

# cDNA cloning of carboxyl ester lipase from human pancreas reveals a unique proline-rich repeat unit

Karen Reue,<sup>1,\*</sup> Johanne Zambaux,<sup>\*</sup> Howard Wong,<sup>\*</sup> Gregory Lee,<sup>\*</sup> Thomas H. Leete,<sup>\*</sup> Michael Ronk,<sup>†</sup> John E. Shively,<sup>†</sup> Berit Sternby,<sup>\*\*</sup> Bengt Borgström,<sup>\*\*</sup> Detlev Ameis,<sup>2,\*</sup> and Michael C. Schotz<sup>\*</sup>

Lipid Research, VA Wadsworth Medical Center,<sup>\*</sup> Los Angeles, CA 90073 and UCLA Department of Medicine,<sup>\*</sup> Los Angeles, CA 90024; Division of Immunology,<sup>†</sup> Beckman Research Institute of City of Hope, Duarte, CA 91010; and Department of Physiological Chemistry,<sup>\*\*</sup> University of Lund, Lund, Sweden

**Abstract** We report the isolation and nucleotide sequence of the cDNA for carboxyl ester lipase (CEL) from human pancreas. CEL was purified from human pancreas and microsequence analysis was performed on the amino-terminal and internal peptides. Peptide sequence was used to design oligonucleotide probes for screening a human pancreas cDNA library. Partial length cDNAs for CEL were isolated from the library, and the 5' portion of the cDNA was obtained using the anchored polymerase chain reaction. The deduced amino acid sequence indicates that mature CEL contains 722 amino acids and is synthesized with a 20 amino acid leader peptide. The amino acid sequence is rich in proline (12.2%), with 68% of the proline residues occurring within the final 25% of protein length. This is due to the occurrence of a series of proline-rich tandem repeat units near the carboxyl terminus, and accounts for the previously observed species variation in CEL size and amino acid composition. The primary sequence of CEL shows strong similarity to members of the serine esterase family, including the identical G-E-S-A-G motif at the putative active site. A striking homology also occurs between CEL and acetylcholinesterase and cholinesterase, essential enzymes of the nervous system. Proteins with cholesteryl esterase activity have been detected in extra-pancreatic tissues including liver, intestine, kidney, aorta, macrophage, and in the milk of some species (human, gorilla, cat, dog), but not others (rat, cow). To clarify the structural relationships between these various esterases and CEL, we used the CEL cDNA to study expression in pancreas and liver. CEL mRNA was abundant in pancreas of human and rat, with the human CEL mRNA approximately 300 nucleotides larger than that from rat. CEL mRNA was not detected in human adult or fetal liver, nor in rat liver. ■ These results indicate that CEL is not synthesized in significant amounts in liver, and suggest that the cholesterol esterase activity that has been described in liver may be due to a distinct enzyme, or may be derived from pancreas, as has been proposed for the cholesterol esterase activity in intestine.—Reue, K., J. Zambaux, H. Wong, G. Lee, T. H. Leete, M. Ronk, J. E. Shively, B. Sternby, B. Borgström, D. Ameis, and M. C. Schotz. cDNA cloning of carboxyl ester lipase from human pancreas reveals a unique proline-rich repeat unit. *J. Lipid Res.* 1991. **32**: 267–276.

**Supplementary key words** lipases • cholesterol esterases • bile salt-stimulated lipase

The absorption and digestion of dietary lipids is dependent on pancreatic synthesis and secretion of lipolytic enzymes. The mammalian pancreas produces two major lipolytic enzymes, pancreatic lipase and carboxyl ester lipase (CEL, EC 3.1.1.3). Both enzymes hydrolyze dietary lipids, but differ in several properties. CEL has no co-factor requirement, although its catalytic activity is markedly enhanced by low concentrations of bile salts, whereas pancreatic lipase is inactivated by bile salts unless a co-factor protein, colipase, is present. CEL is synthesized in the acinar cells of the pancreas, stored in zymogen granules, and secreted into the intestinal lumen where it is activated by bile salts. In rat neonates, CEL comprises approximately 90% of total lipase activity present in the pancreatic juice (1); in adult humans, CEL constitutes about 4% of pancreatic juice proteins (2). CEL has a broad substrate range and is capable of hydrolyzing phospholipids, vitamin esters, cholesteryl esters, and triacylglycerols (3). As a consequence, it has been studied under several names including carboxyl ester lipase, carboxyl ester hydrolase, cholesterol esterase, non-specific lipase, and lysophospholipase. The broad substrate reactivity and abundance of this enzyme suggests that CEL plays a major role in the assimilation of dietary lipids.

Enzymes that resemble CEL in their activation by bile salts, immunological reactivity, and amino acid composition have been detected in rat liver (4, 5) and in human

Abbreviations: CEL, carboxyl ester lipase; BSSL, bile salt-stimulated lipase; kDa, kiloDalton; cDNA, complementary DNA; SDS, sodium dodecyl sulfate; PVDF, polyvinylidenedifluoride; PTH, phenylthiohydantoin; HPLC, high pressure liquid chromatography; nt, nucleotides; kb, kilobases; PCR, polymerase chain reaction.

<sup>1</sup>To whom correspondence should be addressed at: VA Wadsworth Medical Center, Building 113, Room 312, Los Angeles, CA 90073.

<sup>2</sup>Present address: Department of Medicine I, University of Hamburg, Germany.

milk (6, 7). The bile salt-stimulated lipase (BSSL) present in milk of higher primates, cat, and dog is thought to play a role in fat hydrolysis and absorption in the intestine of newborns (8, 9). BSSL from milk is indistinguishable from pancreatic CEL immunologically and has the identical sequence of 30 N-terminal amino acids (7, 10). This has led to the proposal that these enzymes are identical and suggests that CEL may also function in utilization of milk fats and maintaining lipid homeostasis in the newborn.

Pancreatic CEL from the rat, pig, cow, and human have been purified to homogeneity. The molecular mass differs significantly among species: human CEL is 100 kDa (2); pig CEL occurs both as a 74 kDa monomer and as a dimer (11); bovine CEL occurs predominantly as a 72 kDa protein (12); and rat CEL is 67 kDa (5). The molecular weight differences may be due to carbohydrate content; however, the amino acid compositions of these enzymes vary considerably (7). The human enzyme contains a high proportion of proline (13%) not found in other species. These data suggest that there is substantial species variation in the structure of CEL and that the human enzyme may contain functional or structural domains not found in other species.

To facilitate studies of CEL structure and function, we have isolated and characterized the cDNA for human CEL. Protein and DNA sequence data indicate that human CEL contains 722 amino acids and is synthesized with a 20 amino acid leader peptide. Comparison of the cDNA-encoded human CEL sequence with that of rat (13, 14) and cow (12) revealed that the human protein is 20% larger due to extensive repetition of a proline-rich sequence at the carboxyl end of human CEL. Thus, a difference in the primary protein sequence accounts for interspecies variation in CEL amino acid composition and contributes to the difference in molecular weight. Human CEL displays no significant homology with other lipases, aside from the putative active site motif, G-X-S-X-G. CEL does, however, strongly resemble members of the serine esterase family, including acetylcholinesterase and cholinesterase, essential enzymes of the nervous system.

## METHODS

### Protein purification

Human CEL was purified from human pancreatic juice (2) and determined to be homogeneous by electrophoresis on SDS 7–14% polyacrylamide gels developed with silver nitrate.

### N-terminal sequence analysis

Amino terminal sequence of CEL was determined as described (15). Briefly, CEL was electrophoresed on SDS 7–14% polyacrylamide gels, and electrotransferred to Im-

mobulon PVDF sheets (polyvinylidenedifluoride; Millipore). The sheets were stained with 1.25% Coomassie blue R-250 in water-methanol-acetic acid 50:40:10 (v/v/v) for 5 min, then destained in solution without dye for 10 min. The PVDF region corresponding to the migration position of CEL (approximately 100 kDa) was excised and placed into the reaction vessel of a gas-phase microsequenator (Beckman Instruments). Protein sequencing was performed as previously described (16). Typically, 100 pmol of CEL loaded on polyacrylamide gels gave 20 pmol PTH-amino acid signals.

### Internal peptide sequence analyses

Purified CEL (500 pmol) was reduced and alkylated with dithiothreitol and iodoacetamide in the presence of 6 M guanidine HCl (16). The solution was exhaustively dialyzed against 50 mM ammonium bicarbonate buffer, pH 8.4, then trypsinized at a 23.5:1 substrate to enzyme ratio for 5 h at 22°C. Proteolysis was inhibited by reducing the pH to 2 with concentrated HCl. Peptide fragments were separated by HPLC (BrownLee) utilizing a 2.1 mm × 25 cm C<sub>4</sub> reverse-phase column (Vydac). Fractionated peptides were collected manually and aliquots were spotted on Immobulon prior to placement into the sequenator.

### Isolation of cDNA clones

An oligo(dT) primed human pancreas cDNA library constructed in a  $\lambda$ gt10 vector was prepared and provided by Dr. H. Okamoto (Sendai, Japan). This library was amplified in *E. coli* strain Y1088. Two synthetic oligonucleotides were designed based on sequence derived from the protein N-terminus and an internal tryptic peptide. The oligonucleotides were synthesized in accordance with published codon usage tables (17; Genetic Designs, The Woodlands, Texas), and had the following sequences: (N-terminus) 5'-CAGCTTCT/GTGTTCCTCCCTCCACAAAGCCG CCTCT/GGTGTAGACAGCTCCCAGCTTGGC-3', and (internal) 5'-GCCATCCATGTTGTTGGTGCCAGCAA TGTAGTCAATGTCAAAGGCATTGGC-3'. Approximately  $4 \times 10^5$  recombinants were screened on NYTRAN filters (Schleicher and Schuell) with <sup>32</sup>P-end-labeled oligonucleotide probes (18). Hybridization was at 37°C in  $5 \times$  SSC ( $1 \times = 0.15$  M NaCl/15 mM sodium citrate), 20% formamide,  $4 \times$  Denhardt's solution ( $1 \times = 0.02\%$  Ficoll/0.02% polyvinylpyrrolidone/0.02% bovine serum albumin), 0.1% SDS, 0.12 M sodium phosphate, pH 6.8, 2 mM EDTA, pH 7.2, 1% sarcosyl, and 50  $\mu$ g/ml salmon sperm DNA. Filters were washed twice in  $2 \times$  SSC/0.1% SDS for 15 min at room temperature, followed by a 60-min wash at 42°C. cDNA inserts from hybridizing clones were excised with EcoRI and subcloned into the plasmid pSPT19 (Boehringer Mannheim). Sequence determinations were by the dideoxy chain termination method (19).

## Generation of cDNA 5' end by polymerase chain reaction

cDNA was synthesized from RNA and amplified by the polymerase chain reaction (PCR) (20). Briefly, 10  $\mu$ g total human pancreas RNA (generously provided by Dr. W. Hunziker, Hoffman-LaRoche, Basel) was primed for cDNA synthesis with 100 pmol random hexamer oligonucleotides (Pharmacia) in a 50  $\mu$ l reaction containing 1 $\times$  reverse transcriptase buffer (50 mM Tris-HCl, pH 8.3, 75 mM KCl, 10 mM DTT, 3 mM MgCl<sub>2</sub>), 20 units RNasin (Promega), 2.5 mM each dATP, dCTP, dGTP, and dTTP, and 200 units M-MLV reverse transcriptase (BRL). The reaction was at 37°C for 60 min. The resulting cDNA was purified on a Centricon 30 column (Amicon Corp.) and tailed with dGTP in a 25  $\mu$ l reaction containing 0.2 M potassium cacodylate, pH 7.2, 4 mM MgCl<sub>2</sub>, 1 mM 2-mercaptoethanol, 0.5 mg/ml BSA, 0.25 mM dGTP, and 15 units terminal deoxynucleotidyl transferase (BRL). The G-tailed cDNA was then used as a template for PCR with a CEL-specific primer (5'-GATCGAATTCAAGCTTTTGAAGGTGACCACGATGACGTTT-3'; complementary to nucleotides 482 to 504 in Fig. 1) and an "anchor primer" containing a tract of C residues (5'-GATCGAATTCAAGCTTC<sub>14</sub>-3'). Both primers contain *EcoRI* and *HindIII* restriction enzyme recognition sites to facilitate cloning of PCR products. For PCR, samples were denatured at 94°C (1 min), annealed at 50°C (2 min), and extended at 72°C (40 min), followed by 40 cycles of 94°C (1 min), 55°C (2 min), and 72°C (2 min), with a final extension of 10 min. PCR products were extracted with phenol-chloroform precipitated with ethanol, restricted with *EcoRI* (which cuts at sites in both primers used for amplification), and isolated on a non-denaturing 5% polyacrylamide gel. DNA was excised and electroeluted from the gel, and subcloned into pGEM2 (Promega). Two independent clones were isolated and sequenced in both directions.

## RNA analysis

Rat pancreas RNA was prepared by guanidinium-isothiocyanate extraction (21). Human pancreas RNA was provided by Dr. W. Hunziker (Hoffman-LaRoche, Basel). Poly A<sup>+</sup> RNA from adult and fetal human liver and from rat liver was from Clontech (Palo Alto, CA). RNA samples were fractionated on formaldehyde gels and transferred to a nylon membrane (22). Blots were hybridized to the CEL 2 kb cDNA and to a human alpha-tubulin cDNA (23) labeled with <sup>32</sup>P (10<sup>8</sup> cpm/ $\mu$ g). After hybridization for 16 h at 55°C in 0.5 M sodium phosphate, pH 7.2, 7% SDS, and 1 mM EDTA, blots were washed at room temperature in 2 $\times$  SSC/0.1% SDS, and at 55°C in 1 $\times$  SSC/0.1% SDS and 0.5 $\times$  SSC/0.1% SDS. RNA size standards ranging from 0.24 to 9.5 kb (BRL) were used to estimate mRNA size.

## RESULTS

### CEL purification and peptide sequence analysis

CEL was isolated from human pancreatic juice as described (2). Purified CEL demonstrated a single band with an apparent molecular weight of 100,000 on SDS-polyacrylamide gels stained with silver. Microsequence analysis was performed on the amino-terminal and eight internal tryptic peptides. Peptides sequenced are indicated in Fig. 1. The amino-terminal sequence is identical to the 10 N-terminal residues of human CEL isolated from pancreatic extracts as determined by Wang (24), and to 18 of the first 19 N-terminal amino acids of both rat (7) and porcine (11) CEL.

### Isolation of human CEL cDNA

Peptide sequences from the N-terminus and from an internal peptide (residues 305–321 in Fig. 1) were used to design oligonucleotide probes in accordance with codon frequency usage data (17). The synthetic oligonucleotides were used to screen an oligo(dT)-primed human pancreas cDNA library. Three recombinants were isolated as a result of hybridization to the internal oligonucleotide, none of which hybridized to the N-terminal oligonucleotide. The clone containing the longest cDNA insert, approximately 2.0 kb, was sequenced and the corresponding amino acid sequence was deduced. The identity between the amino acid sequence and internal peptide sequences confirmed that the clone encodes CEL. Failure of this clone to hybridize to the N-terminal oligo, however, indicated that it did not contain the full length CEL cDNA.

After extensive screening of the cDNA library failed to yield a full length clone, the missing 5' portion of the CEL cDNA was cloned directly from human pancreas RNA using the anchored polymerase chain reaction technique (20). Briefly, human pancreas RNA was used as the template for cDNA synthesis and tailed with poly(dG). This cDNA was then amplified by PCR using an oligonucleotide consisting of a poly(dC) tail as the 5' primer, and a CEL specific oligonucleotide as the 3' primer (see Methods). The amplification product extended 447 bases beyond the sequence contained in the 2.0 kb clone. The presence of CEL 5' sequence in the PCR product was confirmed by hybridization to an oligonucleotide corresponding to the N-terminal peptide sequence (not shown). This PCR product was then subcloned into a plasmid vector for sequence determinations.

### CEL cDNA sequence analysis

The nucleotide sequence of the 2.0 kb CEL cDNA and the anchor clone were determined in both orientations using a combination of CEL- and plasmid-specific oligonucleotide primers. To guard against errors that could result from misincorporation during PCR, two in-





```

- 20      Met Gly Arg Leu Gln Leu Val Val Leu Gly Leu Thr Cys Cys Trp Ala Val Ala Ser Ala Ala Lys
1         CCCAGAGGCTGATGCTCAC ATG GGG CGC CTG CAA CTG GTT GTG TTG GGC CTC ACC TGC TGC TGG GCA GTG GCG AGT GCC GCG AAG
3         Leu Gly Ala Val Tyr Thr Glu Gly Gly Phe Val Glu Gly Val Asn Lys Lys Leu Gly Leu Leu Gly Asp Ser Val Asp Ile
86        CTG GGC GCC GTG TAC ACA GAA GGT GGG TTC GTG GAA GGC GTC AAT AAG AAG CTC GGC CTC CTG GGT GAC TCT GTG GAC ATC
30        Phe Lys Gly Ile Pro Phe Ala Ala Pro Thr Lys Ala Leu Glu Asn Pro Gln Pro His Pro Gly Trp Gln Gly Thr Leu Lys
167       TTC AAG GGC ATC CCC TTC GCA GCT CCC ACC AAG GCC CTG GAA AAT CCT CAG CCA CAT CCT GGC TGG CAA GGG ACC CTG AAG
57        Ala Lys Asn Phe Lys Lys Arg Cys Leu Gln Ala Thr Ile Thr Gln Asp Ser Thr Tyr Gly Asp Glu Asp Cys Leu Tyr Leu
248       GCC AAG AAC TTC AAG AAG AGA TGC CTG CAG GCC ACC ATC ACC CAG GAC AGC ACC TAC TGG GAT GAA GAC TGC CTG TAC CTC
84        Asn Ile Trp Val Pro Gln Gly Arg Lys Gln Val Ser Arg Asp Leu Pro Val Met Ile Trp Ile Tyr Gly Gly Ala Phe Leu
329       AAC ATT TGG GTG CCC CAG GGC AGG AAG CAA GTC TCC CGG GAC CTG CCC GTT ATG ATC TGG ATC TAT GGA GGC GCC TTC CTC
111       Met Gly Ser Gly His Gly Ala Asn Phe Leu Asn Asn Tyr Leu Tyr Asp Gly Glu Glu Ile Ala Thr Arg Gly Asn Val Ile
410       ATG GGG TCC GGC CAT GGG GCC AAC TTC CTC AAC AAC TAC CTG TAT GAC GGC GAG GAG ATC GCC ACA CGC GGA AAC GTC ATC
138       Val Val Thr Phe Asn Tyr Arg Val Gly Pro Leu Gly Phe Leu Ser Thr Gly Asp Ala Asn Leu Pro Gly Asn Tyr Gly Leu
491       GTG GTC ACC TTC AAC TAC CGT GTC GGC CCC CTT GGG TTC CTC AGC ACT GGG GAC GCC AAT CTG CCA GGT AAC TAT GGT CTT
165       Arg Asp Gln His Met Ala Ile Ala Trp Val Lys Arg Asn Ile Ala Ala Phe Gly Gly Asp Pro Asn Asn Ile Thr Leu Phe
572       CCG GAT CAG CAC ATG GCA TGC TTT GCT TGG GTG AAG AGG AAT ATC GCG GCC TTC GGG GGG GAC CCC AAC AAC ATC ACC TCT TTC
192       Gly Glu Ser Ala Gly Gly Ala Ser Val Ser Leu Gln Thr Leu Ser Pro Tyr Asn Lys Gly Leu Ile Arg Arg Ala Ile Ser
653       GGG GAG TCT GCT GGA GGT GCC AGC GTC TCT CTG CAG ACC CTC TCC CCC TAC AAC AAG GGC CTC ATC CCG CGA GCC ATC AGC
219       Gln Ser Gly Val Ala Leu Ser Pro Trp Val Ile Gln Lys Asn Pro Leu Phe Trp Ala Lys Lys Val Ala Glu Lys Val Gly
734       CAG AGC GGC GTG GCC CTG AGT CCC TGG GTC ATC CAG AAA AAC CCA CTC TTC TGG GCC AAA AAG GTG GCT GAG AAG GTG GGT
246       Cys Pro Val Gly Asp Ala Ala Arg Met Ala Gln Cys Leu Lys Val Thr Asp Pro Arg Ala Leu Thr Leu Ala Tyr Lys Val
815       TGC CCT GTG GGT GAT GCC GCC AGG ATG GCC CAG TGT CTG AAG GTT ACT GAT CCC CGA GCC CTG ACG TAT GGC TAT AAG GTG
273       Pro Leu Ala Gly Leu Glu Tyr Pro Met Leu His Tyr Val Gly Phe Val Pro Val Ile Asp Gly Asp Phe Ile Pro Ala Asp
896       CCG CTG GCA GGC CTG GAG TAC CCC ATG CTG CAC TAT GTG GGC TTC GTC CCT GTC ATT GAT GGA GAC TTC ATC CCC GCT GAC
300       Pro Ile Asn Leu Tyr Ala Asn Ala Ala Asp Ile Asp Tyr Ile Ala Gly Thr Asn Asn Met Asp Gly His Ile Phe Ala Ser
977       CCG ATC AAC CTG TAC GCC AAC GCC GCC GAC ATC GAC TAT ATA GCA GGC ACC AAC AAC ATG GAC GGC CAC ATC TTC GCC AGC
1027      Ile Asp Met Pro Ala Ile Asn Lys Gly Asn Lys Lys Val Thr Gln Glu Asp Phe Tyr Lys Leu Val Ser Glu Phe Thr Ile
1058      ATC GAC ATG CCT GCC ATC AAC AAG GGC AAC AAG AAA GTC ACG GAG GAG GAC TTC TAC AAG CTG GTC AGT GAG TTC ACA ATC
354       Thr Lys Gly Leu Arg Gly Ala Lys Thr Thr Phe Asp Val Tyr Thr Glu Ser Trp Ala Gln Asp Pro Ser Gln Glu Asn Lys
1139      ACC AAG GGG CTC AGA GGC GCC AAG ACG ACC TTT GAT GTC TAC ACC GAG TCC TGG GCC CAG GAC CCA TCC CAG GAG AAT AAG
381       Lys Lys Thr Val Val Asp Phe Glu Thr Asp Val Leu Phe Leu Val Pro Thr Glu Ile Ala Leu Ala Gln His Arg Ala Asn
1220      AAG AAG ACT GTG GTG GAC TTT GAG ACC GAT GTC CTC TTC CTG GTG CCC ACC GAG ATT GCC CTA GCC CAG CAC AGA GCC AAT
408       Ala Lys Ser Ala Lys Thr Tyr Ala Tyr Leu Phe Ser His Pro Ser Arg Met Pro Val Tyr Pro Lys Trp Val Gly Ala Asp
1301      GCC AAG AGT GCC AAG ACC TAC GCC TAC CTG TTT TCC CAT CCC TCT CGG ATG CCC GTC TAC CCC AAA TGG GTG GGG GCC GAC
435       His Ala Asp Asp Ile Gln Tyr Val Phe Gly Lys Pro Phe Ala Thr Pro Thr Gly Tyr Arg Pro Gln Asp Arg Thr Val Ser
1382      CAT GCA GAT GAC ATT CAG TAC GTT TTC GGG AAG CCC TTC GCC ACC CCC ACG GGC TAC CCG CCC CAA GAG ACC ACA TCT TCT
462       Lys Ala Met Ile Ala Tyr Trp Thr Asn Phe Ala Lys Thr Gly Asp Pro Asn Met Gly Asp Ser Ala Val Pro Thr His Trp
1463      AAG GCC ATG ATC GCC TAC TGG ACC AAC TTT GCC AAA ACA GGG GAC CCC AAC ATG GGC GAC TCG GCT GTG CCC ACA CAC TGG
489       Glu Pro Tyr Thr Thr Glu Asn Ser Gly Tyr Leu Glu Ile Thr Lys Lys Met Gly Ser Ser Ser Met Lys Arg Ser Leu Arg
1544      GAA CCC TAC ACT ACG GAA AAC AGC GGC TAC CTG GAG ATC ACC AAG AAG ATG GGC AGC AGC TCC ATG AAG CCG AGC CTG AGA
516       Thr Asn Phe Leu Arg Gly Tyr Trp Thr Leu Thr Tyr Leu Ala Leu Pro Thr Val Thr Asp Gln Glu Ala Thr Pro Val Pro Pro
1625      ACC AAC TTC CTG CCG TAC TGG ACC CTC ACC TAT CTG CCG CTG CCC ACA GTG ACC GAC CAG GAG GCC ACC CCT GTG CCC CCC
543       Thr Gly Asp Ser Glu Ala Thr Pro Val Pro Pro Thr Gly Asp Ser Glu Thr Ala Pro Val Pro Pro Thr Gly Asp Ser Gly
1706      ACA GGG GAC TCC GAG GCC ACT CCC GTG CCC CCC ACG GGT GAC TCC GAG ACC GCC CCG GTG CCG CCC ACG GGT GAC TCC GGG
570       Ala Pro Pro Val Pro Pro Thr Gly Asp Ser Gly Ala Pro Pro Val Pro Pro Thr Gly Asp Ser Gly Ala Pro Pro Val Pro
1787      GCC CCC CCC GTG CCG CCC ACG GGT GAC TCC GGG GCC CCC CCC GTG CCG CCC ACG GGT GAC TCC GGG GCC CCC CCC GTG CCG
597       Pro Thr Gly Asp Ser Gly Ala Pro Pro Val Pro Pro Thr Gly Asp Ser Gly Ala Pro Pro Val Pro Pro Thr Gly Asp Ser
1868      CCC ACG GGT GAC TCC GGG GCC CCC CCC CTG CCG CCC ACG GGT GAC TCC GGG GCC CCC CCC GTG CCG CCC ACG GGT GAC TCC
624       Gly Ala Pro Pro Val Pro Pro Thr Gly Asp Ser Gly Ala Pro Pro Val Pro Pro Thr Gly Asp Ala Gly Pro Pro Pro Val
1949      GGG GCC CCC CCC GTG CCG CCC ACG GGT GAC TCC GGG GCC CCC CCC GTG CCG CCC ACG GGT GAC GCC GGG CCC CCC CCC GTG
651       Pro Pro Thr Gly Asp Ser Gly Ala Pro Pro Val Pro Pro Thr Gly Asp Ser Gly Ala Pro Pro Val Thr Pro Thr Gly Asp
2030      CCG CCC ACG GGT GAC TCC GGG GCC CCC CCC GTG CCG CCC ACG GGT GAC TCC GGG GCC CCC CCC GTG ACC CCC ACG GGT GAC
676       Ser Glu Thr Ala Pro Val Pro Pro Thr Gly Asp Ser Gly Ala Pro Pro Val Pro Pro Thr Gly Asp Ser Glu Ala Ala Pro
2111      TCC GAG ACC GCC CCC GTG CCG CCC ACG GGT GAC TCC GGG GCC CCC CCT GTG CCC CCC ACG GGT GAC TCT GAG GCT GCC CCT
705       Val Pro Pro Thr Asp Asp Ser Lys Glu Ala Gln Met Pro Ala Val Ile Arg Phe End
2192      GTG CCC CCC ACA GAT GAC TCC AAG GAA GCT CAG ATG CCT GCA GTC ATT AGG TTT TAG CGTCCATGAGCCTTGATCAAGAGGCCAC
2280      AAGAGTGGGACCCAGGGCTCCCTCCCATCTTGAGCTCTTCTGT AATAA AGCCCTCATACCCCAAAAAAAAAAAAAA

```

Fig. 1. The complete nucleotide sequence and predicted amino acid sequence for human CEL cDNA. Nucleotide numbering begins with the first base of the full length cDNA. The N-terminal amino acid of the signal peptide is designated -20, and the mature protein begins at residue 1. The cleavage site for the signal peptide is indicated by a vertical arrow. Peptides that were sequenced directly are underlined. An asterisk indicates the potential N-linked glycosylation site at Asn 187. The putative serine active site is boxed, and the acetylcholine binding site homology is underscored with a broken line. The 16 proline-rich repeats are indicated with numbers above the first residue in each unit. The polyadenylation consensus sequence is underlined.

dependent anchor clones were sequenced and found to be identical, as were overlapping sequences from the anchor clones and the 2.0 kb clone. Fig. 1 shows the complete cDNA nucleotide sequence and the deduced amino acid sequence. The cDNA is 2343 nucleotides and contains two in-frame methionine codons (at nucleotide positions 12 and 20) upstream of the amino-terminal peptide sequence. The ATG triplet at position 20 is likely to be the translational start because the sequence immediately surrounding it (TCACCATGG) strongly agrees with the CCA/GCCATGG consensus sequence for translation initiation determined by Kozak (25), while the sequence surrounding the upstream ATG (GGCTGATGC) does not. The sequence specifies an open reading frame of 742 amino acids which is preceded by 19 nucleotides of 5' untranslated sequence and followed by a 98 nucleotide 3' untranslated region. A polyadenylation signal occurs at nucleotides 2326 to 2331.

Based on the position of the N-terminal peptide sequence, a 20 amino acid leader peptide and a mature protein of 722 amino acids is predicted. Sequences of the eight tryptic peptides were also located in the predicted amino acid sequence, verifying the identity of the clone. The predicted nonglycosylated molecular weight of the protein is 76,271. The larger apparent molecular weight of the purified CEL on SDS-polyacrylamide gels is most likely attributable to the presence of carbohydrate (26). A single potential N-linked glycosylation site occurs at asparagine residue 187.

Comparison of the sequence for human CEL with those for the rat and bovine enzymes reveals an interesting species difference. cDNA clones for the rat protein, which has been known variously as lysophospholipase (13) and pancreatic cholesterol esterase (14), are nearly 300 nucleotides shorter and code for a protein 130 amino acids shorter than human CEL. Despite this difference, alignment of human and rat sequences shows a high degree of homology, with 72% amino acid identity through the length of the rat protein (Fig. 2a). A cDNA clone for bovine CEL has also been reported (12). Although the bovine cDNA is not full length at the 5' end, it appears slightly smaller than that of the rat, coding for a mature protein that is 143 amino acids shorter than human CEL. The human and bovine CEL amino acid sequences have 77% identity (Fig. 2a).

The size discrepancy between CEL from human and the other two species is due to a series of proline-rich tandem repeats near the carboxyl terminus of the protein (Figs. 1 and 2). Each repeat unit consists of 11 amino acids and contains 3 to 5 proline residues. There are 16 repeat units in human CEL, while CEL from rat contains only 3 and a half, a cow only one and a half such units. Alignment of the repeat units from human CEL shows a high degree of conservation between repeats (Fig. 2b).

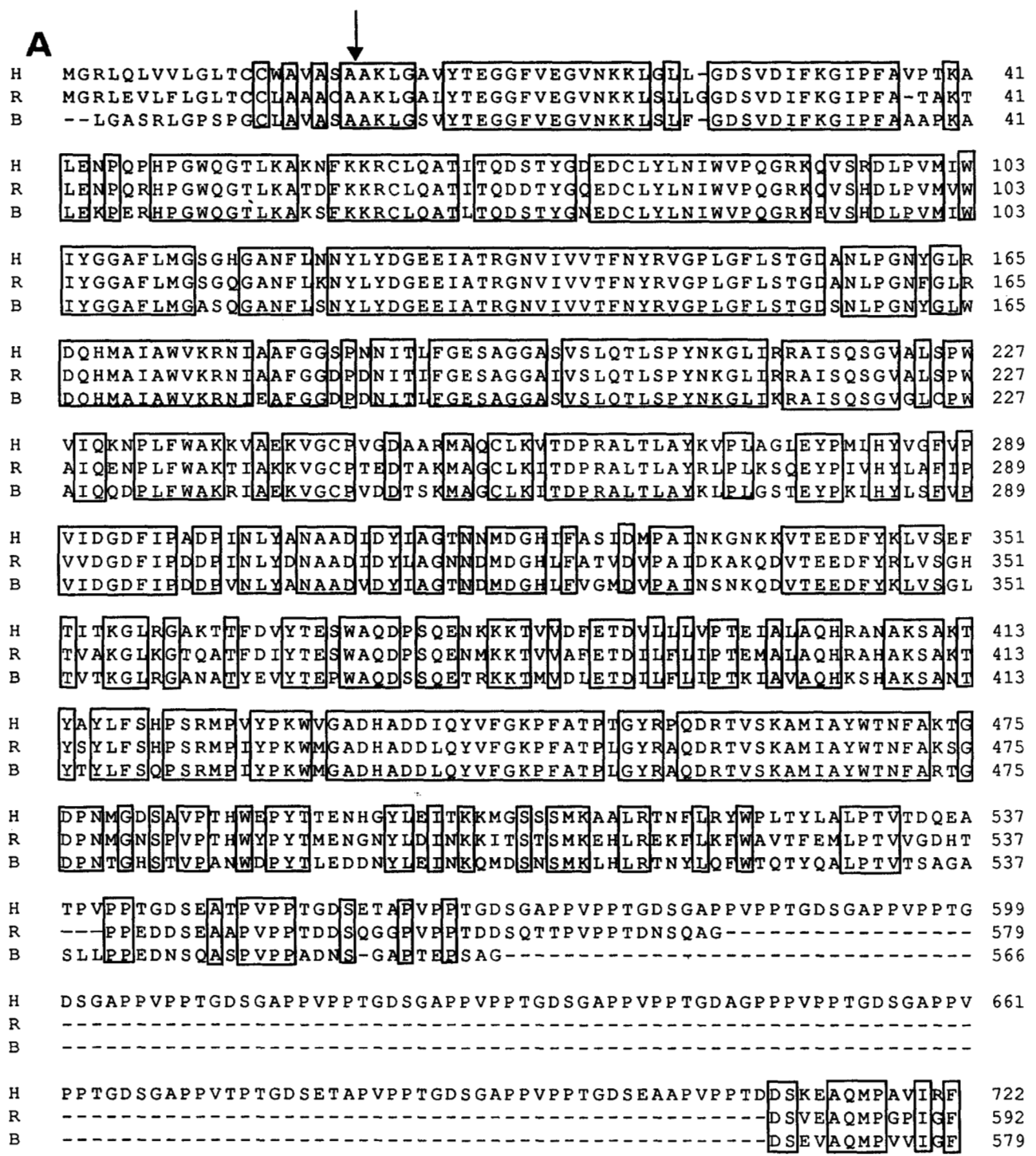
Among the 16 repeat units, however, there is some variation leading to the occurrence of seven types of repeat unit (Fig. 2b, A-G), with a consensus sequence of G-A-P-P-V-P-P-T-G-D-S. The repeat units from rat CEL resemble those from human in both sequence and in the degree of variation among repeats. The first full repeat unit of rat CEL is identical to the final repeat from human CEL. The open reading frame in human CEL continues for 11 amino acids beyond the last full repeat, while rat continues for 16 and the bovine for 19 residues before the termination codon. All three proteins are identical at 8 of the 13 carboxyl-terminal amino acid residues.

Both human and rat CEL cDNAs have short 5' untranslated sequences of 19 and 17-21 nucleotides (13, 14), respectively. A short 5' noncoding leader sequence (11-21 nucleotides) is common to a number of mRNAs encoding pancreatic secretory proteins, including amylase (27), trypsin (28), chymotrypsin B (29), elastases I and II (30), and carboxypeptidase A (31). Although the 3' untranslated sequence from the rat CEL cDNA is longer than that for the human cDNA (203-207 nt for rat vs. 98 nt for human), the 12 additional proline-rich repeats in the human sequence result in a cDNA that is 293 nt longer than rat.

#### CEL mRNA expression

CEL is synthesized in the acinar cells of the pancreas and accounts for 4% of total protein mass in the pancreatic juice (2). Cholesterol esterases with substrate specificities and amino acid compositions similar to CEL have also been detected in rat liver (4, 5) and in human milk (7). To clarify possible structural relationships between the cholesterol esterase activities in pancreas and other tissues, we used the CEL cDNA to study expression of CEL mRNA.

CEL mRNA was abundant in pancreas of both human and rat, as indicated by the intense hybridization signal obtained with 10  $\mu$ g of total pancreas RNA (Fig. 3, top). The human CEL mRNA is approximately 2.3 kb while that for rat is approximately 2.05 kb. This size difference is consistent with the larger size of the human cDNA due to the additional number of proline-rich repeat units. The CEL probe also hybridizes to a much larger RNA species in human and rat pancreas which corresponds in size to 28S RNA. CEL mRNA was not detected in 3  $\mu$ g poly(A+) RNA from adult or fetal human liver, even after 5 days of autoradiography (not shown). However, a cDNA for human alpha-tubulin did hybridize to the liver samples, demonstrating that the liver RNA samples are intact (Fig. 3, bottom). A very faint hybridization signal is visible with rat liver poly (A+) RNA, but this appears as a nonspecific smear throughout the lane, with no appreciable signal corresponding to the size of CEL mRNA. Hybridization of the tubulin cDNA occurs with both rat pan-



**Fig. 2.** Comparison of human, rat, and bovine CEL amino acid sequence and alignment of proline rich repeat units. A: Alignment of human (H), rat (R), and bovine (B) CEL amino acid sequences. Amino acids that are identical for human, rat, and bovine are enclosed in boxes. Horizontal dashes indicate gaps introduced for maximum alignment. Numbers to right indicate position in the mature protein and the vertical arrow indicates signal peptide cleavage site. Rat sequence is from references 13 and 14; bovine sequence is from reference 12. B, upper panel: Human CEL amino acid residues 536-711 are aligned in blocks of 11 amino acid repeat units. Repeats are listed in order as they occur in the protein; consecutive repeats that are identical are not written out, but are indicated by numbers at left. Although there is a total of 16 repeats units, there are only seven different types of repeat sequence. Each type is arbitrarily designated by A-G at the right of each line. Based on the amino acid prevalence at each position, a consensus sequence equivalent to repeat type C was derived. Residues which conform to the consensus are enclosed in the box. B, lower panel: Consensus sequences for the human, rat, and bovine proline-rich repeat unit are compared. Residues common to all three species are boxed. Positions indicated by "X" in the rat sequence are residues for which no consensus residue could be determined.



## B HUMAN CEL REPEATS

REPEATS		TYPE
1-2	E <b>A</b> T P V P P T G D S	A
3	E T A P V P P T G D S	B
4-9	G A P P V P P T G D S	C
10	G A P P V P P T G D <b>A</b>	D
11	G <b>P</b> P P V P P T G D S	E
12	G A P P V P P T G D S	C
13	G A P P V <b>T</b> P T G D S	F
14	E T A P V P P T G D S	B
15	G A P P V P P T G D S	C
16	E <b>A</b> <b>A</b> P V P P T <b>D</b> D S	G

## REPEAT CONSENSUS SEQUENCES

HUMAN	G A P <b>P V P P</b> T G D S
RAT	Q X X <b>P V P P</b> T D D S
BOVINE	Q A S <b>P V P P</b> A D N S

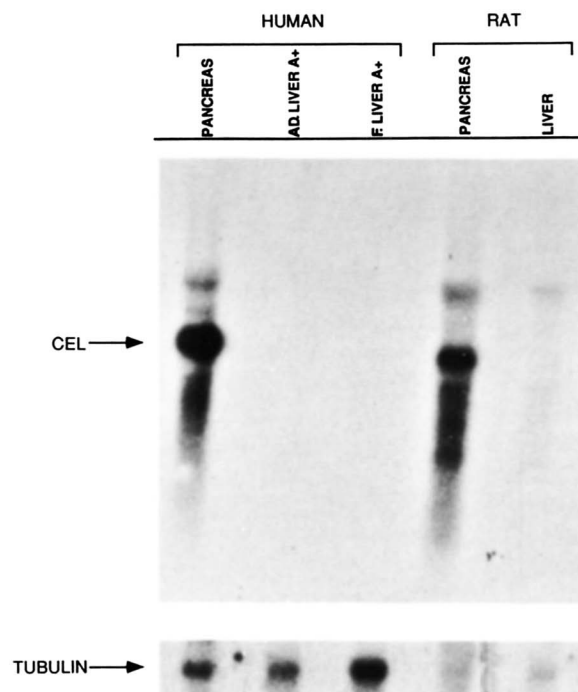
creas and liver RNA, although the signal is less intense than with human samples, presumably due to species sequence differences. CEL mRNA was not detected in other tissues from rat including intestine, spleen, kidney, brain, heart, and testes (not shown). These results suggest that the carboxyl ester lipase activity in liver and intestine may be due to enzymes distinct from pancreatic CEL. Alternatively, the carboxyl ester lipase in these tissues may not be synthesized in situ, but derived from pancreas, as has been previously proposed for the carboxyl ester lipase activity in intestine (32).

## DISCUSSION

We have isolated and characterized the cDNA for human carboxyl ester lipase. The deduced amino acid sequence indicates that mature CEL contains 722 amino acids and is synthesized with a 20 amino acid leader peptide. The predicted nonglycosylated molecular weight of the protein is 76,271. CEL is known to be a glycoprotein

(2) and contains a potential N-linked glycosylation site at Asn residue 187. The deduced sequence is rich in proline (12.2%) in agreement with experimental determinations of amino acid composition of the purified enzyme (2, 7). Interestingly, 68% of the proline residues occur within the final 25% of the protein length. This is due to the tandem repetition of a proline-rich sequence at the carboxy terminus of the protein. The repeat unit, which occurs 16 times, consist of 11 amino acid residues having the consensus sequence G-A-P-P-V-P-P-T-G-D-S (Fig. 2b).

A comparison of human, rat, and bovine CEL sequences revealed a significant size difference between the species. The proline-rich repeats which occur 16 times in human CEL appear only 3 and a half times in rat and one and a half times in bovine CEL, leading to a human protein that is 130 and 143 amino acids longer than rat and bovine, respectively. The difference in apparent molecular weights reported previously for human (100 kDa (2, 7)), rat (65 kDa (7)), and bovine (67 kDa and 72 kDa (12)) CEL is greater than would be accounted for by the additional 130 to 143 residues in the human protein. The increased mass of human CEL may be due to additional carbohydrate chains, possibly in the form of O-linked glycans. While both the human and rat proteins have a single



**Fig. 3.** Northern blot analysis of CEL mRNA expression in pancreas and liver. Ten  $\mu$ g total pancreas RNA (human and rat), or 3  $\mu$ g poly (A+) liver RNA (human adult liver (AD), human fetal liver (F), and rat liver) was fractionated on a denaturing agarose gel, and hybridized with radiolabeled CEL 2 kb cDNA (upper) or human alpha-tubulin cDNA (lower) as described in Methods. Autoradiograms shown were exposed at  $-70^{\circ}\text{C}$  for 48 h (CEL) or 24 h (tubulin). CEL mRNA size was estimated by comparison to RNA size standards (not shown).

potential site for N-linked glycosylation, the abundance of Ser and Thr residues in the proline-rich repeat sequence may provide sites for O-linked glycosylation which are not as prevalent in rat and bovine CEL (33). Whether these sites are in fact glycosylated in human CEL has yet to be determined.

The nature of the species difference in the carboxy terminal portion of CEL prompted us to question whether the prolific repeat sequence in the human cDNA might be a cloning artifact. To verify the sequence in the repeat region, CEL mRNA from an individual other than that used in construction of the cDNA library was amplified by PCR and directly sequenced. This sequence also contained the 16 repeats confirming the authenticity of the clone. Furthermore, both protein composition data and mRNA size measurements are consistent with the cDNA sequence. Experimental determinations of amino acid composition of purified CEL report a proline content of 13.4% for the human protein and 8.3% for rat (7). Comparison of the human and rat cDNA sequences reveals that the additional proline-rich repeats present in the human sequence fully account for this substantial species difference in proline content. The presence of an additional 293 nucleotides in the human cDNA is also evident in the larger mRNA for human CEL compared to rat (Fig. 3).

The primary amino acid sequence of CEL shows strong similarity to members of the serine esterase family. Liver microsomes possess a cholesterol esterase activity distinct from that found in liver cytosol. Human CEL shows 31% amino acid identity with this rat liver microsomal carboxyl esterase, including exact conservation of the G-E-S-A-G motif containing the putative active-site serine. A striking homology also occurs between CEL and acetylcholinesterase (34) and cholinesterase (35), essential enzymes of the nervous system. The homology between CEL and acetylcholinesterase extends through the first 500 amino acids of both proteins with 31% identity and includes the G-E-S-A-G motif. A further similarity between CEL, microsomal carboxyl esterase, acetylcholinesterase, and cholinesterase is the occurrence of four cysteine residues at analogous positions. It has been demonstrated that these cysteines form internal disulfide bridges in human cholinesterase (35). The conservation of these cysteine residues suggests a possible similarity in protein folding among the four proteins. CEL and acetylcholinesterase also share an identical stretch of 11 amino acids (E-D-C-L-Y-L-N-I-W-V-P, residues 78–88 in Fig. 1) thought to play a role in binding of acetylcholine to acetylcholinesterase. The function of an acetylcholine binding site in CEL is unclear, but it may have structural as well as ligand interactive properties (36).

Interestingly, none of the similarities between CEL and the other esterases extend into the proline-rich repeat region at the carboxyl end of CEL. Although the function

of these repeats is not known, the high proline content indicates that this region of CEL cannot fold into a compact secondary structure, and probably adopts a flexible open configuration. The availability of the primary sequence of human CEL provides insight into previously described physical properties of the protein. For example, circular dichroism studies have indicated that the secondary structure of CEL is characterized by a large proportion of  $\beta$ -sheets and  $\beta$ -turns (37). Human CEL also exhibits an anomalously high molecular weight (300 kDa) on gel filtration chromatography which has been attributed to a nonglobular configuration rather than to polymerization (2, 7). These physical properties do not appear to be characteristic of the rat enzyme which differs from human CEL in its lack of an extended proline-rich repeat domain. Proline-rich domains, such as those found in the middle part of IgG heavy chain (38), dystrophin (39), and in cGMP-dependent protein kinase (40), are often referred to as “hinge” regions because they allow flexibility between different portions of the protein which may be required for interaction between molecules. The position of the CEL proline-rich region at the end of the protein, however, makes it unlikely to function as a hinge in the same sense as in the proteins mentioned above.

Proteins with cholesteryl esterase activity have been detected in tissues other than pancreas, including liver, intestine, kidney, brain, aorta, macrophages, and steroidogenic tissues (reviewed in (41)). Immunological studies show that antibodies directed against purified rat pancreatic cholesterol esterase share epitopes with a neutral cholesterol esterase found in liver (4, 5), and with human milk bile salt-stimulated lipase (BSSL) (7). Furthermore, amino terminal sequences of these proteins are very similar or identical to pancreatic CEL and it has been suggested that these esterases originate in the pancreas. To address this question, we examined RNA from extra-pancreatic tissues for evidence of CEL expression. CEL mRNA was not detected in any of the extra-pancreatic tissues examined (human adult and fetal liver, rat liver, intestine, spleen, kidney, brain, heart, and testes). The lack of expression of CEL in rat intestine is in agreement with physiological studies suggesting that the bile salt-stimulated esterase activity in intestine is derived from pancreas (32). The inability to detect CEL mRNA in 3  $\mu$ g poly (A+) RNA from rat liver is in contrast to a recent report in which CEL mRNA was detected in 20  $\mu$ g total rat liver RNA (14). Since poly (A+) RNA accounts for 2–5% of total RNA, the 3  $\mu$ g of poly (A+) liver RNA used in the present study should be equivalent to the amount occurring in 60–150  $\mu$ g of total RNA, and thus should be ample for detection of CEL mRNA. The reason for this discrepancy is not clear, as the human CEL cDNA did cross-hybridize strongly with rat pancreas RNA, and liver RNA samples were demonstrated to be intact by hybridization to an alpha-tubulin probe (Fig. 3). One possible



explanation is that CEL mRNA occurs as an unpolysadenylated mRNA; however, we have not detected CEL mRNA in total liver RNA from either rat or monkey (not shown).

The characterization of a cDNA clone for human CEL should help clarify the relationship between pancreatic CEL and BSSL from human milk. BSSL, which is present in human milk and is thought to have a role in hydrolysis of milk triacylglycerols in the intestine of the newborn, is indistinguishable from CEL by several immunochemical and kinetic criteria. N-terminal analysis of CEL and BSSL shows identical sequence for the 30 amino acids determined; however, the two proteins differ in apparent molecular weight, with CEL estimated at 100 kDa, and BSSL 107 kDa on SDS-polyacrylamide gels (6). This discrepancy may be due to differences in glycosylation or limited proteolysis of the pancreatic enzyme (7, 10). Another intriguing observation is that, like CEL, BSSL exhibits size heterogeneity between human and other species (cat BSSL is 80 kDa) (42). It was suggested that BSSL activity in milk might be derived from CEL synthesized in pancreas and transported through the blood to the lactating mammary gland, but it has since been demonstrated by immunoprecipitation that BSSL is synthesized in lactating human mammary gland (43). With the cDNA clone for human CEL, it will now be possible to clarify the relationship between mRNAs for the pancreatic and milk enzymes, and to determine whether the two originate from a single gene that is normally repressed in mammary gland, but is activated during lactation in humans and a few other species. ■■

Note added in proof: The isolation of a cDNA for human milk bile salt-stimulated lipase was recently reported (Nilsson, J., L. Bläckberg, P. Carlsson, S. Enerbäck, O. Hernell, and G. Bjursell. 1990. *Eur. J. Biochem.* **192**: 543–550). A comparison of protein sequences shows that pancreatic CEL and the milk BSSL in humans are identical.

We wish to thank Dr. H. Okamoto for providing the human pancreas cDNA library and Dr. W. Hunziker for human pancreas RNA. We are especially grateful to Dr. C-S. Wang for making data available prior to publication. We also thank A. Nuckels for assistance in preparation of figures. This work was supported by grants from the National Institutes of Health (HL 28481), the Veterans Administration, and funds from Hoffmann-LaRoche Ltd., Basel, Switzerland. D. A. was supported by Deutsche Forschungsgemeinschaft Grant Am 65/2-1. Manuscript received 17 July 1990 and in revised form 26 October 1990.

## REFERENCES

1. Bradshaw, W. S., and W. J. Rutter. 1972. Multiple pancreatic lipases. Tissue distribution and pattern of accumulation during embryological development. *Biochemistry* **11**: 1517–1528.
2. Lombardo, D., O. Guy, and C. Figarella. 1978. Purification and characterization of carboxyl ester hydrolase from human pancreatic juice. *Biochim. Biophys. Acta.* **527**: 142–149.
3. Rudd, E. A., and H. L. Brockman. 1984. Pancreatic carboxyl ester lipase (cholesterol esterase). In *Lipases*. B. Borgström and H. L. Brockman, editors. Elsevier Science Publishers, Amsterdam. 185–204.
4. Harrison, E. H. 1988. Bile salt-dependent, neutral cholesterol ester hydrolase of rat liver: possible relationship with pancreatic cholesteryl ester hydrolase. *Biochim. Biophys. Acta.* **963**: 28–34.
5. Camulli, E. D., M. J. Linke, H. L. Brockman, and D. Y. Hui. 1989. Identity of a cytosolic neutral cholesterol esterase in rat liver with the bile salt-stimulated cholesterol esterase in pancreas. *Biochim. Biophys. Acta.* **1005**: 177–182.
6. Bläckberg, L., and O. Hernell. 1981. The bile salt-stimulated lipase in human milk: purification and characterization. *Eur. J. Biochem.* **116**: 221–225.
7. Abouakil, N., E. Rogalska, J. Bonicel, and D. Lombardo. 1989. Purification of pancreatic carboxyl-ester hydrolase by immunoaffinity and its application to the human bile salt-stimulated lipase. *Biochim. Biophys. Acta.* **961**: 299–308.
8. Hernell, O. 1975. Human milk lipase. III. Physiological implications of the bile salt-stimulated lipase. *Eur. J. Clin. Invest.* **5**: 267–272.
9. Wang, C-S., M. E. Martindale, M. M. King, and J. Tang. 1989. The nutritional effect of bile salt-activated lipase on kitten growth. *Am. J. Clin. Nutr.* **49**: 457–463.
10. Wang, C-S., and K. Johnson. 1983. Purification of milk bile salt-activated lipase. *Anal. Biochem.* **133**: 456–461.
11. Rudd, E. A., N. K. Mizuno, and H. L. Brockman. 1987. Isolation of two forms of carboxylester lipase (cholesterol esterase) from porcine pancreas. *Biochim. Biophys. Acta.* **918**: 106–114.
12. Kyger, E. M., R. C. Wiegand, and L. G. Lange. 1989. Cloning of the bovine pancreatic cholesterol esterase/lysophospholipase. *Biochem. Biophys. Res. Commun.* **164**: 1302–1309.
13. Han, J. H., C. Stratowa, and W. Rutter. 1987. Isolation of full-length putative rat lysophospholipase cDNA using improved methods for mRNA isolation and cDNA cloning. *Biochemistry.* **26**: 1617–1625.
14. Kissel, J. A., R. N. Fontaine, C. W. Turck, H. L. Brockman, and D. Y. Hui. 1989. Molecular cloning and expression of cDNA for rat pancreatic cholesterol esterase. *Biochim. Biophys. Acta.* **1006**: 227–236.
15. Xu, Q-Y., and J. E. Shively. 1988. Microsequence analysis of peptides and proteins. VIII. Improved electroblotting of proteins onto membranes and derivatized glass-fiber sheets. *Anal. Biochem.* **170**: 19–30.
16. Ben-Zeev, O., C. M. Ben-Avram, H. Wong, J. Nikazy, J. E. Shively, and M. C. Schotz. 1987. Hepatic lipase: a member of a family of structurally related lipases. *Biochim. Biophys. Acta.* **919**: 13–20.
17. Lathe, R. 1985. Synthetic oligonucleotide probes deduced from amino acid sequence data: theoretical and practical considerations. *J. Mol. Biol.* **183**: 1–12.
18. Maniatis, T., E. F. Fritsch, and J. Sambrook. 1982. *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.
19. Sanger, F., S. Nicklen, and A. R. Coulson. 1977. DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. USA.* **74**: 5463–5467.
20. Loh, E. Y., J. F. Elliott, S. Cwiria, L. L. Lanier, and M. M. Davis. 1989. Polymerase chain reaction with single-sided

- specificity: analysis of T cell receptor delta chain. *Science*. **243**: 217-220.
21. Chomczynski, P., and N. Sacchi. 1987. Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. *Anal. Biochem.* **162**: 156-159.
  22. Kroczek, R. A., and E. Siebert. 1990. Optimization of Northern analysis by vacuum-blotting, RNA-transfer visualization, and ultraviolet fixation. *Anal. Biochem.* **194**: 90-95.
  23. Salser, W. A., C-H. Chou, R. C. Davis, J. P. Slavin, and A. Thomason. 1985. In *Leukemia Research: Recent Advances in Biology and Treatment*. R. P. Gale and D. W. Golde, editors. Alan R. Liss, Inc., New York. 349-374.
  24. Wang, C-S. 1988. Purification of carboxyl ester lipase from human pancreas and the amino acid sequence of the N-terminal region. *Biochem. Biophys. Res. Commun.* **155**: 950-955.
  25. Kozak, M. 1987. An analysis of 5'-noncoding sequences from 699 vertebrate messenger RNAs. *Nucleic Acids Res.* **15**: 8125-8148.
  26. Wang, C-S. 1981. Human milk bile salt-activated lipase: further characterization and kinetic studies. *J. Biol. Chem.* **256**: 10198-10202.
  27. MacDonald, R. J., M. M. Crerar, W. F. Swain, R. L. Pictet, G. Thomas, and W. J. Rutter. 1982. Structure of a family of rat amylase genes. *Nature*. **287**: 117-122.
  28. MacDonald, R. J., S. J. Stary, and G. Swift. 1982. Two similar but nonallelic rat pancreatic trypsinogens: nucleotide sequences of the cloned cDNAs. *J. Biol. Chem.* **257**: 9724-9732.
  29. Bell, G. I., C. Quinto, M. Quiroga, P. Valenzuela, C. S. Craik, and W. J. Rutter. 1984. Isolation and sequence of a rat chymotrypsin B gene. *J. Biol. Chem.* **259**: 14265-14270.
  30. MacDonald, R. J., G. H. Swift, C. Quinta, W. Swain, R. Pictet, Jr., W. Nikovits, and W. J. Rutter. 1982. Primary structure of two distinct rat pancreatic preproelastases determined by sequence analysis of the complete cloned messenger ribonucleic acid sequences. *Biochemistry*. **21**: 1453-1463.
  31. Quinto, C., M. Quiroga, W. F. Swain, W. C. Nikovits, D. N. Stondring, R. L. Pictet, P. Valenzuela, and W. J. Rutter. 1982. rat pre-procarboxypeptidase A: cDNA sequence and preliminary characterization of the gene. *Proc. Natl. Acad. Sci. USA*. **79**: 31-35.
  32. Lechene de la Porte, P., N. Abouakil, H. Lafont, and D. Lombardo. 1987. Subcellular localization of cholesterol ester hydrolase in the human intestine. *Biochim. Biophys. Acta*. **920**: 237-246.
  33. Kornfeld, R., and S. Kornfeld. 1976. Comparative aspects of glycoprotein structure. *Annu. Rev. Biochem.* **45**: 217-237.
  34. Schumaker, M., S. Camp, Y. Maulet, M. Newton, K. MacPhee-Quigley, S. S. Taylor, T. Friedmann, and P. Taylor. 1986. Primary structure of *Torpedo californica* acetylcholinesterase deduced from its cDNA sequence. *Nature*. **319**: 407-409.
  35. Lockridge, O., C. F. Bartels, T. A. Vaughan, C. K. Wong, S. E. Norton, and L. L. Johnson. 1987. Complete amino acid sequence of human serum cholinesterase. *J. Biol. Chem.* **262**: 549-557.
  36. Peterson, G. L. 1989. Consensus residues at the acetylcholine binding site of cholinergic proteins. *J. Neurosci. Res.* **22**: 488-503.
  37. Guy, O., D. Lombardo, and J. G. Brahms. 1981. Structure and conformation of human pancreatic carboxyl ester hydrolase. *Eur. J. Biochem.* **117**: 457-460.
  38. Michaelsen, T. E., B. Frangione, and E. C. Franklin. 1977. Primary structure of the "hinge" region of human IgG3. *J. Biol. Chem.* **252**: 883-889.
  39. Koenig, M., and L. M. Kunkel. 1990. Detailed analysis of the repeat domain of dystrophin reveals four potential hinge segments that may confer flexibility. *J. Biol. Chem.* **265**: 4560-4566.
  40. Takio, K., S. B. Smith, K. A. Walsh, E. G. Krebs, and K. Titani. 1983. Amino acid sequence around a "hinge" region and its "autophosphorylation" site in bovine lung cGMP-dependent protein kinase. *J. Biol. Chem.* **258**: 5531-5536.
  41. Brockerhoff, H., and R. G. Jensen. 1974. *Lipolytic Enzymes*. Academic Press, New York. 177-193.
  42. Hernell, O., L. Bläckberg, and T. Lindberg. 1989. Human milk enzymes with emphasis on the lipases. In *Textbook of Gastroenterology and Nutrition in Infancy*. E. Lebenthal, editor. Raven Press, New York. 209-217.
  43. Bläckberg, L., K-A. Ängquist, and O. Hernell. 1987. Bile salt-stimulated lipase in human milk: evidence for its synthesis in the lactating mammary gland. *FEBS Lett.* **217**: 37-41.