, 1971). For immuno-blotting, proteins were transferred electrophoretically to nitrocellulose as described by Towbin *et al.* (1979). The filters were blocked with 3% gelatin in Tris-buffered saline and incubated with the affinity-selected antibodies. The bound antibodies were located by enzymic staining with the reagents supplied by Dakopatts as described above.

Isolation of RNA

Total rat liver RNA (Wistar rats) was prepared as described by Chirgwin *et al.* (1979), by extraction with guanidinium thiocyanate and centrifugation through 5.7 M-CsCl. mRNA was isolated according to the procedure of Aviv & Leder (1972) by two cycles of binding to oligo(dT)–cellulose type 2 (Collaborative Research, Lexington, MA, U.S.A.). The cDNA library was made as described by Octave *et al.* (1987).

Northern and Southern blotting

The polyadenylated RNA was heated for 5 min at 65 °C

in 2.2 M-formaldehyde/50% (v/v) formamide/1 mM-EDTA/25 mM-Mops/NaOH buffer, pH 7.0, and size-fractionated by electrophoresis through a denaturing 1.5% agarose gel containing 25 mM-Mops, 0.66 M-formaldehyde and 0.5 μ g of ethidium bromide/ml.

The restricted recombinants were electrophoresed, together with DNA size markers, through a non-denaturing 1.5% agarose gel containing ethidium bromide. After the run, the gel was denatured and neutralized as described by Amersham International.

The resolved components were transferred to nylon membranes by passive diffusion. The blots were probed as described above, with the [α - ^{32}P]dCTP-labelled cDNA insert IA_1 .

Subcloning of λ gt11 cDNA inserts and DNA sequencing

Bacteriophage recombinant DNA was purified from each of the positive clones by the plate lysis procedure (Maniatis *et al.*, 1982) and digested with *Eco*RI. The most appropriate inserts were subcloned in the plasmid vectors pTZ18R and 19R. Transfections were performed by the Hanahan (1985) procedure with *E. coli* JM 109 as host bacterium. Vectors and recombinant DNAs were purified by CsCl centrifugation after alkaline lysis of the cells (Maniatis *et al.*, 1982). Recombinant DNAs were characterized by restriction mapping. Fragments of suitable size for sequencing were further subcloned into pTZ18R and 19R. Single-stranded DNAs were purified and sequenced by the dideoxy chain-termination method of Sanger *et al.* (1977) with T7 DNA polymerase (Sequenase; U.S. Biochemical Corp., Cleveland, OH, U.S.A., and Pharmacia, Gent, Belgium) and [α - ^{35}S]thio[dATP]. DNA sequence data were analysed by using the DNA Strider 1.0 program (Marck, 1988).

Restriction enzymes and T4 ligase were purchased from Bethesda Research Laboratories (Gent, Belgium) or Boehringer (Mannheim, Germany) and used in accordance with the manufacturers' instructions.

RESULTS

Characterization of immunologically detected cDNA clones for pI 6.1 esterase

Screening of the commercial rat liver cDNA library (Sprague–Dawley) in λ gt11 with a rabbit antiserum to pI 6.1 esterase yielded four positive clones out of 2.2×10^6 . These immunoreactive clones were designated IA_1 to IA_4 . They contained inserts ranging in size from 0.9 to 1.1 kbp and overlapped extensively (result not shown).

That these clones indeed harbour cDNA inserts encoding fragments of the pI 6.1 esterase is demonstrated in Fig. 1. The antibodies affinity-purified by adsorption on nitrocellulose filters carrying the fusion proteins derived from the immunoreactive clones were identified on immunoblots. They reacted clearly with pI 6.1 esterase (lanes 2–5). The filter derived from a plate with the parent vector λ gt11 carrying no insert (lane 6) did not release such antibodies.

Northern-blot analysis

The size of rat liver pI 6.1 esterase mRNA was determined by polyadenylated RNA blot analysis. Radiolabelled IA_1 hybridized to a single mRNA of approx. 2200 nucleotide residues (Fig. 2).

Search for full-length pI 6.1 esterase cDNA and sequencing

The library was rescreened with radiolabelled IA_1 as a probe. Thirty strongly positive clones were obtained; they were designated HA_1 to HA_{30} . Analysis on agarose gel of *Eco*RI

Fig. 1. Identification of the antibodies reacting with the immunoreactive recombinant clones

Polyclonal anti-(pI 6.1 esterase) serum was incubated with nitrocellulose filters derived from confluent plates of each of the four immunologically detected cDNA clones and the parent vector λ gt11. The epitope-selected antibodies were eluted as described in the Materials and methods section. Six identical samples of crude 0.2% Triton X-100 extracts from rat liver microsomal fractions (corresponding to 100 μ g of microsomal protein each) were resolved by non-denaturing PAGE: lane 1, gel stained for esterase activity with α -naphthyl acetate as substrate; lanes 2–6, samples transferred to nitrocellulose and treated separately with the antibodies retained by four pI 6.1-esterase-positive recombinant clones (lanes 2–5) or the parent vector λ gt11 (lane 6). The bound antibodies were treated with pig anti-(rabbit IgG) antibody followed by peroxidase–rabbit anti-peroxidase complex. Immunoreactive proteins were located with 4-chloro-1-naphthol and H_2O_2 .

Fig. 2. Northern-blot analysis of pI 6.1 esterase mRNA in rat liver

Rat liver polyadenylated RNA (6 μ g) was size-fractionated on a 1.5% agarose gel containing formaldehyde and transferred to a nylon membrane. pI 6.1 esterase mRNA was detected by hybridization with 32 P-labelled cDNA insert 6.1-IA₁ (referred to below as IA₁) as described in the text. Positions of ethidium bromide-stained 28S (5.5 kb) and 18S (2.1 kb) rat liver rRNAs are indicated by arrows. The arrowhead points to pI 6.1 esterase mRNA, the size of which was estimated to be 2.2 kb.

digests of these clones revealed that they contain inserts ranging in size from 0.6 to 1.9 kbp (results not shown). The two longest, HA₂₆ and HA₂₉, which by restriction mapping were found to be very similar, and the immunologically detected clone IA₁ were

Fig. 3. Restriction map of overlapping pI 6.1 esterase clones and sequencing strategy

The restriction sites used in subcloning in the plasmic vectors pTZ 18R and 19R of clones IA₁, HA₂₆, HA₂₉ and HB₁ are shown (key: P, *Pst*I; B, *Bam*HI; A, *Ava*II; S, *Sau*3AI; Hd, *Hind*III; K, *Kpn*I; N, *Nsp*7524; Hf, *Hinf*I). The DNA sequence was determined by the dideoxy chain-termination method (Sanger *et al.*, 1977) with single-stranded DNA templates. Stippled and open bars indicate coding regions for signal peptide and mature pI 6.1 esterase subunit respectively; single lines indicate 5'-end and 3'-end non-coding regions. Arrows indicate direction and extent of DNA sequencing of restriction fragments; arrow starting with open circle indicates sequence primed with 24-mer used to screen the second library.

retained for further characterization and sequencing (Fig. 3). These clones still did not encompass the complete coding region of the precursor of pI 6.1 esterase subunit.

A cDNA clone containing the full-length coding region was obtained from another rat liver cDNA library (Wistar rats) in λ gt11, by screening with a 24-mer oligonucleotide corresponding to the 5'-end of HA₂₆ and HA₂₉ (Fig. 3). This yielded a single positive clone, HB₁. Sequencing of this clone confirmed the earlier data and contributed 2 bp of the initiator codon and 48 bp of the 5'-end untranslated region not present in the clones isolated from the commercial library (Fig. 3).

The cDNA sequence of clone HB₁ (see Fig. 3) and the predicted amino acid sequence for pI 6.1 esterase subunits are shown in Fig. 4. The first ATG codon in this sequence is probably the start site for translation, being located in the consensus sequence for eukaryotic initiation codons (Kozak, 1986) and followed by an open reading frame of 1695 bp (i.e. 18 amino acid residues for signal peptide and 547 amino acid residues for mature subunit). Thus the sequence encompasses part of the 5'-end non-coding region (48 bp) and the 3'-end non-coding region (212 bp), which includes two polyadenylation signals and ends with a poly(A) tail. That the cleavage site for signal peptidase is after glycine at position -1 is barely questionable: (1) the small uncharged residues at positions -1 and -3 are separated by a bulky residue, which is consistent with the

		CAAAGGAACAAATAGCCACAGATCCGAGAACCCTTGTCCT																				42
43	TCCACA	ATG	CGC	CTC	TAC	CCT	CTG	GTC	TGG	CTT	TTT	CTT	GCT	CGC	TGC	ACA	GCT	TGG	GGG	↓	102	
-18		Met	Arg	Leu	Tyr	Pro	Leu	Val	Trp	Leu	Phe	Leu	Ala	Ala	Cys	Thr	Ala	Trp	Gly		-1	
103	TAC	CCA	TCC	TCA	CCA	CCT	GTG	GTG	AAC	ACT	GTT	AAA	GGC	AAA	GTC	CTG	GGG	AAG	TAT	GTC	162	
1	Tyr	Pro	Ser	Ser	Pro	Pro	Val	Val	Asn	Thr	Val	Lys	Gly	Lys	Val	Leu	Gly	Lys	Tyr	Val	20	
163	AAT	TTG	GAA	GGA	TTT	GCA	CAG	CCT	GTG	GCT	GTT	TTC	CTG	GGA	ATC	CCC	TTC	GCC	AAG	CCC	222	
21	Asn	Leu	Glu	Gly	Phe	Ala	Gln	Pro	Val	Ala	Val	Phe	Leu	Gly	Ile	Pro	Phe	Ala	Lys	Pro	40	
223	CCT	CTT	GGC	TCC	TTG	AGG	TTT	GCT	CCA	CCA	CAG	CCT	GCA	GAG	CCT	TGG	AAC	TTT	GTG	AAG	282	
41	Pro	Leu	Gly	Ser	Leu	Arg	Phe	Ala	Pro	Pro	Gln	Pro	Ala	Glu	Pro	Trp	Asn	Phe	Val	Lys	60	
283	AAT	ACT	ACC	TCC	TAC	CCA	CCT	ATG	TGC	TCT	CAA	GAT	GCT	GTT	GGA	GGG	CAG	GTT	CTC	TCA	342	
61	Asn	Thr	Thr	Ser	Tyr	Pro	Pro	Met	Cys	Ser	Gln	Asp	Ala	Val	Gly	Gly	Gln	Val	Leu	Ser	80	
343	GAG	CTT	TTC	ACC	AAC	AGG	AAG	GAA	AAC	ATT	CCT	TTA	CAG	TTT	TCT	GAA	GAC	TGC	CTC	TAC	402	
81	Glu	Leu	Phe	Thr	Asn	Arg	Lys	Glu	Asn	Ile	Pro	Leu	Gln	Phe	Ser	Glu	Asp	Cys	Leu	Tyr	100	
403	CTG	AAC	GTT	TAT	ACT	CCG	GCT	GAC	TTG	ACA	AAG	AAC	AGC	CGG	CTA	CCA	GTG	ATG	GTG	TGG	462	
101	Leu	Asn	Val	Tyr	Thr	Pro	Ala	Asp	Leu	Thr	Lys	Asn	Ser	Arg	Leu	Pro	Val	Met	Val	Trp	120	
463	ATC	CAT	GGA	GGT	GGA	CTG	GTA	GTG	GGT	GGA	GCA	TCC	ACC	TAT	GAT	GGA	CAG	GTC	CTC	TCT	522	
121	Ile	His	Gly	Gly	Gly	Leu	Val	Val	Gly	Gly	Ala	Ser	Thr	Tyr	Asp	Gly	Gln	Val	Leu	Ser	140	
523	GCC	CAT	GAA	AAT	GTG	GTG	GTG	GTG	ACC	ATT	CAG	TAT	CGC	CTT	GGC	ATC	TGG	GGA	TTC	TTC	582	
141	Ala	His	Glu	Asn	Val	Val	Val	Val	Thr	Ile	Gln	Tyr	Arg	Leu	Gly	Ile	Trp	Gly	Phe	Phe	160	
583	AGC	ACA	GGG	GAT	GAA	CAC	AGC	CGG	GGC	AAC	TGG	GGT	CAC	TTG	GAC	CAG	GTG	GCT	GCA	CTA	642	
161	Ser	Thr	Gly	Asp	Glu	His	Ser	Arg	Gly	Asn	Trp	Gly	His	Leu	Asp	Gln	Val	Ala	Ala	Leu	180	
643	CAC	TGG	GTC	CAG	GAC	AAC	ATT	GCC	AAC	TTT	GGG	GGT	AAC	CCA	GGC	TCT	GTG	ACC	ATC	TTT	702	
181	His	Trp	Val	Gln	Asp	Asn	Ile	Ala	Asn	Phe	Gly	Gly	Asn	Pro	Gly	Ser	Val	Thr	Ile	Phe	200	
703	GGA	GAA	TCT	GCA	GGA	GGT	TTC	AGT	GTC	TCT	GCT	CTT	GTG	TTA	TCT	CCT	CTG	GCC	AAG	AAC	762	
201	Gly	Glu	Ser	Ala	Gly	Gly	Phe	Ser	Val	Ser	Ala	Leu	Val	Leu	Ser	Pro	Leu	Ala	Lys	Asn	220	
763	CTC	TTC	CAC	AGG	GCC	ATT	TCT	GAG	AGT	GGT	GTG	GTC	CTC	ACT	TCT	GCT	CTG	ATT	ACA	ACA	822	
221	Leu	Phe	His	Arg	Ala	Ile	Ser	Glu	Ser	Gly	Val	Val	Leu	Thr	Ser	Ala	Leu	Ile	Thr	Thr	240	
823	GAT	AGC	AAG	CCC	ATT	GCT	AAT	CTG	ATT	GCT	ACA	CTT	TCT	GGG	TGT	AAA	ACC	ACC	ACA	TCA	882	
241	Asp	Ser	Lys	Pro	Ile	Ala	Asn	Leu	Ile	Ala	Thr	Leu	Ser	Gly	Cys	Lys	Thr	Thr	Thr	Ser	260	
883	GCT	GTT	ATG	GTT	CAT	TGC	CTG	CGC	CAG	AAG	ACA	GAG	GAT	GAA	CTC	CTG	GAG	ACT	TCA	TTA	942	
261	Ala	Val	Met	Val	His	Cys	Leu	Arg	Gln	Lys	Thr	Glu	Asp	Glu	Leu	Leu	Glu	Thr	Ser	Leu	280	
943	AAA	TTG	AAT	CTT	TTC	AAA	CTG	GAC	TTA	CTT	GGA	AAC	CCA	AAA	GAG	AGC	TAT	CCC	TTC	CTA	1002	
281	Lys	Leu	Asn	Leu	Phe	Lys	Leu	Asp	Leu	Leu	Gly	Asn	Pro	Lys	Glu	Ser	Tyr	Pro	Phe	Leu	300	
1003	CCT	ACT	GTG	ATT	GAC	GGA	GTG	GTG	CTG	CCA	AAG	ACA	CCA	GAA	GAG	ATC	CTG	GCT	GAG	AAG	1062	
301	Pro	Thr	Val	Ile	Asp	Gly	Val	Val	Leu	Pro	Lys	Thr	Pro	Glu	Glu	Ile	Leu	Ala	Glu	Lys	320	
1063	AGT	TTC	AAC	ACA	GTC	CCC	TAC	ATA	GTG	GGC	ATC	AAC	AAG	CAA	GAG	TTT	GGC	TGG	ATC	ATT	1122	
321	Ser	Phe	Asn	Thr	Val	Pro	Tyr	Ile	Val	Gly	Ile	Asn	Lys	Gln	Glu	Phe	Gly	Trp	Ile	Ile	340	
1123	CCA	ACG	CTT	ATG	GGC	TAT	CCA	CTC	TCC	GAA	GGC	AAA	CTG	GAC	CAG	AAA	ACA	GCC	AAA	TCC	1182	
341	Pro	Thr	Leu	Met	Gly	Tyr	Pro	Leu	Ser	Glu	Gly	Lys	Leu	Asp	Gln	Lys	Thr	Ala	Lys	Ser	360	
1183	CTC	TTG	TGG	AAG	TCC	TAC	CCA	ACA	CTG	AAA	ATC	TCT	GAG	AAA	ATG	ATT	CCA	GTG	GTT	GCT	1242	
361	Leu	Leu	Trp	Lys	Ser	Tyr	Pro	Thr	Leu	Lys	Ile	Ser	Glu	Lys	Met	Ile	Pro	Val	Val	Ala	380	
1243	GAG	AAG	TAC	TTC	GGA	GGG	ACA	GAT	GAC	CCT	GCC	CAA	AGG	AAA	GAC	CTG	TTC	CAG	GAC	TTG	1302	
381	Glu	Lys	Tyr	Phe	Gly	Gly	Thr	Asp	Asp	Pro	Ala	Lys	Arg	Lys	Asp	Leu	Phe	Gln	Asp	Leu	400	
1303	GTT	GCA	GAT	GTG	ATA	TTT	GGT	GTC	CCA	TCA	GTA	ATG	GTG	TCT	CGA	AGT	CAC	AGA	GAT	GCT	1362	
401	Val	Ala	Asp	Val	Ile	Phe	Gly	Val	Pro	Ser	Val	Met	Val	Ser	Arg	Ser	His	Arg	Asp	Ala	420	
1363	GGA	GCC	CCC	ACC	TTC	ATG	TAT	GAA	TTT	GAG	TAT	CGC	CCA	AGC	TTT	GTA	TCA	GCC	ATG	AGG	1422	
421	Gly	Ala	Pro	Thr	Phe	Met	Tyr	Glu	Phe	Glu	Tyr	Arg	Pro	Ser	Phe	Val	Ser	Ala	Met	Arg	440	
1423	CCC	AAG	ACA	GTG	ATC	GGA	GAC	CAT	GGT	GAT	GAA	CTC	TTC	TCA	GTA	TTT	GGA	TCT	CCA	TTT	1482	
441	Pro	Lys	Thr	Val	Ile	Gly	Asp	His	Gly	Asp	Glu	Leu	Phe	Ser	Val	Phe	Gly	Ser	Pro	Phe	460	
1483	TTA	AAA	GAT	GGT	GCC	TCA	GAA	GAG	GAG	ACC	AAT	CTC	AGC	AAA	ATG	GTG	ATG	AAA	TAC	TGG	1542	
461	Leu	Lys	Asp	Gly	Ala	Ser	Glu	Glu	Glu	Thr	Asn	Leu	Ser	Lys	Met	Val	Met	Lys	Tyr	Trp	480	

1543	GCC AAC TTT GCT CGG AAT GGG AGC CCT AAT GGG GGA GGG CTG CCC CAT TGG CCA GAA TAT	1602
481	Ala Asn Phe Ala Arg Asn Gly Ser Pro Asn Gly Gly Gly Leu Pro His Trp Pro Glu Tyr	500
1603	GAC CAG AAG GAA GGG TAC CTG AAG ATT GGT GCC TCA ACT CAG GCA GCC CAG AGG CTG AAG	1662
501	Asp Gln Lys Glu Gly Tyr Leu Lys Ile Gly Ala Ser Thr Gln Ala Ala Gln Arg Leu Lys	520
1663	GAC AAA GAA GTG GCT TTT TGG TCT GAG CTC AGG GCC AAG GAG GCA GCA GAG GAA CCA TCC	1722
521	Asp Lys Glu Val Ala Phe Trp Ser Glu Leu Arg Ala Lys Glu Ala Ala Glu Glu Pro Ser	540
1723	CAC TGG AAA CAT GTT GAG CTC TGA TCAGGAGGGTCAGCCATGTTTGAGAACCTGAAGCTAAAGGGGAATTA	1793
541	His Trp Lys <u>His Val Glu Leu</u> ***	547
1791	TTCCACAGAAGATTTTGTAAGACATAACACTTCTTGTCTTTGAGACTATAACATCAGGTATTTTGTACAAATGCA	1872
1873	<u>TTAAAGGGAAATACTTAACCTTATTGCTTCAACTGTAAATAAAACATTGTTTAGAAATTCAAAAAAAAAAAAA</u>	1951
1952	AAAA	1955

Fig. 4. Amino acid sequence of precursor of rat liver pI 6.1 esterase subunit inferred from the nucleotide sequence of clone HB₁

The signal-sequence cleavage site is indicated by an arrow and potential *N*-glycosylation sites (Asn-Xaa-Ser/Thr) are underlined. Ser and His residues presumed to be part of the active site are labelled *; Cys residues are labelled .. The C-terminal tetrapeptide is doubly underlined. The in-frame stop codon (TGA) is labelled *** and the polyadenylation signals (ATTAAG and AATAAG) in the 3'-end untranslated region are underlined (waved).

1	YPSSPPVNTVKGVLGKYNLEGFAPVAVFLGIPFAKPPLGSLRFAPPQPAEPWNFVK	
	H*****D*T*****S*****T*****V*****E*****S***	
	Y++++++N+V+++++D++N+++A	
	Y.....N.V.....N....	
61	NTTSYPPMCSQDAVGGQVLSELFTRNKENIPLQFSEDCLYNVYTPADLTKNRLPVMVW	
	T**G*V*KL*ADMLSTG*S**E*****I*S*****S*****	
121	IHGGGLVVGGASTYDQVLSAHENVVVTIQYRLGIWGFSTGDEHSRGNWGHLDQVAAL	
	*****II***P*S*LA*****F*G*L*****A*****L***	
181	HWQDNIANFGGNPGSVTIFGESAGGFSVSALVLSPLAKNLFHRAISESGVVLTSALITT	
	R*****D*****V*****TN*DKK	
	...F...L....	
241	DSKPIAKNIATLSGCKTTTSAMVMHCLRQKTEDELLETSLKLNLFKLDLLGNPKESYPFL	
	NTQAV*QM*****NN*S**A**Q*****A***LTV*-----NTSM	
301	PTVIDGVVLPKTPPEILAEKSFNTVPYIVGINKQEFGWIIPTLMGYPLSEGKLDQKTAKS	
	S*****T*****F*****M**NL**RMNE*M*S*	
	+++++++	
	A.....I..	
360	LLWKSYPYTLKISEKMIPVVAEKYFGGTDDPAKRKDLFQDLVADVIFGVPSVMVSRSHRDA	
	F*KRFS*N* <u>N</u> *SV**AII**LR*****K*E*LL*MFS**F*I*A*LM**L***	
		
420	GAPTFMYEFYRPSFVSAMRFKTVIGDHGDELFSVFGSPFLKDGASEEETNL SKMVMKYW	
	YQ*****DQ**Q**Q*****I*****T***E*****L***F*	
	++I+++++L+	
	...F.....G... ..I...EDL....	
480	ANFARNGSPNGGGLPHWPEYDQKEGYLKIGASTQAAQRLKDKEVAFWSELRAKEAAEAPS	
	*****N**E*****K*****Q**T**Q**K**GE*****T**L**NPPQTEH	
	..DK.....S..R	
540	HWKHVEL	
	TEHT	

Fig. 5. Comparison between the amino acid sequences predicted for the mature rat liver pI 6.1 esterase subunit and for other rat liver esterases (Long *et al.*, 1988; Takagi *et al.*, 1988)

The sequences are shown in single-letter code. Amino acid residues corresponding to pI 6.1 esterase subunit (first lines) are numbered in the left-hand margin. The second lines compare that sequence with that deduced from cDNA cloning by Takagi *et al.* (1988), which is identical with that reported simultaneously by Long *et al.* (1988) except for residues 156–157, which were Phe-Gly instead of Ile-Trp. Hyphens indicate gaps introduced for maximal alignment. Amino acids common with pI 6.1 esterase are labelled *; His and Ser residues of the active site are labelled *; Cys residues are labelled ..; potential *N*-glycosylation sites are underlined and C-terminal tetrapeptides are in bold characters. The structure of purified peptide fragments that were sequenced by Takagi *et al.* (1988) and by Long *et al.* (1988) are shown by (+) and (·) respectively, interspaced by residues that were inconsistent with their cDNA sequence.

```

1  YPSSPPVNTVKGKVLGKYVNLGFAQPVAVFLGIPFAKPLGLSLRFAPPQPAEPWNFVK
H**A****D*****F*S*****V*****SH**
QD*AS*IR*HT*Q*R*SL*HV**TDAG*HT*****P*****E**A*A*G*R

61  NNTSYPPMCSQD-AVGGO-VLSELFNTNRKENIPLQFSEDCLYLNVTYTPADLTKNRSLPVM
*****S---S*H-M*****K*****I*****RG****
DG*L*A*LGNL*IMDQDV*LLH*PP--S*M---*****S*HAREGSD****

119 VVIHGGGLVVGASTYDQVLSAHENVVVTIQYRLGIWGFSTGDEHSRGNWGHLDQVA
*****M*****LA*****G**GFNI*-----*FL**
*****TM*M*M*SA*A*F*D*****VL*****Q*AT*H*Y*****

179 ALHWVQDNIANFGGNPGSVTIFGESAGGFSVSALVLSPLAKNLFHRAISESGVLTSAI
VNR*****D*****Q**I*L****T*****A*L*S*F
**R***K***H*****R*****T***SH***MSQG***G**M*L*A*LPG**

239 TTDKPIAKNIATLSGCKTTTSAMVHCLRQKTEDELLETSLKLNLFKLDLLGNPKESYP
RKNT*SL*EK**IEA*****E**M*VT**MKFMA**V*D***NTA
*SS*EVVSTVV*N**R*QQVD*ETL*R**A*S*E*MLAITQV---*M---*IPG-----

299 FLPTVIDGVVLPKTPPEEILAEKSFNTVPYIVGINKQEFGWIIPTLM-GYPLSEGKLDQKT
**T*****L**A*A**YE**KYNML**M***Q*****MQ*L*****
---*V**F**RH**L**LAD*QP**S*I**ND*Y*****K*LLAIDPQ*ER-*RQA

358 AKSLWLKSYPTLKIASEKMPVVAEKYFGGTDDPAKRKDLFQDLVADVIFGVPSVMVSRSH
*TE*****IVNV*KELT*AT**L*****V*K***L*ML**LL*****N*A*H*
MREIMHQATKQ*MLPPA*GDLLMDE*M*SNE**KHLMAQ*QEMM*A*VM*ALR*AHLQ

418 RDAGAPTFMYEFYRPSFVSAMRFKTVIGDHGDELFSVFGSPFL--KDGASEEETNLSKM
*****Y***YR*****S*D*****I**L*A***--*E**T***IK****
*SH-***YF***QH***TKDL*PH*RA*****VVF**R*HLFGS*VPLT***EL**RR

476 VMKYWANFARNGSPNGGLPHWPEYDQKEGYLKIGASTQAAQRKDKQEVAFWSELRAKEA
*****N***E**Q**A*Y*****Q**T***K*****T*W****
*****RN***E**A**LF*LDQR**QLNMQPAVG*A**ARRLQ**THTLPQRV

536 AEEP SHWK--HVEL
R-RETE--I**
Q*-LRGTEQK*T**

```

Fig. 6. Comparison of the predicted amino acid sequence of the mature pI 6.1 esterase subunit with the protein sequences of purified rabbit liver esterase forms 1 and 2 (Korza & Ozols, 1988; Ozols, 1989)

The sequences are shown in single-letter code. Amino acid residues corresponding to pI 6.1 esterase subunit (first lines) are numbered in the left-hand margin. The sequences of rabbit esterase forms 1 and 2 are shown on the second and third lines respectively. Gaps, homologies, putative residues of the active site, Cys residues, potential *N*-glycosylation sites and the terminal tetrapeptide are indicated as in Fig. 5.

rules for signal-peptide cleavage sites (von Heijne, 1983); (2) the size of the leader peptide is compatible with the difference in molecular mass between the precursor and the proteolytically processed polypeptide chains (Robbi & Beaufay, 1987); (3) the *N*-terminus of the mature protein is Tyr, as shown by Edman degradation of the purified protein (Mentlein *et al.*, 1980). The polypeptide chain contains two potential *N*-glycosylation sites, one of which is used and bears a high-mannose-type oligosaccharide (Robbi & Beaufay, 1987; Natsuka *et al.*, 1988). As discussed below, the C-terminal tetrapeptide is not identical to but is related to the localization signal for the ER lumen in animal cells (Munro & Pelham, 1987) and yeast (Pelham *et al.*, 1988).

Only four differences between the clones derived from the commercial library (Sprague-Dawley strain) and HB₁ were noted in the coding region: Gln-168 (CAG) instead of Arg-168 (CGG), Lys-247 (AAG) instead of Asn-247 (AAT), Met-405 (ATG) instead of Ile-405 (ATA) and Asn-488 (AAC) instead of Ser-488 (AGC). A single-base substitution was also detected in the 3' non-coding sequence: G-1777 instead of A-1777. Whether these differences are due to strain or allelic variations is unknown.

Takagi *et al.* (1988) and Long *et al.* (1988) have reported cDNA sequences coding for rat liver carboxylesterases. The nucleotide and amino acid sequences reported by these two groups are nearly identical, but they differ noticeably from the sequence predicted in our work (Fig. 5). The amino acid sequences predicted by Long *et al.* (1988) and Takagi *et al.* (1988), however, do not satisfactorily match the amino acid sequences they

determined experimentally from peptide fragments of the purified protein. These peptide fragments match much better with the sequence deduced from our cDNA clones. Presumably, we all attempted to sequence the cDNA coding for the same protein, but Long *et al.* (1988) and Takagi *et al.* (1988) extracted from their libraries cDNA clones coding for another still-to-be-identified isoenzyme.

DISCUSSION

Here we report the first cDNA sequence coding for the full length of the precursor polypeptide of a well-characterized microsomal carboxylesterase. The encoded protein is the pI 6.1 enzyme [identical with esterase E1 or ES-10 (Mentlein *et al.*, 1987)]. Indeed, the fusion proteins produced by the recombinant bacteriophages bind rabbit antibodies that, after release from the antigen-antibody complex, specifically react with the authentic protein isolated by non-denaturing PAGE. The amino acid sequence predicted starting from the first ATG codon is fully compatible with a preprotein containing a removable signal peptide (first 18 residues) (von Heijne, 1983), and with earlier data about the structure and properties of the mature protein (next 547 residues). The pI value calculated from the amino acid composition (6.08) is identical with that determined by isoelectric focusing (Robbi & Beaufay, 1983). However, the structure of the C-terminus is different from that reported by Mentlein *et al.* (1984b). As there is no predicted AVL sequence near the C-terminal end, the discrepancy cannot be accounted for by partial

proteolytic degradation of the protein. Probably the method used earlier to sequence the C-terminus of the purified protein has failed.

Korza & Ozols (1988) and Ozols (1989) have established the primary structure of two microsomal esterases purified from rabbit liver and designated 60 kDa esterase forms 1 and 2. Comparison of these protein sequences with the predicted amino acid sequence for pI 6.1 esterase from rat liver reveals 77% identity with form 1 and 56% identity with form 2 (Fig. 6). The Ser and His residues identified as part of the active site by diisopropyl phosphate binding (Korza & Ozols, 1988) and four Cys residues are present in the three proteins. The pI 6.1 and form 1 enzymes share one of their two potential N-glycosylation sites (Asn-61). Presumably this site bears the single high-mannose-type oligosaccharide chain present in each pI 6.1 esterase subunit. Esterase form 2 has a single glycosylatable Asn residue located elsewhere in the molecule. The pI 6.1 and form 1 esterases show a remarkable similarity, spread over their whole length, except for a short segment (residues 161–182) where the structure of pI 6.1 esterase is closer to that of esterase form 2. The C-terminal ends of the three enzymes are highly charged. The C-terminal tetrapeptides are identical, except for the antepenultimate residue. Hydropathy plots and α -helix and β -sheet amphiplicity distributions (results not shown) suggest that pI 6.1 and form 1 esterases are closely related enzymes in the two animal species.

Rabbit 60 kDa esterase form 1 behaves as a monomeric transmembrane protein with its C-terminal end accessible to trypsin on the cytoplasmic side (Korza & Ozols, 1988; Ozols, 1989), whereas rat pI 6.1 esterase is a trimer and behaves as a lumenally exposed peripheral protein. It requires a detergent for extraction from microsomal membranes but not for the subsequent purification steps (Mentlein *et al.*, 1980; Robbi & Beaufay, 1983). In this earlier work we have also shown that the activity of rat liver esterase pI 6.1 is highly sensitive to trypsin and chymotrypsin when the protein is free, but remarkably resistant in sealed microsomal vesicles. This implies that any membrane-spanning segment connecting the catalytic domain to a putative cytoplasmically exposed C-terminus should be located after residue His-448. In that part of the sequence the maximum hydrophobicity value on a segment of 21 residues is below 0.42 in the rabbit and rat proteins, making the existence of a membrane-spanning helix highly unlikely (Eisenberg *et al.*, 1984). Thus the structural analysis strengthens our earlier conclusion that pI 6.1 esterase is not inserted in the phospholipid bilayer, but resides in the ER lumen.

The various esterases sequenced so far each harbour a different C-terminal tetrapeptide, none of which is identical with the -KDEL signal that retains several intraluminal proteins in animal cells (Munro & Pelham, 1987). The known C-terminal tetrapeptides of rat and rabbit esterases have the structure HXEL, in which X is V, I or T. In fact, they are closer to the sequence -HDEL, identified as an ER-targeting signal in yeast cells (Pelham *et al.*, 1988). The meaning of this similarity is puzzling, because those authors have shown that the sequence HDEL does not lead to the retrieval of manipulated proteins from the secretory traffic in animal cells. At present we still lack direct evidence to decide whether or not the C-terminal ends of esterases function as an ER retention signal. The C-terminus (-TEHT) of the protein cloned by Long *et al.* (1988) and by Takagi *et al.* (1988) is unrelated to -KDEL or -HXEL, but the subcellular topology of that protein is unknown. Therefore a striking similarity remains between the C-termini of three well-characterized esterases and sequences identified as topogenic signals for other ER-associated proteins. If the terminal tetrapeptide sequence of esterases is indeed responsible for their

retention in the ER, some variation is tolerated for the antepenultimate residue, and the species-specificity of H- and K-primed sequences is less sharp than originally thought, or depends on other structural features of the protein.

Note added in proof (received 21 May 1990)

Our interpretation that the C-terminal dipeptide -EL is the most stringent requirement for retention of certain proteins in the ER lumen is based on the structure of several natural esterases. It is now strengthened by the results reported by Andres *et al.* (1990), who show that variants of proneuropeptide Y are retained in cells of the AtT-20 line when one of the tetrapeptides KDEL, DKEL, RDEL or KNEL is added to the C-terminus.

We are grateful to Dr. P. Michels and Dr. M. Darville (International Institute of Cellular and Molecular Pathology, Brussels) for helpful advice and valuable discussions. This research was supported by the Belgian Fonds de la Recherche Fondamentale Collective (no. 2.4549.88), by the Belgian State-Prime Minister's Office-Science Policy Programming (Poles d'Attraction Interuniversitaires Grant no. 7, and Actions Concertées Grant no. 83/93-122) and by the Belgian Fonds National de la Recherche Scientifique.

REFERENCES

- Amar-Costesec, A., Wibo, M., Thinès-Sempoux, D., Beaufay, H. & Berthet, J. (1974) *J. Cell Biol.* **62**, 717–745
- Andres, D. A., Dickerson, J. M. & Dixon, J. E. (1990) *J. Biol. Chem.* **265**, 5952–5955
- Aviv, H. & Leder, P. (1972) *Proc. Natl. Acad. Sci. U.S.A.* **69**, 1408–1412
- Beaufay, H., Amar-Costesec, A., Feytmans, E., Thinès-Sempoux, D., Wibo, M., Robbi, M. & Berthet, J. (1974) *J. Cell Biol.* **61**, 188–200
- Chirgwin, J. M., Przybyla, A. E., MacDonald, R. J. & Rutter, W. J. (1979) *Biochemistry* **18**, 5294–5298
- Eisenberg, D., Schwarz, E., Komaromy, M. & Wall, R. (1984) *J. Mol. Biol.* **179**, 125–142
- Hanahan, D. (1985) *DNA Cloning: A Practical Approach*, vol. 1, pp. 109–135, IRL Press, Oxford and Washington
- Korza, G. & Ozols, J. (1988) *J. Biol. Chem.* **263**, 3486–3495
- Kozak, M. (1986) *Cell* **44**, 283–292
- Laemmli, U. K. (1970) *Nature (London)* **227**, 680–685
- Ljungquist, A. & Augustinsson, K. B. (1971) *Eur. J. Biochem.* **23**, 303–313
- Long, R. M., Satoh, H., Martin, B. M., Kimura, S., Gonzalez, F. J. & Pohl, L. R. (1988) *Biochem. Biophys. Res. Commun.* **156**, 866–873
- Maniatis, T., Fritsch, E. F. & Sambrook, J. (1982) *Molecular Cloning: A Laboratory Manual*, pp. 368–369, Cold Spring Harbor Laboratory Press, Cold Spring Harbor
- Marck, C. (1988) *Nucleic Acids Res.* **16**, 1829–1836
- Mentlein, R. & Heymann, E. (1984) *Biochem. Pharmacol.* **33**, 1243–1248
- Mentlein, R., Heiland, S. & Heymann, E. (1980) *Arch. Biochem. Biophys.* **200**, 547–559
- Mentlein, R., Suttorp, M. & Heymann, E. (1984a) *Arch. Biochem. Biophys.* **228**, 230–246
- Mentlein, R., Schumann, M. & Heymann, E. (1984b) *Arch. Biochem. Biophys.* **234**, 612–621
- Mentlein, R., Ronai, A., Robbi, M., Heymann, E. & von Deimling, O. (1987) *Biochim. Biophys. Acta* **913**, 27–38
- Munro, S. & Pelham, H. R. B. (1987) *Cell* **48**, 899–907
- Natsuka, S., Himeno, M., Hase, S., Ito, H., Ueda, T., Kato, K. & Ikenaka, T. (1988) *J. Biochem. (Tokyo)* **103**, 986–991
- Octave, J.-N., Macq, A.-F., de Sauvage, F., Maloteaux, J.-M., Laterre, E. C. & Trouet, A. (1987) *Protides Biol. Fluids* **35**, 67–70
- Ozols, J. (1989) *J. Biol. Chem.* **264**, 12533–12545
- Pelham, H. R. B., Hardwick, K. G. & Lewis, M. J. (1988) *EMBO J.* **7**, 1757–1762
- Robbi, M. & Beaufay, H. (1983) *Eur. J. Biochem.* **137**, 293–301
- Robbi, M. & Beaufay, H. (1986) *Eur. J. Biochem.* **158**, 187–194
- Robbi, M. & Beaufay, H. (1987) *Biochem. J.* **248**, 545–550
- Robbi, M. & Beaufay, H. (1988) *Biochem. J.* **254**, 51–57
- Robbi, M. & Beaufay, H. (1989) *Biol. Chem. Hoppe-Seyler* **370**, 988
- Sanger, F., Nicklen, S. & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. U.S.A.* **74**, 5463–5467

- Simon, B., de Looze, S., Ronai, A. & von Deimling, O. (1985) *Electrophoresis* **6**, 575–582
- Takagi, Y., Morohashi, K., Kawabata, S., Go, M. & Omura, T. (1988) *J. Biochem. (Tokyo)* **104**, 801–806
- Towbin, H., Staehelin, T. & Gordon, J. (1979) *Proc. Natl. Acad. Sci. U.S.A.* **76**, 4350–4354
- von Heijne, G. (1983) *Eur. J. Biochem.* **133**, 17–21
- Weinberger, C., Hollenberg, S. M., Ong, E. S., Haruon, J. M., Brower, S. T., Cidlowski, J., Thompson, E. B., Rosenfeld, M. G. & Evans, R. M. (1985) *Science* **228**, 740–742
- Young, R. A. & Davis, R. W. (1980) *Science* **222**, 778–782

Received 25 September 1989/4 January 1990; accepted 18 January 1990