Phillips, W. C., & Phillips, G. N. (1985) *J. Appl. Crystallogr. 18*, 3-7.
Press, W. H., Flannery, B. P., Teukolsky, S. A., & Vetterling, W. T. (1989) *Numerical Recipes: The Art of Scientific Computing*, pp 529-538, Cambridge University Press, Cambridge.
Runquist, E. A., & Helmkamp, G. M., Jr. (1988) *Biochim. Biophys. Acta 940*, 10-20.
Selinger, Z., & Lapidot, Y. (1966) *J. Lipid Res. 7*, 174-175.
Shannon, C. E. (1949) *Proc. Inst. Radio Eng. N.Y. 37*, 10-21.
Silvius, J. R. (1990) *Biochemistry 29*, 2930-2938.
Small, D. M. (1986) *The Physical Chemistry of Lipids*, pp 51-56, Plenum Press, New York.
Torbet, J., & Wilkins, M. H. F. (1976) *J. Theor. Biol. 62*, 447-458.
Trumbore, M., Chester, D. W., Moring, J., Rhodes, D., & Herbette, L. G. (1988) *Biophys. J. 54*, 535-543.
Warren, B. E. (1969) *X-Ray Diffraction*, pp 35-38, Addison-Wesley, Reading, MA.
White, S. H., Jacobs, R. E., & King, G. I. (1987) *Biophys. J. 52*, 663-665.
Wiener, M. C., & White, S. H. (1991a) *Biophys. J. 59*, 162-173.
Wiener, M. C., & White, S. H. (1991b) *Biophys. J. 59*, 174-185.
Wiener, M. C., King, G. I., & White, S. H. (1991) *Biophys. J.* (in press).
Worcester, D. L., & Franks, N. P. (1976) *J. Mol. Biol. 100*, 359-378.
Worthington, C. R. (1969) *Biophys. J. 9*, 222-234.
Yeager, M. D., & Feigenson, G. W. (1990) *Biochemistry 29*, 4380-4392.
Zaccai, G., Blasie, J. K., & Schoenborn, B. P. (1975) *Proc. Natl. Acad. Sci. U.S.A. 72*, 376-380.

Structure of the Rat Pancreatic Cholesterol Esterase Gene[†,‡]

Robert N. Fontaine, Christopher P. Carter, and David Y. Hui*
Department of Pathology and Laboratory Medicine, University of Cincinnati College of Medicine, Cincinnati, Ohio 45267-0529
Received November 21, 1990; Revised Manuscript Received May 9, 1991

ABSTRACT: The gene encoding the rat pancreatic cholesterol esterase has been isolated and characterized. Analysis of overlapping genomic clones showed that the cholesterol esterase gene spans approximately 8 kb, containing 11 exons interrupted by 10 introns. The exons ranged in size from 83 to 201 bp except for the last exon, which was 548 bp in length. A TAAATA sequence was present at −31 nucleotides from the transcriptional initiation site. A putative pancreas-specific enhancer sequence was found at −90 bp upstream from the CAP site. Although cholesterol esterase shares three domains of similarity with cholinesterase and acetylcholinesterase, these domains were found to be localized in distinct exons of the cholesterol esterase gene. The organization of the cholesterol esterase gene suggests its divergent evolution with other members of the serine esterase gene family.

The cholesterol esterase of the pancreas, also called carboxyl ester lipase, bile salt stimulated lipase, or nonspecific lipase, catalyzes the hydrolysis of cholesteryl esters to free cholesterol and fatty acids. The enzyme is synthesized in the acinar cells of the pancreas and is released into the intestinal lumen via the pancreatic duct (Guy & Figarella, 1981). The cholesterol esterase is one of the most abundant proteins in the pancreatic juice (Rudd & Brockman, 1984).

Current data from several laboratories (Gallo et al., 1984; Williams et al., 1989) have suggested a role of the cholesterol esterase in mediating cholesterol absorption in the gut. The cholesterol esterase has also been shown to act in concert with pancreatic lipase in lipid absorption (Lindstrom et al., 1988). In addition, the cholesterol esterase is the only enzyme in the pancreatic juice capable of hydrolyzing vitamin esters (Rudd & Brockman, 1984), suggesting its role in catalyzing the lymphatic absorption of fat-soluble vitamins from the diet.

In view of observations that the rate and efficiency of cholesterol absorption may be important determinants in regulating plasma cholesterol level and hypercholesterolemia (Kesaniemi & Miettinen, 1987), it is important to understand the regulatory mechanism(s) that may control pancreatic cholesterol esterase gene expression. Unfortunately, very little information is currently available concerning the regulation of cholesterol esterase biosynthesis under physiological conditions. The lack of progress may be related to the limited availability of information on the structure of the cholesterol esterase gene. In recent studies, progress has been made in several laboratories by cloning and sequencing of the cDNA for pancreatic cholesterol esterase from different species (Han et al., 1987; Kissel et al., 1989; Kyger et al., 1989; Hui & Kissel, 1990). In this paper, we report the isolation and characterization of the rat cholesterol esterase gene. Information obtained in this study will be useful for future studies aimed at understanding the mechanism governing the regulation of cholesterol esterase gene expression. The results of this study also suggest that the cholesterol esterase gene may have evolved differently from other members of the serine esterase gene family.

Experimental Procedures

Genomic Cloning. A rat genomic library in λ-DASH vector was obtained from Stratagene (catalog no. 247211). The genomic library was screened by the filter hybridization

[†]This research was supported by Grant DK 40917 from the National Institutes of Health.
[‡]The nucleic acid sequence in this paper has been submitted to the GenBank/EMBL database under Accession Number M69157.
*To whom correspondence should be addressed at the Department of Pathology.

0006-2960/91/0430-7008$02.50/0

technique as described by Benton and Davis (1977). Two different cDNA probes corresponding to the 5′ 400-bp and 3′ 1-kb fragments of the full-length rat cholesterol esterase cDNA were used as probes for screening of the library. The probes were isolated by *Sma*I and *Bal*I digestion of the full-length rat cholesterol esterase cDNA (Kissel et al., 1989). The cDNA fragments were labeled with ^{32}P by the random priming method to a specific activity of approximately 1×10^8 cpm/μg. The conditions of hybridization were 50% formamide/5× concentrated SSC (1× SSC is 150 mM NaCl and 15 mM sodium citrate)/5× Denhardt's solution/0.1% SDS/2 mg/mL of denatured salmon sperm DNA at 40 °C. Filters were washed with buffer containing 1× SSC and 0.1% SDS at 40 °C. The filters were exposed to Kodak XAR-2 films for 18 h to identify the positive clones.

Subcloning and Sequencing of the Genomic Clones. Five recombinant clones containing cholesterol esterase sequences were identified from screening of 2×10^6 plaques. The DNA inserts of these clones were digested with various restriction endonucleases, electrophoresed on 1% agarose gels, and then transferred to nitrocellulose paper. The DNA were blotted with various fragments, or oligonucleotides, corresponding to different regions of the cholesterol esterase cDNA. Hybridization with cDNA probes was performed as described above. For hybridization studies with oligonucleotide probes, the synthetic oligonucleotides were labeled with ^{32}P by using the 5′ kination method and hybridization was performed in the absence of formamide. Restriction fragments hybridizing to the probes were isolated and subcloned into pTZ-18U or pTZ-19U vectors for propagation and further analysis. The recombinant plasmids were used to transfect *Escherichia coli* JM101 cells. Single-stranded DNA in filamentous phage particles was isolated from the medium of transfected cells cultured in the presence of M13KO7 helper phages. The nucleotide sequences of the inserts in the single-stranded DNA were then determined by the dideoxy chain-termination method (Sanger et al., 1977) with modified T7 DNA polymerase obtained from Pharmacia Biotechnology. Sequencing reactions were initiated either by using a universal M13 sequencing primer or with synthetic oligonucleotide primers corresponding to known sequences within the cholesterol esterase cDNA. Sequencing information was obtained from overlapping clones covering the entire cholesterol esterase gene.

Southern Blot Analysis. Rat genomic DNA was isolated from the liver of a male Sprague-Dawley rat by using the rapid DNA preparation procedure described by Davis et al. (1986). The cellular DNA and a cloned λ DNA were digested with restriction enzymes and subjected to blot analysis with the full-length rat cholesterol esterase cDNA (Kissel et al., 1989) used as the probe. The conditions used for hybridization were identical with those described above for screening of the genomic library. The filters were washed with 0.1× SSC and 0.1% SDS at 60 °C.

Results and Discussion

Five overlapping genomic clones containing rat cholesterol esterase sequences were isolated and characterized. The restriction map of the insert sequences was determined for five enzymes (*Sac*I, *Bam*HI, *Hin*dIII, *Pst*I, and *Acc*I) and showed the relatedness of the five clones (data not shown). Several synthetic oligonucleotides corresponding to various regions of the cholesterol esterase cDNA were used as hybridization probes to estimate the size of the DNA inserts and to confirm that the DNA was derived from overlapping clones. The synthetic oligonucleotide probes used in the experiment included the 5′ untranslated sequence of cholesterol esterase cDNA, the sequence corresponding to the signal peptide, and sequences surrounding the active-site serine (DiPersio et al., 1990) and histidine (DiPersio et al., 1991) in the catalytic triad of cholesterol esterase. Results indicated that three of the four oligonucleotide probes hybridized to DNA fragments obtained from the five recombinant phages. Therefore, these clones must contain sequences of the cholesterol esterase gene. Moreover, the DNA from one of these clones showed a positive reaction with all of the probes, including the probe corresponding to the 5′ untranslated region of cholesterol esterase mRNA. This result indicated that the sequence of the cholesterol esterase gene may be obtained by sequencing overlapping inserts from these recombinant phages.

Table I: Differences in Nucleotide Sequence between the Cholesterol Esterase cDNA and the Gene

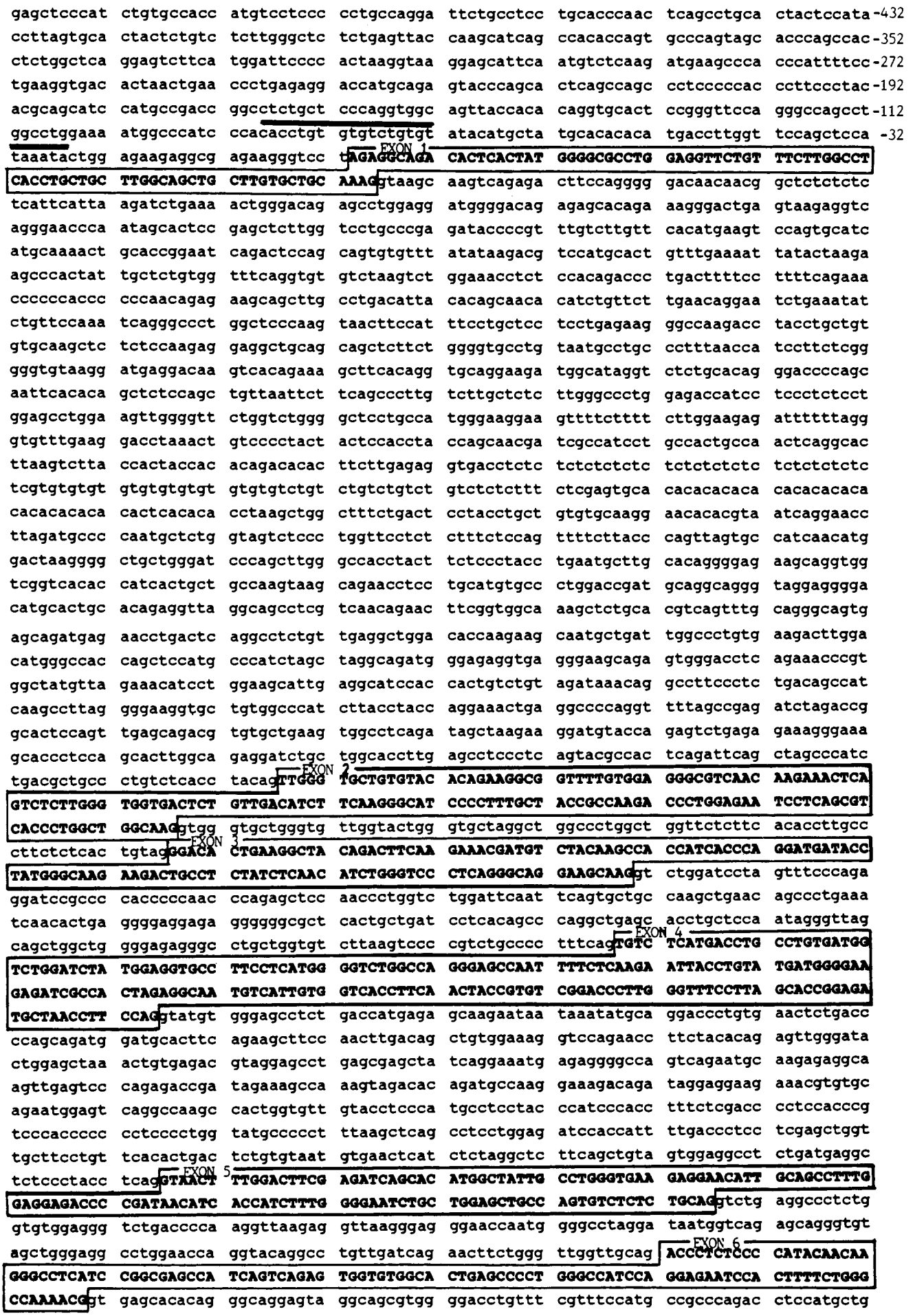
residue no.[a]	cDNA[b]	gene[c]
493	Thr (ACG)	Met (ATG)
588	Gly (GGT)	Val (GTC)
589	Pro (CCC)	Ala (GCC)
591	Gly (GGC)	Ala (GCG)

[a] The amino acid residues are numbered with the first amino acid of the nascent protein as 1. [b] From Kissel et al. (1989). [c] Present study.

Restriction fragments yielding positive reaction upon hybridization with cholesterol esterase cDNA were then isolated and subcloned into pTZ vectors for sequence determination. The entire sequence of the cholesterol esterase gene was determined (Figure 1). The data revealed that the rat cholesterol esterase gene spans approximately 8 kb. Furthermore, when the cholesterol esterase gene sequence was compared with that of the cDNA (Kissel et al., 1989), results revealed a gene structure containing 11 exons interrupted by 10 introns (Figure 2). Southern blot analysis of genomic DNA isolated from rat liver revealed a restriction map similar to that observed with restriction digestion of the cloned DNA (Figure 2). This result strongly suggests that the pancreatic cholesterol esterase gene is a single-copy gene in the rat genome.

The genomic sequence obtained in this study included 494 bp of the 5′ flanking region of the cholesterol esterase gene. A TATA-box-like sequence, TAAATA, was found at −31 bp from the transcriptional initiation site as determined by Han et al. (1987). No CAAT-box structure or GC boxes were evident in this region of the cholesterol esterase gene. However, it is interesting to note that at −90 bp from the CAP initiation site is a sequence with 64% similarity to the consensus sequence of rat pancreas-specific enhancer element (Boulet et al., 1986). Whether this sequence participates in the tissue-specific expression of the cholesterol esterase gene in the pancreas remains to be determined.

The analysis of the cholesterol esterase gene organization revealed that all the exons and introns are small and are less than 2 kb. In fact, exons 1–10 ranged in size from 83 to 201 bp. Intron 1 was the largest intron, containing 1991 bp, while exon 11 was the largest exon in the cholesterol esterase gene with 548 bp. The nucleotide sequence in each exon matched completely with the cholesterol esterase cDNA sequence with four exceptions. These differences in cDNA and genomic sequences are outlined in Table I. In each case, these differences may be due to DNA polymorphism, possible sequencing error, or differences in the source of the mRNA and genomic DNA used for the construction of the cDNA and genomic DNA libraries. The latter possibility is supported by the fact that the cDNA library was constructed by priming of mRNA isolated from Sprague-Dawley rats while the genomic library was made by using DNA from Wistar rats. In any event, sequence comparison between the cholesterol est-


EXON 1
EXON 2
EXON 3
EXON 4
EXON 5
EXON 6

tcccagacct tctgtttgcc cttagtcatt acaagcctca ccaccggcct tctctaacct cctgcctctg cctttcccta
EXON 7
GATCGCTAAG AAGGTGGGAT GCCCCACAGA GGATACCGCC AAGATGGCTG GGTGTCTGAA GATCACAGAT CCCCGAGCCT
TGACACTGGC CTACAGGTTG CCCTTGAAAA GCCAGGAGgt gtgagttggc gctcctgggc gggaggcatg tagggtggaa
gaggaaggca ccttgtaaga caaggctgag ctcctgacaa gacagaaagg tgccacactg gcctgtctct accctctagc
aggccccaat ggtaaagaag tgtctcccgg tactccaggg acttttaaga accgatatat ggccaggcac acacctgtaa
tcgcagcaga aaggaggcag aagcaggcag atctccgtgg gttcaaagct ggtctgttct acacagtgag aggacagcca
gagctacaca gtgggactat ctttaaaaaa aaaaaaagat aggaagttag ggagggaggg agggatagaa ggaatgagga
aaggagtcaa acattaggtc gggcagcaga aacccatacg tagtttccgg agatatgtgc ccaacccaag ctttgtaaag
tggcaatcat ttgtaaattt ccgaacacct cctgtggcat ttggttcttc cagctgaggc atgaagaacc agcctgcgtg
cctaggacac gggacttgtc tctgcttacc cttccgaagc aaaccaggca gcctgagcat ccacaccttc tgcctttctt
ttttctttcc atgccgtaga ccaaacccag agcctgtgct ctgcctccga gcacctcgcc acccaaatgg ctattttatt
tggagagcca ctgaagaaaa cctggctgtg ctggccccaa ggaagccagg gtgtctgcaa gtctagaggc caggcagcga
ccacacagcg gccagcgacc ccgtatccct ggtgtaggtt caataaggag agagcgagca gctgggggtg ctcaagcagg
aggacagaca gccgtggtcc tcagagcagt ggactctcca tgtggtcgtg agcaaagtct gggaatgcac cggaagtgtg
ccagtgacat ctagcaggtt acactaggga cactgctcag catcccacac ggaggacagt caccccctgtc actagctgac
tgtggccaaa ggtcctgggg ctggggttga ggaaccttgg tctggtaaga aaagaaagag gtcaagtcta gtacaaaagg
ccaagaggag gaagaggcag gcacttgaag gtgttggctc tccacccagT ACCCCATTGT GCACTACCTG GCCTTCATCC
EXON 8
CTGTCGTCGA TGGTGACTTC ATTCCTGATG ATCCCATCAA CCTGTACGAC AACGCTGCTG ACATTGACTA CTTAGCGGGT
ATTAATGACA TGGATGGCCA CCTGTTTGCT ACAGTTGACG TGCCCGCCAT CGACAAGGCC AAGCAGGATG TCACAGAGGA
Ggtggtgctg tgttgggaga gcagggaaga catgctccgg tccacaccag tgtagtacct ccatgtttga tcattctcat
EXON 9
gcagggagGA CTTCTACAGG CTAGTCAGTG GACACACTGT CGCCAAGGGG CTTAAAGGCA CCCAAGCCAC CTTTGACATC
TACACTGAGT CCTGGGCCCA GGACCCGTCC CAGGAGAACA TGAAGAAGAC AGTGGTGGCC TTTGAGACTG ACATACTCTT
CCTGATCCCC ACAGAGACGG CTCTGGCCCA GCACAGAGCC CATGCCAAgt aaggttctgc agagcgtctt gagagggctg
cctaggcttc ttggctcact aaacaatgaa catagctctg ggcttcctat cagtcaaggt taaaagtata ataggagtat
atactttgtc taatcatgtg agaaatgtga gtttaggtgg aaaatgaatg acccttgatc ctggattatg ggaagaagga
ggcttctgaa tattgtaggg tggagacacg aacattgaat gtgggctatg gtgtttcttt cccttactat caagaacttc
agttcagacg tgggggtggt ggcacatact tttaatccca gccactcaca aggctgatct tctgtgagtt caagaccagt
ctggcttata tagtaagttc gaggccggcc aaggctacca taaaaagacc ctgtcttaaa caaacaaaca aacaaacaaa
caaacaaaca aacaaacaag ccctagtttt ttggtgagtc tcactggtcc cctcaggccc tgctctgttc ctcctgactt
EXON 10
ctctccaggg ccacctccca agggagcgcc ctaccaccca gctgcctctg cttcccacag GAGTGCCAAG ACCTACTCTT
ACCTGTTTTC CCACCCTTCA CGAATGCCTA TCTACCCAAA ATGGATGGGG GCAGACCACG CTGATGACCT CCAGTACGTC
TTTGGGAAGC CCTTTGCCAC CCCACTGGGC TACCGGGCCC AAGACAGGAC TGTCTCCAAG GCCATGATTG CCTACTGGAC
CAACTTTGCC AAGAGTGGgt aagatgtgtg gtggagtcca ggaccaaggg cagcaggctt gagggctcct ccatcgcttt
ggcctgtagg tctgtctctc acttggcaaa cttcttatct caagccttag tagctgcaca tgagcagaag agcttcagcc
ccagccccag ccccagcccc agccccagcc ccagggccga acacatgctc tccatgacag cctcgagagg ttggctagga
tggccaagct gtctgtgcag agctgtctgg gagcacaggc ctccagtctc gttgcaggcc cctgctcatt ctcttcttac
EXON 11
tctgcagGGA CCCCAACATG GGCAACTCAC CCGTGCCCAC ACACTGGTAC CCTTATACCA TGGAGAATGG TAACTACCTG
GACATCAATA AGAAAATAAC CAGCACCTCC ATGAAGGAGC ACCTAAGGGA AAAGTTCCTC AAGTTCTGGG CTGTGACATT
CGAGATGCTG CCCACTGTGG TTGGTGACCA CACTCCCCCT GAGGATGACT CAGAGGCTGC CCCCGTCCCA CCTACAGACG
ACTCCCAGGT TGTTCCTGTC CCACCTACAG ATGACTCTCA GACAACACCG GTGCCCCCAA CAGACAACTC TCAGGCTGGT
GACTCTGTGG AGGCTCAGAT GCCTGTCGCC ATTGCGTTCT AAAGTCCTAT AAACCGGGGC TAGAGATGGC TCAGGAGCTA
AGAGCTCTTC TTCCACTGTT CTTCTGAAGG TCCTGTGTTC AATTCCCAGC ACCCACATTC GTGCTTACAG CTGTCTGATT
CCGTCTACTG GTGTGCAGAT GTACACGCAG ACAAGCGCCT ATATATAGAA AATTCATACA TAAATAAAGT CTTATAAgcc
ttggtcccag gtcctcctct tattagctct ccgtcaagaa agcctcagcc attaccatct ggtatctgat ttgctcctga
aacagcctgg aggaaggagg gacaccctgt gaatggtgac agcacttggc ccattccctc tgagcccaag gctgcacgag
tcaaaggact ctctttgggg atgggtcagt agcagtgtag tggggagcga ggggttttgg ttctttttt cggagctggg
gactgaaccc agggcctctg cgttcctagg caagcgctct accactgagc taaatcccca acccgtgagg ggaggctctt
tcaggccatg agcaacaccc tgggattctg gaggatccca gctatgatgg cgatcctgca gcccttggcc agcatgtgtg
gcagagagcg gggccatctg aggcacaggc agctatagga agctggaacc tgcttcatag gccacaggat gttaatggga

FIGURE 1: Nucleotide sequence of the rat pancreatic cholesterol esterase gene. The exons are numbered and their sequences are indicated in bold uppercase letters. The intronic sequences are represented in lowercase letters. The sequence of the 5′ flanking region is numbered with the first base preceding the transcription initiation site as −1 (Han et al., 1987). The putative pancreatic tissue-specific enhancer, the TATA box, the translation termination codon, and the polyadenylation signal are marked with heavy lines above the respective sequences.

Table II: Exon–Intron Organization of the Rat Pancreatic Cholesterol Esterase Gene

exon no.	exon size (bp)	sequence at exon–intron junction							intron length (bp)	preceding or interrupted amino acid
		5′ splice donor				3′ splice acceptor				
1	83	GCT	GCA	AAG	gtaagc....ctgtctcacctacag	TTG	GGT	GCT	1991	Lys
2	151	TGG	CAA	G	gtgggt....cttctctcactgtag	GG	ACA	CTG	79	Gly
3	123	AAG	CAA	G	gtctgg....gtctgccccttcag	TG	TCT	CAT	238	Val
4	198	CTT	CCA	G	gtatgt....ctctccctacctcag	GT	AAC	TTT	563	Gly
5	131	TCT	CTG	CAG	gtctga....ctgggttggttgcag	ACC	CTC	TCC	155	Gln
6	108	GCC	AAA	ACG	gtgagc....tctgcctttccctag	ATC	GCT	AAG	153	Thr
7	117	AGC	CAG	GAG	gtgtga....tggctctccaccag	TAC	CCC	ATT	1130	Glu
8	192	ACA	GAG	GAG	gtggtg....ttctcatgcagggag	GAC	TTC	TAC	87	Glu
9	201	CAT	GCC	AA	gtaagg....ctctgcttcccacag	G	AGT	GCC	572	Lys
10	198	AAg	AGT	GG	gtaaga....cttcttactctgcag	G	GAC	CCC	309	Gly
11	548									

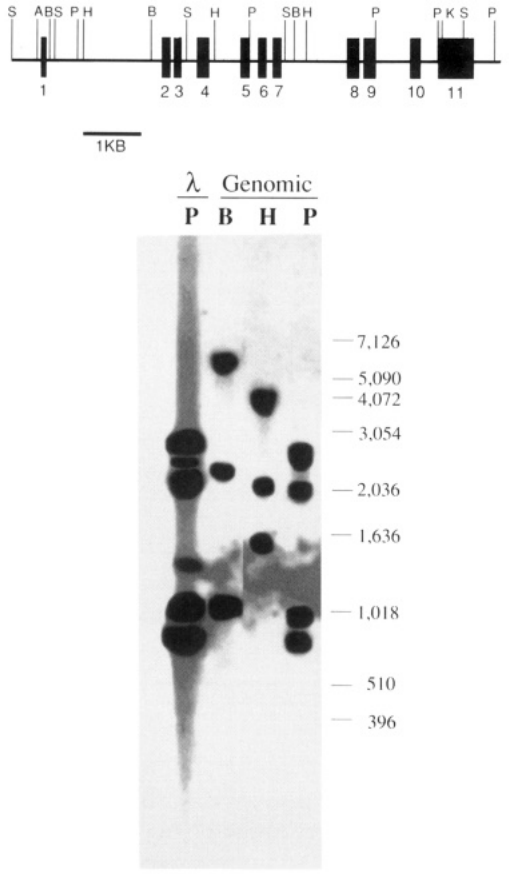


FIGURE 2: Structure of the rat pancreatic cholesterol esterase gene. The schematic structure of the rat pancreatic cholesterol esterase gene is presented with exons indicated by numbered blocks and introns indicated by connecting lines. The major restriction sites utilized for subcloning are shown with letters: S, *Sac*I; A, *Acc*I; B, *Bgl*II; H, *Hin*dIII; and K, *Kpn*I. The bottom panel shows the Southern blot analysis of rat genomic DNA and its comparison with a single digest of the cloned DNA. Rat liver DNA (genomic) or the λ DNA containing the cholesterol esterase gene (λ) were digested with the following restriction enzymes: P, *Pst*I; B, *Bgl*II; or H, *Hin*dIII. The Southern blot was probed with a full-length ^{32}P-labeled cholesterol esterase cDNA. The hybridizing DNA bands were visualized by exposure to X-ray films. The size of the DNA bands hybridizing to the cDNA probe was determined by comparison with standard DNA. The numbers to the right of the blot refer to the size, in base pairs, of the DNA standards.

erase cDNA and its gene revealed that the nucleotide sequences surrounding the intron–exon junctions conformed to the consensus sequence for splice junctions in eukaryotic genes (Mount, 1982). Each intron was found to begin with the dinucleotide GT and end with the dinucleotide AG (Table II).

The position of introns and exons in eukaryotic genes often signals the structural and functional domains of the protein (Sudhof et al., 1985). The structural organization of the cholesterol esterase gene appears to support this hypothesis. Exon 1 of the cholesterol esterase gene encodes the 5′ untranslated region and the signal peptide region of the mRNA. Exon 3 contains the putative heparin-binding domain of the cholesterol esterase (Kissel et al., 1989) and an 11 amino acid domain with complete sequence identity with other serine esterases (Hui & Kissel, 1990). A second domain of similarity between cholesterol esterase and other serine esterases was found to be localized to exon 5. This domain encodes the active-site serine of the catalytic triad. Recent studies using chemical modification and site-specific mutagenesis techniques have identified serine-194, within this domain, as the active-site serine of cholesterol esterase (DiPersio et al., 1990). The third domain of similarity between the cholesterol esterase and other serine esterases encompasses the active-site histidine residue, histidine-435, of cholesterol esterase (DiPersio et al., 1991). This domain was found to be encoded by exon 10 in the cholesterol esterase gene.

In addition to the domains with similarities to serine esterases, another domain that has been identified previously contains repeating units enriched with proline (P), glutamate (E), serine (S), and threonine (T). This domain was postulated to signal for rapid degradation of the protein (Kissel et al., 1989). Interestingly, all the repeating sequences were found to be localized within exon 11. This exon, the largest exon in the cholesterol esterase gene, may also display the greatest divergence among species since the rat and human cholesterol esterases differ in the number of such proline-rich repeating units. Previous studies have shown that the rat cholesterol esterase contains four PEST sequences (Kissel et al., 1989) while the human protein contains 16 such repeats (Hui & Kissel, 1990). It is not known whether the repeating sequences of cholesterol esterase in each species are encoded by one large exon or are the result of exon duplication events.

Analysis of the intron–exon organization of cholesterol esterase gene evoked some interesting speculation concerning the evolutionary development of the gene. For example, although cholesterol esterase has been shown to contain domains with significant homology with other serine esterases, the gene structure of cholesterol esterase and cholinesterase are strikingly different. The human cholinesterase gene (Arpagaus et al., 1990) contains only four exons and spans over 70 kb. Likewise, the *Torpedo* acetylcholinesterase gene also contains four exons (with alternatively spliced exons 3 and 4) and spans over 40 kB (Maulet et al., 1990). Moreover, the serine, histidine, and acidic amino acid domains of the catalytic triad of cholinesterase and acetylcholinesterase are all encoded by the same exon (exon 2) in these two genes (Arpagaus et al., 1990; Gibney et al., 1990; Maulet et al., 1990). In contrast, the cholesterol esterase gene, consisting of 11 exons spanning

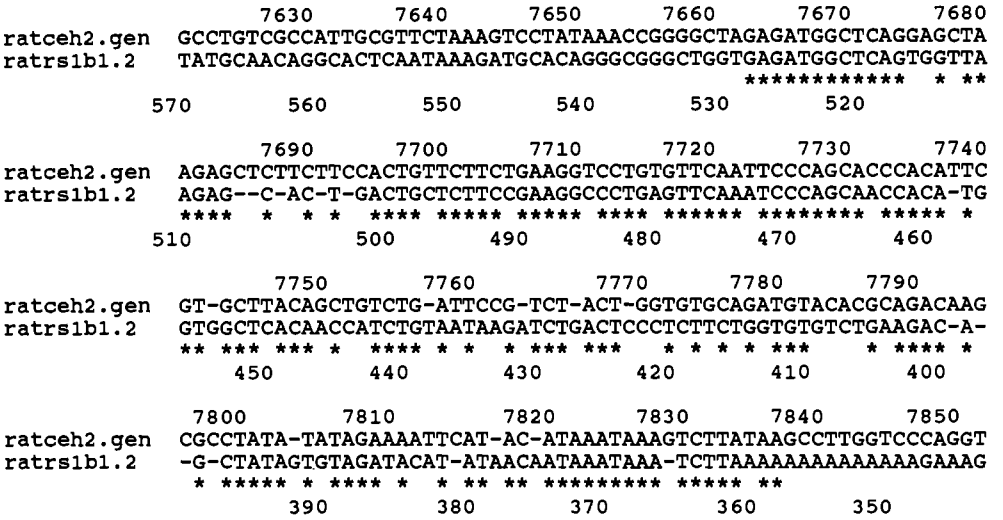


FIGURE 3: Sequence similarities between rat cholesterol esterase gene (ratceh2.gen) and repetitive sequence 1b1 (ratrs1b1.2). The region in the rat cholesterol esterase gene that is similar to the 1b1 repetitive sequence is shown. The numbers above the sequence represent the nucleotide numbers in the cholesterol esterase gene. The numbers below the sequence are the nucleotide numbers for the consensus 1b1 repetitive sequence. Asterisks indicates identical sequences.

Table III: *rat55rep* Repetitive Sequences in Cholesterol Esterase Gene

intron	nucleotides	similarity (%)	length (bp)
5′ flanking (R/C)[a]	104–146	62	42
1	1565–1730	72	165
1	2476–2547	60	71
2	2723–2770	63	47
3	3141–3186	60	45
4 (R/C)[a]	3714–3754	60	40
5	4060–4208	60	148
6	4458–4510	61	52
7	4905–4961	61	56
9	6423–6586	60	163
9 (R/C)[a]	6536–6665	61	129
10	7027–7259	61	232
3′ flanking (R/C)[a]	7924–8189	62	265

[a] R/C = Reversed and complement strand.

8 kB, encodes the three catalytic domains in at least two, and most likely three, separate exons.

The multiple intron and exon structure of the cholesterol esterase gene resembles more closely the organization of the genes important for lipid modification enzymes. For example, human hepatic lipase, lipoprotein lipase, cholesteryl ester transfer protein, and canine pancreatic lipase are encoded by genes containing 9, 10, 16, and 13 exons, respectively (Cai et al., 1989; Kirgessner et al., 1989; Agellon et al., 1990; Mickel et al., 1989). However, no sequence homology could be identified between the cholesterol esterase and these proteins. Thus, these comparative studies suggest that cholesterol esterase may have evolved by the shuffling of specific exons, such as exons containing the active-site domains of esterases, to a lipaselike gene. If the latter hypothesis is correct, then the cholesterol esterase gene is a product of convergent evolution between serine esterase and lipase gene families. In support of this hypothesis, the *rat55rep* repetitive sequence was found to be present in 9 of the 10 introns and in the flanking regions of the cholesterol esterase gene (Table III). The repetitive sequences are present in various lengths, ranging from 40 to 272 bp, and share a >60% nucleotide identity with the *rat55rep* sequence. Since many of these short dispersed repetitive sequences resemble transposable genetic elements in microorganisms (Hardman, 1986), it is possible that these short *rat55rep* sequences played a role in exon shuffling and the development of the cholesterol esterase gene.

In addition to the *rat55rep* repeating sequences, exon 11 of the cholesterol esterase gene contains a 175-bp domain that shares a 70% similarity with the *rat1b1* repetitive sequence (Figure 3). This domain corresponds to the entire 3′ untranslated region of cholesterol esterase mRNA. Although the presence of repetitive sequences within exons is unusual, their role in regulation of gene expression has been documented in at least two cases (Lone et al., 1986; Glaichenhaus & Cuzin, 1987). Recent research in our laboratory have shown that cholesterol esterase gene expression may be regulated at the level of translation (Huang & Hui, 1991); it is possible that the repetitive sequence at the 3′ untranslated region of the cholesterol esterase mRNA may be involved with translational control via interaction with translational-control cytoplasmic RNA (Sarkar, 1984).

ACKNOWLEDGMENTS

We acknowledge the superb technical assistance of James Kissel. We are grateful to Mike Hughes in the Department of Molecular Genetics, Biochemistry, and Microbiology for the synthesis of oligonucleotide probes used in this study.

REFERENCES

Agellon, L. B., Quinet, E. M., Gillette, T. G., et al. (1990) *Biochemistry 29*, 1372–1376.

Arpagaus, M., Kott, M., Vatsis, K. P., et al. (1990) *Biochemistry 29*, 124–131.

Benton, W. D., & Davis, R. W. (1977) *Science 196*, 180–182.

Boulet, A. M., Erwin, C. R., & Rutter, W. J. (1986) *Proc. Natl. Acad. Sci. U.S.A. 83*, 3599–3603.

Cai, S. J., Wong, D. M., Chen, S. H., et al. (1989) *Biochemistry 28*, 8966–8971.

Davis, L. G., Dibner, M. D., & Battey, J. F. (1986) in *Basic Methods in Molecular Biology*, Elsevier Press, New York.

DiPersio, L. P., Fontaine, R. N., & Hui, D. Y. (1990) *J. Biol. Chem. 265*, 16801–16806.

DiPersio, L. P., Fontaine, R. N., & Hui, D. Y. (1991) *J. Biol. Chem. 266*, 4033–4036.

Gallo, L. L., Clark, S. B., Myers, S., et al. (1984) *J. Lipid Res. 25*, 604–612.

Gibney, G., Camp, S., Dionne, M., et al. (1990) *Proc. Natl. Acad. Sci. U.S.A. 87*, 7546–7550.

Glaichenhaus, N., & Cuzin, F. (1987) *Cell 50*, 1081–1089.

Guy, O., & Figarella, C. (1981) *Scand. J. Gastroenterol. 16* (Suppl. 67), 59-61.
Han, J. H., Stratowa, C., & Rutter, W. J. (1987) *Biochemistry 26*, 1617-1625.
Hardman, N. (1986) *Biochem. J. 234*, 1-11.
Huang, Y., & Hui, D. Y. (1991) *J. Biol. Chem. 266*, 6720-6725.
Hui, D. Y., & Kissel, J. A. (1990) *FEBS Lett. 276*, 131-134.
Kesaniemi, Y. A., & Miettinen, T. A. (1987) *Eur. J. Clin. Invest. 17*, 391-395.
Kirgessner, T. G., Chuat, J. C., Heinzmann, C., et al. (1989) *Proc. Natl. Acad. Sci. U.S.A. 86*, 9647-9651.
Kissel, J. A., Fontaine, R. N., Turck, C., et al. (1989) *Biochim. Biophys. Acta 1006*, 227-236.
Kyger, E. M., Wiegand, R. C., & Lange, L. C. (1989) *Biochem. Biophys. Res. Commun. 164*, 1302-1309.
Lindstrom, M. B., Sternby, B., & Borgstrom, B. (1988) *Biochim. Biophys. Acta 959*, 178-184.
Lone, Y. C., Simon, M. P., Kahn, A., et al. (1986) *J. Biol. Chem. 261*, 1499-1502.
Maulet, Y., Camp, S., Gibney, G., et al. (1990) *Neuron 4*, 289-301.
Mickel, F. S., Weidenbach, F., Swarovsky, B., et al. (1989) *J. Biol. Chem. 264*, 12895-12901.
Mount, S. M. (1982) *Nucleic Acids Res. 10*, 459-472.
Rudd, E. A., & Brockman, H. L. (1984) in *Lipases* (Borgstrom, B., & Brockman, H. L., Eds.) pp 185-204, Elsevier Science Publishers, Amsterdam.
Sanger, F., Nicklen, S., & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. U.S.A. 80*, 5463-5467.
Sarkar, S. (1984) *Prog. Nucleic Acid Res. Mol. Biol. 31*, 267-293.
Sudhof, T. C., Goldstein, J. L., Brown, M. S., et al. (1985) *Science 228*, 815-822.
Williams, R. J., McCarthy, A. D., & Sutherland, C. D. (1989) *Biochim. Biophys. Acta 1003*, 213-216.

Molecular Cloning, Sequence Analyses, and Expression of Complementary DNA Encoding Murine Progesterone Receptor[†,‡]

David R. Schott,[§] Gopalan Shyamala,[*,§,∥] Wolfgang Schneider,[∥] and Gordon Parry[§]

Division of Cell and Molecular Biology, Lawrence Berkeley Laboratory, University of California, Berkeley, California 94720, and Lady Davis Institute for Medical Research, Sir Mortimer B. Davis Jewish General Hospital, Montreal, Quebec H3T 1E2, Canada

Received February 5, 1991; Revised Manuscript Received April 15, 1991

ABSTRACT: Progesterone receptors exist in two molecular forms commonly designated as "A" and "B" forms, the relative proportion of which can vary among species. In murine tissues, progesterone receptor exists predominantly as the "A" form which, in mammary glands, is also under developmental regulation [Shyamala et al. (1990) *Endocrinology 126*, 2882-2889]. Therefore, toward resolving the molecular mechanisms responsible for the predominance of the "A" form of progesterone receptor in murine tissues and its developmental regulation, we have isolated, sequenced, and expressed the complementary DNA corresponding to the mouse progesterone receptor. Nucleotide sequence analysis revealed two in-frame ATG codons, such that the largest open reading frame beginning with the first codon could encode a polypeptide with an estimated molecular weight of 99 089, while the shorter open reading frame beginning with the second codon could produce a polypeptide with a calculated molecular weight of 81 829. The murine progesterone receptor had complete identity for the DNA binding domain of human and rabbit progesterone receptors and 99% homology with the chicken progesterone receptor; for the steroid binding domain, it had 96% homology with human and rabbit progesterone receptors and 86% homology with chicken progesterone receptors. Expression of the complete complementary DNA in Chinese hamster ovary cells yielded a protein which bound the synthetic progestin promegestone with an equilibrium dissociation constant of approximately 1 nM, and in Western blot analyses revealed both "A" and "B" forms of immunoreactive receptor.

The progesterone receptor (PR) belongs to the superfamily of nuclear steroid receptors which regulate steroid-dependent gene expression by interacting with discrete cis-acting DNA elements present in the 5'-flanking region of target genes [for reviews, see Yamamoto (1985), Evans (1988), Green and Chambon (1988), Beato (1989), and Carson-Jurica et al. (1990)]. Among the various steroid receptors, PR is somewhat unique in that it exists as two forms: a larger molecular form of approximately 108 000-120 000 daltons and a smaller one of approximately 80 000-94 000 daltons (designated as B and A, respectively), the relative proportion of which varies among species [for reviews, see Schrader et al. (1981) and Wei and Horwitz (1986)]. The "A" forms of the receptor repesent an N-terminally truncated naturally occurring variant of the "B" form arising from alternate codon utilization (Conneely et al., 1987a,b; Gronemeyer et al, 1987; Kastner et al., 1990). There

[†] This work was supported by National Institutes of Health Grant HD-22762 and by a grant from the National Cancer Institute of Canada (to G.S.).

[‡] The nucleic acid sequence in this paper has been submitted to GenBank under Accession Number J05333.

* Address correspondence to this author at the Division of Cell and Molecular Biology, Lawrence Berkeley Laboratory, University of California.

[§] University of California.

[∥] Sir Mortimer B. Davis Jewish General Hospital.

0006-2960/91/0430-7014$02.50/0