(2)

AD_____

ISOLATION OF GENOMIC CLONE FOR HUMAN CHOLINESTERASE

ANNUAL SUMMARY REPORT

OKSANA LOCKRIDGE

December 1987

Supported by

U.S. ARMY MEDICAL RESEARCH AND DEVELOPMENT COMMAND
Fort Detrick, Frederick, Maryland 21701-5012

Contract No. DAMD17-86-C-6037

University of Michigan
Department of Pharmacology, Medical Science Building I
Ann Arbor, Michigan 48109-0626

SECURITY CLASSIFICATION OF THIS PAGE

ADA202965

REPORT DOCUMENTATION PAGE

Form Approved OMB No. 0704-0188

1a. REPORT SECURITY CLASSIFICATION: Unclassified	1b. RESTRICTIVE MARKINGS
2a. SECURITY CLASSIFICATION AUTHORITY	3. DISTRIBUTION/AVAILABILITY OF REPORT: Approved for public release; distribution unlimited.
2b. DECLASSIFICATION/DOWNGRADING SCHEDULE	
4. PERFORMING ORGANIZATION REPORT NUMBER(S)	5. MONITORING ORGANIZATION REPORT NUMBER(S)
6a. NAME OF PERFORMING ORGANIZATION: University of Michigan, Pharmaco. Dept., Medical Science I, M6322 / 6b. OFFICE SYMBOL (If applicable)	7a. NAME OF MONITORING ORGANIZATION
6c. ADDRESS (City, State, and ZIP Code): Ann Arbor, Michigan 48109-0626	7b. ADDRESS (City, State, and ZIP Code)
8a. NAME OF FUNDING/SPONSORING ORGANIZATION: U.S. Army Medical Research & Development Command / 8b. OFFICE SYMBOL (If applicable)	9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER: Contract No. DAMD17-86-C-6037
8c. ADDRESS (City, State, and ZIP Code): Fort Detrick, Frederick, Maryland 21701-5012	10. SOURCE OF FUNDING NUMBERS: PROGRAM ELEMENT NO. 62734A; PROJECT NO. 3M16 2734A875; TASK NO. AA; WORK UNIT ACCESSION NO. 523

11. TITLE (Include Security Classification): Isolation of Genomic Clone for Human Cholinesterase

12. PERSONAL AUTHOR(S): Oksana Lockridge, Ph.D.

13a. TYPE OF REPORT	13b. TIME COVERED	14. DATE OF REPORT (Year, Month, Day)	15. PAGE COUNT
Annual Report	FROM 15Nov86 TO 15Nov87	1987 December	43

16. SUPPLEMENTARY NOTATION

17. COSATI CODES			18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number)
FIELD	GROUP	SUB-GROUP	
06	13		RA V; Human cholinesterase gene; butyrylcholinesterase.
06	01		

19. ABSTRACT (Continue on reverse if necessary and identify by block number)

The cDNA for human cholinesterase (EC 3.1.1.8) was isolated from a human basal ganglia library. The clone has a total length of 2426 base pairs (bp), of which 1722 bp code for the mature protein, 212 bp are on the 5' side of the coding sequence, and 492 bp are in the 3' untranslated region. The signal peptide contains 28 amino acids. The amino acid sequence deduced from the cDNA exactly matched the 574 amino acid sequence of the cholinesterase found in human plasma. Therefore, our cDNA represents cholinesterase (EC 3.1.1.8) rather than acetylcholinesterase (EC 3.1.1.7). It was concluded that the amino acid sequence of cholinesterase from two different human tissues, brain and serum, was identical. Hybridization of genomic DNA blots suggested that a single gene, or very few genes, coded for cholinesterase. A genomic clone was isolated using cDNA as a probe. The genomic clone contained exons 1 and 2. There appear to be a minimum of four exons in the cholinesterase gene, and we are searching for other genomic clones containing exons 3 and 4.

20. DISTRIBUTION/AVAILABILITY OF ABSTRACT	21. ABSTRACT SECURITY CLASSIFICATION
☐ UNCLASSIFIED/UNLIMITED ☒ SAME AS RPT. ☐ DTIC USERS	Unclassified

22a. NAME OF RESPONSIBLE INDIVIDUAL	22b. TELEPHONE (Include Area Code)	22c. OFFICE SYMBOL
Mrs. Virginia Miller	301/663-7325	SGRD-RMI-S

DD Form 1473, JUN 86 — Previous editions are obsolete. — SECURITY CLASSIFICATION OF THIS PAGE

AD_____

ISOLATION OF GENOMIC CLONE FOR HUMAN CHOLINESTERASE

ANNUAL SUMMARY REPORT

OKSANA LOCKRIDGE

December 1987

Supported by

U.S. ARMY MEDICAL RESEARCH AND DEVELOPMENT COMMAND
Fort Detrick, Frederick, Maryland 21701-5012

Contract No. DAMD17-86-C-6037

University of Michigan
Department of Pharmacology, Medical Science Building I
Ann Arbor, Michigan 48109-0626

SUMMARY

A human basal ganglia cDNA library was screened with oligonucleotide probes corresponding to portions of the amino acid sequence of human serum cholinesterase (EC 3.1.1.8). Five overlapping clones, representing 2.4 kilobases (kb), were isolated. The sequenced cDNA contained 212 base pairs (bp) of coding sequence 5′ to the N-terminus of the mature protein in which there were four ATG translation start sites in the same reading frame as the protein. Only the ATG coding for Met -28 lay within a favorable consensus sequence for functional initiators. There were 1722 bp of coding sequence corresponding to the soluble, globular, G4 form of cholinesterase. The amino acid sequence deduced from the cDNA exactly matched the 574 amino acid sequence of human serum cholinesterase, which had previously been determined by Edman degradation. Therefore, our clones represented cholinesterase (EC 3.1.1.8) rather than acetylcholinesterase (EC 3.1.1.7). It was concluded that the amino acid sequence of cholinesterase from two different tissues, human brain and human serum, was identical. Expression of a recombinant plasmid in bacterial cells yielded cholinesterase protein but no cholinesterase activity.

Hybridization of genomic DNA blots suggested that a single gene, or very few genes, coded for cholinesterase. Multiple mRNAs of different lengths were found on Northern blot analysis of human liver and brain. A few of these mRNA sizes may be explained by transcription termination at multiple poly(A) sites. One genomic clone was isolated from a human genomic library by screening with cholinesterase cDNA. The clone contained exons 1 and 2. We are presently searching for a genomic clone containing exons 3 and 4. The minimum length of the cholinesterase gene is 17 kb.

FOREWORD

Citations of commercial organizations and trade names in this report do not constitute an official Department of the Army endorsement or approval of the products or services of these organizations.

The investigator(s) have abided by the National Institutes of Health Guidelines for Research Involving Recombinant DNA Molecules (April 1982) and the Administrative Practices Supplements.

For the protection of human subjects, the investigator(s) have adhered to policies of applicable Federal Law 45CFR46.

TABLE OF CONTENTS

INTRODUCTION

In our Annual Report for 1986 (1), we reported isolation and sequencing of one cDNA clone coding for a portion of human cholinesterase. This clone contained a total of 1440 bases, of which 1138 coded for phenylalanine 195 through leucine 574, while 302 bases were in the 3' untranslated region. From amino acid sequencing (2) we knew the cholinesterase subunit contained a total 574 amino acids and that therefore our first clone represented only 66% of the mature cholinesterase protein. In the present report we have isolated overlapping cDNA clones from a human brain cDNA library and have completed sequencing cholinesterase cDNA (3).

Cholinesterase(EC 3.1.1.8) and acetylcholinesterase (EC 3.1.1.7) are generally considered to be products of different genes (4). The evidence includes the observation that there is no immunological cross-reaction between them (5,6). Furthermore, there are known genetic variants of human cholinesterase, including the silent variant with zero activity in serum (7), but no corresponding genetic variants of acetylcholinesterase. The function of cholinesterase is unknown, while the function of acetylcholinesterase is to terminate nerve impulse transmission at cholinergic synapses. Cholinesterase and acetylcholinesterase are highly similar, but not identical, with regard to substrate preference, the types of chemicals that inhibit activity, tissue distribution, and occurrence as globular and asymmetric molecular forms (4). Human cholinesterase and Torpedo acetylcholinesterase are 54% identical in amino acid sequence (2). They have nearly the same number of amino acids per catalytic subunit, 574 and 575 respectively. Their differences in molecular weights are due to differences in glycosylation, acetylcholinesterase having four and cholinesterase having nine carbohydrate chains per subunit (2,8). Their disulfide bonds are located in precisely the same locations, suggesting similar protein folding (9,10). The present report shows the complete nucleotide sequence of human cholinesterase cDNA and the partial sequence of a genomic clone.

This work is of interest to the Army because of its potential application to defense against nerve gas. The cholinesterases are the primary targets of nerve gas. The cholinesterase in serum acts as a scavenger of nerve gas, reacting rapidly to detoxify it. Our present results will be the basis for expression of cholinesterase by recombinant DNA techniques. This may lead to production of a small cholinesterase-like molecule that can detoxify nerve gas.

METHODS

cDNA library screening. A human cDNA library from the basal ganglia of a 1-day-old brain was kindly provided by Dr. Robert A. Lazzarini (11). This same library is now available from the American Type Culture Collection, Rockville, MD. The cDNA was cloned into lambda gt11 via Eco RI linkers. The library contains approximately 1 million independent recombinants. Screening was carried out by the in situ plaque hybridization method of Benton and Davis (12) as modified by Maniatis et al. (13). Table 1 lists the oligonucleotide probes used for finding the first three positive clones, which are called OH57, Z35, and Z3 in Figure 1. Design of the oligonucleotide probes was based on the previously obtained amino acid sequence of human serum cholinesterase (2). Oligonucleotides 5 and 6 were synthesized by the University of Michigan DNA-synthesizing facility. All other oligonucleotides were synthesized by Pharmacia PL-Biochemical Corp., Milwaukee, WI. The oligonucleotides were labeled with gamma-[^{32}P]-ATP using T4 polynucleotide kinase. Prehybridization as well as hybridization was in 0.25% nonfat dry milk (14) dissolved in 6 x SSC (6 x SSC is 0.9 M NaCl and 0.09 M Na Citrate, pH 7.0). Hybridization was for 12-20 h at 36°C for all oligonucleotides except No. 5 which was hybridized at 46°C. Posthybridization washes were at room temperature in 6 x SSC and 0.05% pyrophosphate, followed by a 10-min wash at 41-43°C in 6 x SSC and 0.05% pyrophosphate for probes 3, 4, 7, 8, 9, and 10. More stringent conditions were used for the final wash for probes 5 and 6. Probe 5 was washed for 10 min at 46°C in 2 x SSC and 0.1% sodium dodecyl sulfate (SDS). Probe 6 was washed for 10 min at 46°C in 6 x SSC and 0.05% pyrophosphate.

Additional details of isolation and sequencing of clone Z3 are in our first Annual Report (1). A 130-base pair (bp) fragment of clone Z3 was used as a probe to find the overlapping clone Z2. The 130-bp fragment had been produced by digestion with Sau 3A1. It was subcloned into phage M13 where it was radiolabeled by synthesizing a labeled complement strand. The double-stranded fragment was released from M13 by digestion with Bam HI and Eco RI. It was purified by polyacrylamide gel electrophoresis (PAGE) and electroeluted into a dialysis bag. One hundred thousand plaques were screened by hybridization at 60°C, followed by three brief washes at room temperature in 2 x SSC and 0.1% SDS, and a 10-min wash at 59°C in 0.1 x SSC and 0.2% SDS.

A 270-bp fragment of clone Z2 was used as a probe to find overlapping clone Z13. The 270-bp fragment was produced by digesting clone Z2 with Bam HI and Eco RI. The Eco RI site was in the linker joining human DNA to the lambda arm and was not in the

cholinesterase cDNA. The 270-bp fragment was amplified in plasmid pUC18. The 270-bp fragment was isolated on low melting agarose and radiolabeled by the random oligolabeling method of Feinberg and Vogelstein (15). Hybridization was at 68°C. Posthybridization washes were with 0.1 x SSC and 0.5% SDS. Three brief washes at room temperature were followed by a 120-min wash at 68°C.

DNA sequencing. Sequencing was by the dideoxy method of Sanger et al. (16). Restriction fragments were cloned into M13mp18 and M13mp19. Primers for the sequencing reaction included the universal primer, oligonucleotides 1 and 2 in Table 1, and lambda gt11 primers.

Computer analysis. The Protein Identification Resource, National Biomedical Research Foundation, Georgetown University Medical Center, 3900 Reservoir Road, N.W., Washington, D.C. 20007, contains computer programs which allowed us to store and align our DNA sequence files and to compare the DNA sequences of our clones to those in the computer data bank.

Construction of bacterial expression vector. The plasmid expession vector, pGEM 3ZF-, was purchased from Promega, Madison, WI. We chose this vector because it could be used to achieve two goals: to study expression of cholinesterase in a bacterial host, and to produce messenger RNA for use as riboprobes in Northern blots.

A full-length cholinesterase cDNA was produced by ligating a 0.9-kilobase (kb) Pvu II/Bam HI fragment from the 5′ end of clone Z13 to a 2-kb Bam HI/Eco RI fragment from the 3′ end of clone Z19. Figure 2 is a schematic diagram of clones Z13 and Z19 showing the location of restriction sites used to construct the full-length cDNA. The resulting 2.9-kb fragment was ligated into the Sma I and Eco RI sites of the polylinker region of pGEM 3ZF-. The Pvu II end of the cDNA was attached to the Sma I site of the vector by blunt end ligation. This was possible because both Pvu II and Sma I yield blunt ends. In this forced orientation the 29 amino acids coding for the SP6 promoter, the lac Z peptide, and the polylinker are on the 5′ side of the cholinesterase cDNA. The ligations were carried out simultaneously so that Z13 and Z19 fragments as well as the pGEM vector were in one tube. Transformation was into Escherichia coli strain JM109 (17). Transformants were selected by ampicillin resistance. Colonies containing recombinant plasmid were identified by absence of beta-galactosidase activity which caused colonies to be white rather than blue in the presence of isopropyl-D-thiogalacto pyranoside (IPTG) and 5-bromo-4-chloro-3-indolyl-D-galactoside (X-gal). Seven white colonies resistant to ampicillin were found. Plasmid DNA was prepared from all seven colonies and

subjected to restriction mapping. Only two of the seven colonies had the correct restriction map for cholinesterase. These two recombinant plasmids, containing 2.9-kb of cholinesterase cDNA and 3.2-kb of pGEM DNA, were called pBCHE 1 and pBCHE 2. All further experiments used pBCHE 1 as the two plasmids were identical.

Detection of expression product. pBCHE-transformed cells were grown in 2 ml of LB medium containing 250 ug/ml ampicillin, at 37°C, overnight, in a shaking incubator set at 150 rpm. An aliquot of the overnight culture was diluted to an absorbance at 600 nm of 0.05, using 50 ml of LB medium containing 250 ug/ml ampicillin. The diluted aliquot was grown for approximately 1 h and 15 min to an absorbance at 600 nm of 1.0, after which IPTG was added to a concentration of 0.12 g in 50 ml. Incubation was continued for another 1- 2 h. The IPTG induced the lac promoter to make cholinesterase fusion protein. IPTG had to be added during log phase growth and not earlier because the cholinesterase fusion product was deleterious to growth. Cells were pelleted by centrifugation of the 50-ml culture. Periplasmic proteins were extracted from the cell pellet with 2 ml of 25% sucrose containing 50 mM Tris-Cl, pH 8.0, and 0.5 mg/ml lysozyme. Proteins remaining after another centrifugation were solubilized in 1 ml of 2% SDS, 50 mM Tris-Cl, pH 7.0, and 5% mercaptoethanol.

Cell lysate was tested for the presence of cholinesterase by 1) SDS-PAGE stained for protein with Coomassie blue, 2) SDS-PAGE followed by electrotransfer of proteins to nitrocellulose, washing with cholinesterase rabbit antiserum, washing with horseradish peroxidase-conjugated goat anti-rabbit Ig(H+L) antibody, and staining with hydrogen peroxide and 4-chloro-1-naphthol. The human cholinesterase polyclonal antibody was purchased from Calbiochem, San Diego, CA. The horseradish peroxidase-conjugated goat anti-rabbit Ig(H+L) antibody was from Bio-Rad, Richmond, CA. 3) A third method for detecting cholinesterase was nondenaturing PAGE stained for esterase activity with alpha-naphthyl acetate and Fast Red TR or Fast Blue RR (18).

Riboprobe. Radiolabeled antisense RNA was prepared by cell-free transcription of the recombinant plasmid pBCHE after pBCHE had been linearized with Hind III. Cleavage by Hind III was at the 5' cDNA/plasmid linker junction. RNA synthesis cannot proceed beyond the cleavage point, thus Hind III cleavage defined the end of the transcript. The RNA synthesis reaction in a total volume of 20 ul contained 40 mM Tris-Cl, pH 7.9, 6 mM $MgCl_2$, 2 mM spermidine, 0.5 mM concentration each of rATP, rUTP, and rGTP, 0.012 mM rCTP, 10 mM dithiothreitol, 1 unit/ul of RNasin ribonuclease inhibitor, 50 uCi of alpha-[^{32}P] CTP (Amersham,

Arlington Heights, IL, 410 Ci/mmol), 0.2-1 ug of linearized plasmid DNA, and 10 units of T7 RNA polymerase (Promega, Madison, WI) (19). After 1.5 h reaction at 37°C, the RNA was extracted with phenol/chloroform and precipitated with sodium acetate and ethanol in the presence of carrier RNA.

Northern blot. Poly A(+) RNA from human liver and human brain was purchased from Clontech, Palo Alto, CA. The RNA was denatured with glyoxal and dimethylsulfoxide (13) and electrophoresed on 1.0% agarose. To maintain constant pH, the buffer (12 mM Tris, 6 mM sodium acetate, and 0.3 mM EDTA) was recirculated during electrophoresis. After electrophoresis the gel was blotted onto Gene Screen Plus Membrane (Du Pont NEN, Wilmington, DE). Prehybridization was in 50% formamide, 1% SDS, 1 M NaCl, and 10% dextran sulfate at 37°C for 6 h. Hybridization with 200 uCi of riboprobe in 10 ml of fresh solution was for 48 h at 37°C. Posthybridization washes were in 2 x SSC and 1% SDS at 50°C for 30 min, 0.1 X SSC at 25°C for 30 min, 2 X SSC and 1% SDS at 65°C for 30 min, and 0.1 X SSC at 50°C for 30 min (20).

Screening a human genomic library. A genomic library was kindly provided by Dr. Patrick Venta of the University of Michigan. The library was constructed from white cell DNA purified from the blood of a single donor. A partial Sau 3A digest was ligated into the Bam HI site of the polylinker region of the vector EMBL 3 (21). Human DNA inserts could be removed by digestion with Sal I. The average size of the inserts was 15-20 kb. The library contains 1 million independent recombinants.

The library was plated at a density of 50,000 plaques/ 150-mm plate on 40 plates. Screening was carried out by the *in situ* hybridization method of Benton and Davis as modified by Maniatis et al. (12,13). After baking under vacuum at 80°C for 2 h, the nitrocellulose filters (Millipore, Bedford, MA or Schleicher & Schuell, Keene, NH) were washed at 37°C in 3 X SSC and 0.1% SDS for 6 h and then prehybridized at 68°C in 6 X SSC and 0.25% nonfat dry milk overnight. Filters were hybridized 18 h with 2-4 ng/ml of probe at 68°C in 6 X SSC and 0.25% nonfat dry milk. After hybridization the filters were washed in 2 x SSC and 0.1% SDS at 68°C for 4 h with three wash fluid changes.

Purification of DNA from EMBL 3 genomic clones. The host was *E. coli* strain LE392. A high titer stock of the plaque-purified clone was prepared by incubating 100 ul of competent LE392 cells with 10^6 phage (one plaque). The phage were adsorbed to the LE392 cells for 20 min at 37°C. After adsorption 4 ml of LB broth supplemented with 10 mM $MgSO_4$ and 0.005% thymidine was added. The liquid culture was incubated at 37°C with vigorous shaking until lysis (6-9 h) occurred, at which time the supernatant had a titer of 10^{10} plaque-forming units/ml. Approximately 5 x 10^6

plaques were used to infect 0.3 ml of a fresh overnight culture of LE392. After 20-min incubation to allow adsorption of the phage to the bacterial cells, the inoculum was added to 500 ml of NZYM medium and grown for 9-12 h. NZYM medium contains in a total volume of 1 liter, 10 g of casein (enzymatic hydrolysate from bovine milk, No. C-0626 from Sigma, St. Louis, MO), 5 g of NaCl, 5 g of Bacto-yeast extract (Difco Laboratories, Detroit, MI), and 2 g of $MgSO_4 \cdot 7H_2O$. After the solids have dissolved the pH is adjusted to 7.5 with NaOH (13). DNA was prepared from the culture supernatant by standard procedures (13).

Genomic blots. Human placental DNA (Sigma, St. Louis, MO) was diluted to 50 ug/ml for digestion with various restriction enzymes. After digestion the samples were extracted with phenol and chloroform and then precipitated with ethanol and glycogen (20 ug). The 50 ug of digested DNA was dissolved in 50 ul of 20 mM Tris-Cl, pH 7.6, and 0.1 mM EDTA buffer. Gel electrophoresis was in 300 ml of 1.2% agarose in an 18 x 30 cm gel form, at 40 volts for 24 h. The DNA fragments were transferred to nitrocellulose by the method of Southern as described by Maniatis et al. (13). Filters were hybridized and washed at 68°C as already described.

RESULTS

The oligonucleotides listed in Table 1 were synthesized to use as probes for human cholinesterase cDNA. Our laboratory had previously determined the complete amino acid sequence of human serum cholinesterase by Edman degradation of overlapping peptides (2). The probes correspond to regions of the amino acid sequence representing minimum codon ambiguity. Our most successful oligonucleotide probe proved to be the unique sequence 53mer (probe 5) because it could be used at the highest stringency. The first clone, OH57, was found by using probe 5 alone. The 53mer probe matched the cDNA clone in 50 of 53 nucleotides.

Clone Z3 in Figure 1 was found by screening 150,000 plaques of the human brain cDNA library after making four nitrocellulose lifts from each plate (1). One filter was hybridized with the 53mer probe (No. 5), one with the 39mer probe (No. 6), one with a mixture of probes 4, 8, and 10, and one filter was hybridized with a mixture of probes 3, 7, and 9. Clone Z3 was positive on all four filters. Clone Z35 was found by hybridizing duplicate filters with two different probes, probes 3 and 4. Overlapping clones which completed the coding sequence at the 5′ end were found by using fragments of clones Z3 and Z2 as probes. The 130-bp fragment of Z3 indicated in Figure 1 was used to screen 100,000 plaques which led to the isolation of clone Z2. The 270-bp fragment of Z2 was used to screen 2 million plaques which led to isolation of clone Z13.

The location of one intron was suggested from the sequencing results of clone Z35. This clone had 105 extra nucleotides at its 5′ end in addition to the 614 nucleotides indicated in Figure 1. The extra 105 nucleotides did not match the known amino acid sequence and were absent in clone Z3 as determined by DNA sequencing. The 420 bp of Z35 immediately downstream from these 105 nucleotides matched clone Z3 in all but two nucleotides. The junction between the 105 nucleotides and the rest of clone Z35 contained the sequence TTCCATATTTTACAGG, which is similar to the consensus sequence present at the 3′ splice junction of an intron (22). Thus, it appears that the 105 nucleotides belong to an intron, and that this intron is located between nucleotides 1600 and 1601 in Figure 3. We considered the possibility that these 105 extra nucleotides represented alternative splicing, leading to a cholinesterase with an extra or different peptide near its carboxyl terminus. However, there were five stop codons within the 105 nucleotides, a result compatible with an intron but not with a coding region. It is not uncommon to find unspliced introns or possible cloning artifacts in cDNA clones (23,24). We have not ruled out the interpretation that these 105 nucleotides may be a cloning artifact rather than an intron.

The nucleotide sequence is shown in Figure 3. The cholinesterase found in human serum corresponds to nucleotides 1 through 1722 and contains 574 amino acids. The amino acid sequence derived from the nucleotide sequence of the brain cDNA is a perfect match with the amino acid sequence of cholinesterase found in human serum. This confirms the amino acid sequence determined by Edman degradation, including those regions where an overlap had not been obtained but depended on homology with Torpedo acetylcholinesterase. Of special interest was the finding that asparagine occurred at residues 17, 57, 106, 241, 256, 341, 455, 481, and 486. We had reported that these nine asparagines were glycosylated (2). Since asparagine cannot be detected by amino acid composition analysis due to hydrolysis of the amide by HCl, and since glycosylated asparagine is not detected as a phenylthiohydantoin-amino acid, the identification of asparagine had been based on indirect evidence. The nucleotide sequence shown here supports the conclusion that these residues are asparagines and that they are linked to carbohydrate chains.

Southern blots of human placental DNA that had been digested with various restriction enzymes yielded the results shown in Figure 4. The enzymes and the DNA fragment sizes were Eco RI (2.45 kb), Hind III (4.95 kb), Bam HI (no fragment), Kpn I (no fragment), Pst I (14 kb), Eco RI and Bam HI (1.9 kb), Bam HI and Hind III (3.1 and 1.85 kb), Eco RI and Hind III (2.45 kb), Eco RI and Xba I (1.9 kb), and Eco RI and Pst I (2.45 kb). Other bands and spots that appear in Figure 4 but are not listed above, were irreproducible and therefore are considered to be artifacts. The simple pattern of bands in Figure 4, where only one or two bands in each lane hybridizes, indicates that there may be a single copy of the cholinesterase gene or very few copies. A more complex pattern would be expected for multiple genes because intron sequences would not be conserved even if exons were highly conserved.

Northern blots using 10 ug poly A(+) mRNA/lane from human liver and human brain were hybridized with a 2.8-kb riboprobe. See Figure 5. The result was that the ribosomal RNA gave the most intense bands. Ribosomal RNA is identified by its size of 1.85 and 4.4 kb and is referred to as 18 S and 28 S RNA. In addition to these intense ribosomal RNA bands, there were six faint bands whose sizes were 1.0, 1.45, 2.45, 2.8, 3.4, and 4.2 kb. The banding pattern was exactly the same for human liver mRNA and human brain mRNA. We consider these results tentative because the riboprobe hybridized with the ribosomal RNA and was therefore not completely specific.

Expression in bacteria. Bacterial cells transfected with pBCHE expressed cholinesterase protein as judged by

cross-reaction with human cholinesterase antibodies. The size of the immunoreactive protein was 65,000 on SDS gel electrophoresis. See Figure 6. This size is consistent with a cholinesterase subunit from which the 28 amino acids of the signal peptide and the 29 amino acids of the vector have been cleaved off. The size of 65,000 daltons is also consistent with a subunit free of carbohydrates. The fusion protein would have had a molecular weight of 75,000 per subunit.

The expression product was tested for cholinesterase activity by electrophoresing extracts on nondenaturing polyacrylamide gel and staining for activity with alpha-naphthyl acetate. No cholinesterase activity was detected. Two bacterial esterases were present in all samples, even in JM109 cells containing no recombinant plasmid, and these two esterases migrated much faster than cholinesterase. The yield of cholinesterase protein was estimated to be 3 ug/ 50-ml culture, from gels stained with Coomassie blue.

Isolation of genomic clone. One positive clone, P117N, was found by screening the genomic library with probe OH5~ Figure 7 shows a partial restriction map of clone P117N. Its total length is 19 kb. Clone P117N contains the first two exons of cholinesterase but does not contain exons coding for amino acids 479-574 and the 3' untranslated region. The position of the two exons was determined by restriction mapping and hybridization. To establish the exact length of exon 2 and to confirm that exon 2 was uninterrupted, all of exon 2 was sequenced. The intron/exon junctions for exon 2 were also sequenced. We have not yet completed sequencing the intron/exon junctions of exon 1.

Figure 8 shows the nucleotide sequences at the intron/exon boundaries as we know them to date. Intron 1 is located between nucleotides -93 and -92 and has a length of 4.0 kb. Intron 2 is located between nucleotides 1433 and 1434 and has a length of 8.5 kb. Intron 3 is located between nucleotides 1600 and 1601 and has a length of 2.3 kb. The sizes of introns 2 and 3 are based on restriction mapping of total genomic DNA and are tentative. We expect to get more definitive sizes for these introns once we have isolated a genomic clone containing exons 3 and 4. The intron/exon boundary information for exon 4 is from the unprocessed cDNA clone Z35.

DNA sequencing of clone P117N has added 67 nucleotides to the 5' end of the known cDNA sequence. It is likely that these 67 nucleotides belong to exon 1 because they do not contain the TATA box, which is expected to precede exon 1 by 20-30 nucleotides. Figure 9 shows all of the nucleotide sequence information we have to date for the cholinesterase gene including the partial sequences of introns. Figure 10 shows a tentative map of the

cholinesterase gene and suggests that the total length of the gene is 17 kb or greater.

DISCUSSION

The coding region of the cDNA clone isolated from human brain is a perfect match with the amino acid sequence of human serum cholinesterase (2). Therefore, we conclude that the amino acid sequence of cholinesterase from two different human tissues, from brain and serum, is identical. The cholinesterase present in human serum is thought to be synthesized in the liver (25,26). Prody et al. (27) isolated cholinesterase cDNA from a human liver library, which demonstrates that human liver has mRNA for the same protein found in serum. The cholinesterase in serum is the soluble, globular, tetrameric G4 form. Human fetal and adult brain contain G4 and G1 forms of cholinesterase, with G4 being the most abundant (28,29). It is not known whether G4 and G1 are coded by the same or different mRNAs. Our cDNA represents the soluble G4 form of cholinesterase.

To date the evidence leads us to expect that cholinesterase and acetylcholinesterase will be shown to have different primary amino acid sequences. The amino terminus of acetylcholinesterase from human red blood cells, Glu-Gly-Ala-Glu-Asp-Ala (30), is different from the Glu-Asp-Asp-Ile-Ile-Ile for human serum cholinesterase. The 18 amino acid sequence represented by our 53mer probe is highly conserved in human cholinesterase and in Torpedo and Drosophila acetylcholinesterases (2,8,24,31). This suggests that these same 18 amino acids are likely to be conserved in human acetylcholinesterase. Acetylcholinesterase is more abundant than cholinesterase in human brain, regardless of developmental stage (28,29). For these reasons, we expected to find acetylcholinesterase as well as cholinesterase cDNA using our 53mer oligonucleotide as a probe. However, we have found no acetylcholinesterase candidate clones that differ from the sequence in Figure 3.

Two possible genetic variations were found in clone Z35. Nucleotide 1615 was A in clone Z35, but was G in clone Z3. Nucleotide 1914 was G in Z35, but was A in Z3. Figure 3 shows G at nucleotide 1615, coding for Ala 539, because amino acid sequencing had shown Ala 539. An A at nucleotide 1615 would code for Thr. A rare variant would not have been detected by amino acid sequencing because the cholinesterase protein had been purified from pooled human serum. The cDNA library was from mRNA of a single person. If the donor was heterozygous, the cDNA library would contain an equal number of clones representing the rare and normal variants. The possibility that these nucleotide differences are cloning artifacts must also be considered.

Upstream of the beginning of the mature protein, in the signal peptide region, are 212 nucleotides coding for 69 amino acids. There are four ATG triplets in the same reading frame as

the protein and one ATG triplet in a different reading frame. In 95% of all cases the ATG triplet nearest the 5′ end of the mRNA is the initiator signal (32). This would suggest that ATG at position -207, corresponding to Met -69, could be the codon which initiates translation. However, this ATG does not lie in a consensus sequence for initiator signals. Only the ATG coding for Met -28 lies within a favorable consensus sequence, AATATGC, commonly found in functional initiators (32). Another reason favoring Met -28 as the functional initiator is that signal peptides are known to range in length from 15 to 36 amino acids (33). A signal peptide with 69 amino acids would be unusually long. It is possible that synthesis of cholinesterase is initiated at several sites, and that this has a role in regulation of expression (34). Our signal peptide has the three components characteristic of all signal peptides (33). Residues -1 to -5 are the c region defining the cleavage site for signal peptidase, residues -6 to -17 are the hydrophobic h region, and residues -18 to either Met -28, Met -47, or Met -69 are the n region containing a net positive charge.

The cDNAs for Drosophila and Torpedo marmorata acetylcholinesterase also have 5′ sequences with multiple ATG sites (24,31). In contrast to human cholinesterase cDNA, most of these correpond to very brief coding regions that are terminated by stop codons.

In the 3′ untranslated region, a polyadenylation signal, AATAAA, at nucleotides 2002-2007, is 18 nucleotides away from a poly(A) addition site, CA, at 2025. This agrees with the 10-30 nucleotide separation commonly found for these sites (35). Clone Z3 ends at nucleotide 2027, suggesting that the mRNA for clone Z3 had a poly(A) tail following nucleotide 2026 and suggesting that the poly(A) signal at nucleotide 2002 was functional for clone Z3. Cleavage of pre-mRNA prior to polyadenylation can require the presence of sequences rich in G/T about 30 bp downstream from the CA poly(A) site (35,36). Figure 3 shows that GTG, TTTT, and TGT follow nucleotide 2025 within approximately 30 bases. This gives additional support to the possibility that transcription termination for the mRNA of clone Z3 was directed by the AATAAA signal at nucleotides 2002-2007.

Clone Z35 terminates at the oligo(A) stretch shown in Figure 3, at nucleotide 2214. Clone Z35 may have utilized a nonconsensus poly(A) addition signal, probably ATTAAA (35), at nucleotides 2182-2187. This would suggest termination at alternative poly(A) sites and generation of mRNAs of different lengths but identical coding potential.

Northern blot analysis of mRNA from human liver and human brain showed six sizes for mRNA, 1.0, 1.45, 2.45, 2.8, 3.4, and

4.2 kb. One of these, the 2.45-kb mRNA, fits the cDNA reported in Figure 2. The 1.0- and 1.45-kb mRNA's are too short to code for the 574 amino acids of the cholinesterase subunit and their significance is unclear. The 2.8-, 3.4- and 4.2-kb mRNA's could possibly represent termination at alternative poly(A) sites. Prody et al. (27) reported a single mRNA of 2.5 kb in human liver and human brain. However, multiple mRNA's were found in Torpedo californica and T. marmorata electric organ, as well as in Drosophila. The sizes ranged from 2.2 to 20 kb in T. californica (8) and were 5.5, 10.5, and 14.5 in T. marmorata (24). In Drosophila melanogaster pupae two mRNAs of 4.5 and 4.8 kb were reported, as well as fainter bands both above and below these sizes (31).

The amino acid sequence of human cholinesterase is 54% identical with the amino acid sequence of acetylcholinesterase from the electric organ of T. californica (2,8). The homology extends throughout the two proteins, particularly in the region of amino acids 430-450. The sequences of human cholinesterase and D. melanogaster acetylcholinesterase are 38% identical (10,31). Despite the presence of an active site serine in the cholinesterases, the cholinesterases do not have significant sequence homology with the serine proteases. It appears that the cholinesterases are a specialized type of serine esterase (10), differing from serine proteases, which include the blood coagulation factors and other hydrolytic enzymes. To date the only protein outside of the cholinesterase family which is known to have significant sequence homology with the cholinesterases is thyroglobulin (2,8).

Expression in bacteria. When a recombinant plasmid containing full-length cDNA for cholinesterase was transfected into bacteria, the bacteria synthesized cholinesterase. The size of the bacterial product was 65,000 daltons which is consistent with a cholinesterase subunit free of carbohydrates and processed by removal of the signal peptide. The bacterial product had no cholinesterase activity. This result is not surprising as bacterial systems often produce inactive mammalian enzymes, especially when the mammalian enzyme is a complex glycoprotein with many disulfide bonds. The bacteria seem unable to form correct disulfide bonds. Human cholinesterase has three internal disulfide bonds per subunit and one interchain disulfide bond for a total of 14 disulfide bonds per molecule (10). The importance of the carbohydrate chains is unknown, and a protein made in bacterial cells will not be glycosylated. A mammalian expression system is more likely to yield active cholinesterase enzyme (37).

Cholinesterase gene. Results for the cholinesterase gene are incomplete. We have isolated one genomic clone and this clone

contains exons 1 and 2. Exon 2 is unusually long, containing 1525 nucleotides. Exon 2 codes for 83% of the mature cholinesterase protein. Introns are located between nucleotides -93 and -92 in the 5' untranslated region and between nucleotides 1433 and 1434, and 1600 and 1601 in the coding region. The gene contains a total of at least four exons and three introns. A fifth exon is expected based on results from the laboratory of J. Massoulie for Torpedo acetylcholinesterase (38). They found that the hydrophobic globular form of acetylcholinesterase results from alternative splicing of an exon coding for the carboxyl terminus. The alternatively spliced exon must be a fifth exon. We have no information at present regarding a fifth exon in cholinesterase.

Since cholinesterase and thyroglobulin have extensive amino acid sequence homology, it is of interest to compare the location of introns. Thyroglobulin has 2,748 amino acids (39) which is about 5 times more than the 574 amino acids of the cholinesterase subunit. The sequence homology is only with 540 amino acids at the carboxyl terminus of thyroglobulin and, therefore, the comparison of intron locations is limited to this region. The most striking result is that thyroglobulin has at least nine introns within this region, with one intron occurring approximately every 150 nucleotides (40,41). In contrast, cholinesterase has only two introns within the coding region. These two cholinesterase introns are located in precisely the same positions as the last two introns of thyroglobulin. By comparison with thyroglobulin it seems that seven introns have been lost from the cholinesterase gene. The seven fewer introns in the cholinesterase gene account for the relatively small size of the cholinesterase gene whose minimum size is 17 kb. In contrast, the size of the thyroglobulin gene is 76 kb for the region that corresponds to cholinesterase (40) and 300 kb for the whole thyroglobulin gene (42).

Table 1. Oligodeoxynucleotide probes

Probe	Protein sequence		Probe sequence† (5′ to 3′)
	Amino acids*	Location	
1	WKNQFN	557–562	TGGAARAAYCARTTYAA
2	MMDWKN	554–559	TTYTTCCARTCCATCAT
3	WNNYMMDWKN	550–559	TTITTCCAITCCATCATITAITTITTCCA
4	AEWEWKAGFH	539–548	TGIAAICCIGCITTCCAITCCCAITCIGC
5	WPEWMGVMHGYEIEFVFG	430–447	CCAAAGACAAATTCAATTTCATATCCATGCATGACTCCCATCCATTCAGGCCA
6	EWGNNAFFYYFEH	411–423	ATGTTCAAAATAATAAAAAAAAGCATTATTTCCCCATTC
7	KEFQEG	348–353	AARGARTTYCARGARGG
8	VYGAPGF	331–337	AAICCIGGIGCICCITAIAC
9	GVNKDEGTAF	320–329	AAIGCIGTICCITCITCITTITTIACICC
10	GQFKKTQI	310–317	ATITGIGTITTITTIAAITGICC

*Single letter code for amino acids: A, alanine; D, aspartic acid; E, glutamic acid; F, phenylalanine; G, glycine; H, histidine; I, isoleucine; K, lysine; M, methionine; N, asparagine; P, proline, Q, glutamine; T, threonine; V, valine; W, tryptophan; Y, tyrosine.
†R = A or G; Y = C or T; I, deoxyinosine.

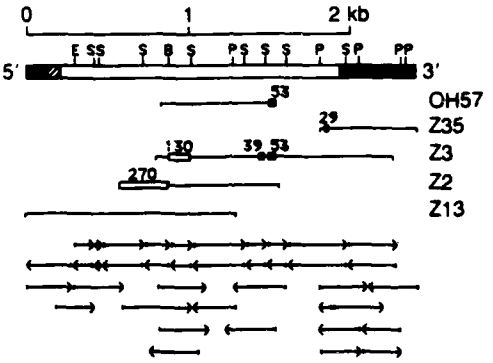


FIG. 1. Restriction map and sequencing strategy for cholinesterase cDNA. Scale at top shows length in kilobases (kb). The open bar represents the 1722 nucleotides that code for the 574 amino acids of mature cholinesterase. The hatched bar represents 84 nucleotides of the signal-peptide region. The solid bar at the 3' end represents 492 nucleotides of the 3' untranslated sequence. Restriction sites used for sequencing are indicated by E (*Eco*RI), S (*Sau*3A1), B (*Bam*HI), and P (*Ssp* I). Lines immediately below the restriction map show the sizes and designations for five cDNA inserts in λgt11. The solid boxes are oligonucleotide probes 53, 29, and 39 nucleotides in length. They are shown on the clones they identified. The open boxes are cDNA probes, 130 and 270 bp in length, prepared from clones Z3 and Z2, respectively. Lines with arrowheads show the direction and extent of DNA sequencing.

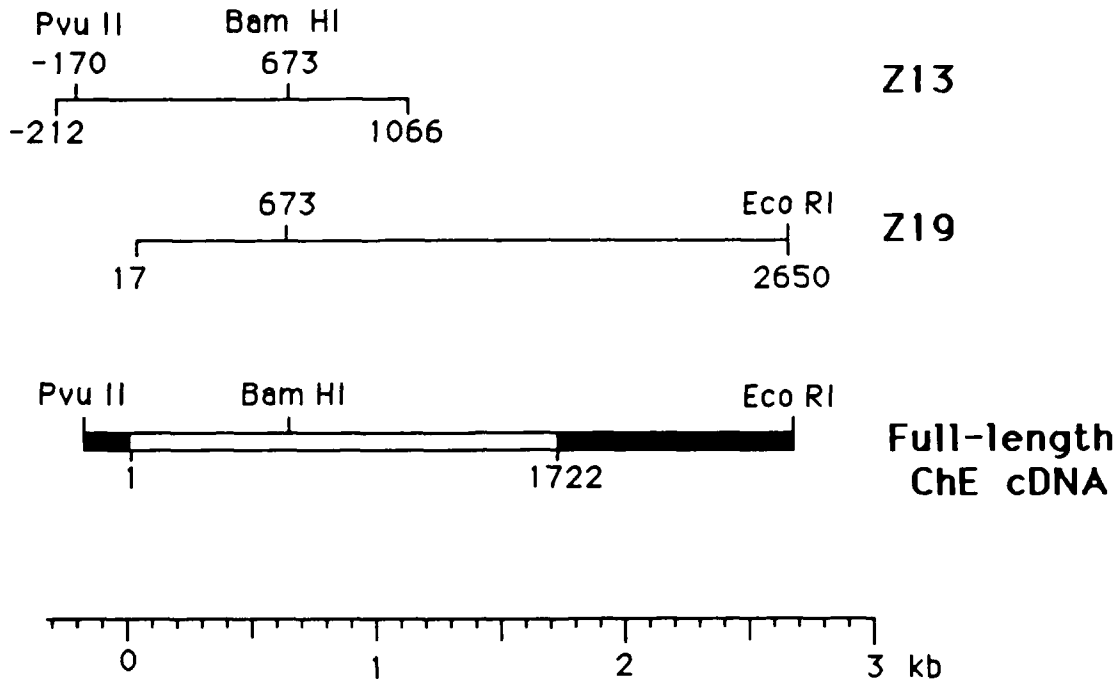


Fig. 2. Schematic diagram of clones used to construct full-length cholinesterase cDNA. The Pvu II/Bam HI fragment from cDNA clone Z13 was ligated to the Bam HI/Eco RI fragment of cDNA clone Z19 to yield full-length cholinesterase cDNA. The Pvu II and Bam HI restriction sites are genuine sites in the cholinesterase gene, but the Eco RI site at nucleotide 2650 is from an Eco RI linker attached during construction of the cDNA library. In the diagram for full-length cDNA the open box from nucleotides 1 to 1722 represents the complete coding region of mature cholinesterase. The filled-in box starting at the Pvu II site is the 5' untranslated region and includes 84 nucleotides of the signal peptide. The filled-in box from nucleotide 1722 to the Eco RI site is the 3' untranslated region.

Fig. 3. Nucleotide and amino acid sequence of human cholinesterase.

```
     *** Met Ser Val Gln Ser Asn Leu Gln Ala Gly Ala Ala Ala Ala Ser Cys Ile Ser  -52
 AC TGA ATG TCA GTG CAG TCC AAT TTA CAG GCT GGA GCA GCA GCT GCA TCC TGC ATT TCC -154

Pro Lys Tyr Tyr Met Ile Phe Thr Pro Cys Lys Leu Tyr His Leu Cys Cys Arg Glu Ser  -32
CCG AAG TAT TAC ATG ATT TTC ACT CCT TGC AAA CTT TAC CAT CTT TGT TGC AGA GAA TCG  -94
                -28
Glu Ile Asn Met His Ser Lys Val Thr Ile Ile Cys Ile Arg Phe Leu Phe Trp Phe Leu  -12
GAA ATC AAT ATG CAT AGC AAA GTC ACA ATC ATA TGC ATC AGA TTT CTC TTT TGG TTT CTT  -34
                                        -1  +1
Leu Leu Cys Met Leu Ile Gly Lys Ser His Thr.Glu Asp Asp Ile Ile Ile Ala Thr Lys    9
TTG CTC TGC ATG CTT ATT GGG AAG TCA CAT ACT.GAA GAT GAC ATC ATA ATT GCA ACA AAG   27

Asn Gly Lys Val Arg Gly Met Asn Leu Thr Val Phe Gly Gly Thr Val Thr Ala Phe Leu   29
AAT GGA AAA GTC AGA GGG ATG AAC TTG ACA GTT TTT GGT GGC ACG GTA ACA GCC TTT CTT   87

Gly Ile Pro Tyr Ala Gln Pro Pro Leu Gly Arg Leu Arg Phe Lys Lys Pro Gln Ser Leu   49
GGA ATT CCC TAT GCA CAG CCA CCT CTT GGT AGA CTT CGA TTC AAA AAG CCA CAG TCT CTG  147

Thr Lys Trp Ser Asp Ile Trp Asn Ala Thr Lys Tyr Ala Asn Ser Cys Cys Gln Asn Ile   69
ACC AAG TGG TCT GAT ATT TGG AAT GCC ACA AAA TAT GCA AAT TCT TGC TGT CAG AAC ATA  207

Asp Gln Ser Phe Pro Gly Phe His Gly Ser Glu Met Trp Asn Pro Asn Thr Asp Leu Ser   89
GAT CAA AGT TTT CCA GGC TTC CAT GGA TCA GAG ATG TGG AAC CCA AAC ACT GAC CTC AGT  267

Glu Asp Cys Leu Tyr Leu Asn Val Trp Ile Pro Ala Pro Lys Pro Lys Asn Ala Thr Val  109
GAA GAC TGT TTA TAT CTA AAT GTA TGG ATT CCA GCA CCT AAA CCA AAA AAT GCC ACT GTA  327

Leu Ile Trp Ile Tyr Gly Gly Gly Phe Gln Thr Gly Thr Ser Ser Leu His Val Tyr Asp  129
TTG ATA TGG ATT TAT GGT GGT GGT TTT CAA ACT GGA ACA TCA TCT TTA CAT GTT TAT GAT  387

Gly Lys Phe Leu Ala Arg Val Glu Arg Val Ile Val Val Ser Met Asn Tyr Arg Val Gly  149
GGC AAG TTT CTG GCT CGG GTT GAA AGA GTT ATT GTA GTG TCA ATG AAC TAT AGG GTG GGT  447

Ala Leu Gly Phe Leu Ala Leu Pro Gly Asn Pro Glu Ala Pro Gly Asn Met Gly Leu Phe  169
GCC CTA GGA TTC TTA GCT TTG CCA GGA AAT CCT GAG GCT CCA GGG AAC ATG GGT TTA TTT  507

Asp Gln Gln Leu Ala Leu Gln Trp Val Gln Lys Asn Ile Ala Ala Phe Gly Gly Asn Pro  189
GAT CAA CAG TTG GCT CTT CAG TGG GTT CAA AAA AAT ATA GCA GCC TTT GGT GGA AAT CCT  567

                    ●●●
Lys Ser Val Thr Leu Phe Gly Glu SER Ala Gly Ala Ala Ser Val Ser Leu His Leu Leu  209
AAA AGT GTA ACT CTC TTT GGA GAA AGT GCA GGA GCA GCT TCA GTT AGC CTG CAT TTG CTT  627

Ser Pro Gly Ser His Ser Leu Phe Thr Arg Ala Ile Leu Gln Ser Gly Ser Phe Asn Ala  229
TCT CCT GGA AGC CAT TCA TTG TTC ACC AGA GCC ATT CTG CAA AGT GGA TCC TTT AAT GCT  687

Pro Trp Ala Val Thr Ser Leu Tyr Glu Ala Arg Asn Arg Thr Leu Asn Leu Ala Lys Leu  249
CCT TGG GCG GTA ACA TCT CTT TAT GAA GCT AGG AAC AGA ACG TTG AAC TTA GCT AAA TTG  747

Thr Gly Cys Ser Arg Glu Asn Glu Thr Glu Ile Ile Lys Cys Leu Arg Asn Lys Asp Pro  269
ACT GGT TGC TCT AGA GAG AAT GAG ACT GAA ATA ATC AAG TGT CTT AGA AAT AAA GAT CCC  807

Gln Glu Ile Leu Leu Asn Glu Ala Phe Val Val Pro Tyr Gly Thr Pro Leu Ser Val Asn  289
CAA GAA ATT CTT CTG AAT GAA GCA TTT GTT GTC CCC TAT GGG ACT CCT TTG TCA GTA AAC  867

Phe Gly Pro Thr Val Asp Gly Asp Phe Leu Thr Asp Met Pro Asp Ile Leu Leu Glu Leu  309
TTT GGT CCG ACC GTG GAT GGT GAT TTT CTC ACT GAC ATG CCA GAC ATA TTA CTT GAA CTT  927

Gly Gln Phe Lys Lys Thr Gln Ile Leu Val Gly Val Asn Lys Asp Glu Gly Thr Ala Phe  329
GGA CAA TTT AAA AAA ACC CAG ATT TTG GTG GGT GTT AAT AAA GAT GAA GGG ACA GCT TTT  987
```

```
Leu Val Tyr Gly Ala Pro Gly Phe Ser Lys Asp Asn Asn Ser Ile Ile Thr Arg Lys Glu  349
TTA GTC TAT GGT GCT CCT GGC TTC AGC AAA GAT AAC AAT AGT ATC ATA ACT AGA AAA GAA 1047

Phe Gln Glu Gly Leu Lys Ile Phe Phe Pro Gly Val Ser Glu Phe Gly Lys Glu Ser Ile  369
TTT CAG GAA GGT TTA AAA ATA TTT TTT CCA GGA GTG AGT GAG TTT GGA AAG GAA TCC ATC 1107

Leu Phe His Tyr Thr Asp Trp Val Asp Asp Gln Arg Pro Glu Asn Tyr Arg Glu Ala Leu  389
CTT TTT CAT TAC ACA GAC TGG GTA GAT GAT CAG AGA CCT GAA AAC TAC CGT GAG GCC TTG 1167

Gly Asp Val Val Gly Asp Tyr Asn Phe Ile Cys Pro Ala Leu Glu Phe Thr Lys Lys Phe  409
GGT GAT GTT GTT GGG GAT TAT AAT TTC ATA TGC CCT GCC TTG GAG TTC ACC AAG AAG TTC 1227

Ser Glu Trp Gly Asn Asn Ala Phe Phe Tyr Tyr Phe Glu His Arg Ser Ser Lys Leu Pro  429
TCA GAA TGG GGA AAT AAT GCC TTT TTC TAC TAT TTT GAA CAC CGA TCC TCC AAA CTT CCG 1287

Trp Pro Glu Trp Met Gly Val Met His Gly Tyr Glu Ile Glu Phe Val Phe Gly Leu Pro  449
TGG CCA GAA TGG ATG GGA GTG ATG CAT GGC TAT GAA ATT GAA TTT GTC TTT GGT TTA CCT 1347

Leu Glu Arg Arg Asp Asn Tyr Thr Lys Ala Glu Glu Ile Leu Ser Arg Ser Ile Val Lys  469
CTG GAA AGA AGA GAT AAT TAC ACA AAA GCC GAG GAA ATT TTG AGT AGA TCC ATA GTG AAA 1407

Arg Trp Ala Asn Phe Ala Lys Tyr Gly Asn Pro Asn Glu Thr Gln Asn Asn Ser Thr Ser  489
CGT TGG GCA AAT TTT GCA AAA TAT GGG AAT CCA AAT GAG ACT CAG AAC AAT AGC ACA AGC 1467

Trp Pro Val Phe Lys Ser Thr Glu Gln Lys Tyr Leu Thr Leu Asn Thr Glu Ser Thr Arg  509
TGG CCT GTC TTC AAA AGC ACT GAA CAA AAA TAT CTA ACC TTG AAT ACA GAG TCA ACA AGA 1527

Ile Met Thr Lys Leu Arg Ala Gln Gln Cys Arg Phe Trp Thr Ser Phe Phe Pro Lys Val  529
ATA ATG ACG AAA CTA CGT GCT CAA CAA TGT CGA TTC TGG ACA TCA TTT TTT CCA AAA GTC 1587

Leu Glu Met Thr Gly Asn Ile Asp Glu Ala Glu Trp Glu Trp Lys Ala Gly Phe His Arg  549
TTG GAA ATG ACA GGA AAT ATT GAT GAA GCA GAA TGG GAG TGG AAA GCA GGA TTC CAT CGC 1647

Trp Asn Asn Tyr Met Met Asp Trp Lys Asn Gln Phe Asn Asp Tyr Thr Ser Lys Lys Glu  569
TGG AAC AAT TAC ATG ATG GAC TGG AAA AAT CAA TTT AAC GAT TAC ACT AGC AAG AAA GAA 1707

Ser Cys Val Gly Leu ***                                                          574
AGT TGT GTG GGT CTC TAA TTAATAGATTTACCCTTTATAGAACATATTTTCCTTTAGATCAAGGCAAAAATAT 1780

CAGGAGCTTTTTTACACACCTACTAAAAAAGTTATTATGTAGCTGAAACAAAAATGCCAGAAGGATAATATTGATTCCT 1859

CACATCTTTAACTTAGTATTTTACCTAGCATTTCAAAACCCAAATGGCTAGAACATGTTTAATTAAATTTCACAATATA 1938

AAGTTCTACAGTTAATTATGTGCATATTAAAACAATGGCCTGGTTCAATTTCTTTCTTTCCTTaataaaTTTAAGTTTT 2017

TTCCCCCCAAAATTATCAGTGCTCTGCTTTTAGTCACGTGTATTTTCATTACCACTCGTAAAAAGGTATCTTTTTTAAA 2096

TGAATTAAATATTGAAACACTGTACACCATAGTTTACAATATTATGTTTCCTAATTAAAATAAGAATTGAATGTCAATA 2175

TGAGATATTAAAATAAGCACAGAAAATCAAAAAAAAAAAA                                         2214
```

Fig. 3. Nucleotide and amino acid sequence of human cholinesterase. Nucleotide 1 is the G of the GAA codon for the amino terminal amino acid (Glu, amino acid 1). The active site serine is Ser-198. The underlined ATG's are possible translation initiation sites. The polyadenylation signal (aataaa) is shown in lowercase letters in the 3' untranslated region. The nucleotide locations of clones used to construct this sequence are as follows: Z13, -212-1066; Z2, 400-1381; Z3, 583-2027; OH57, 625-1341; and Z35, 1600-2214. Nucleotide 1615 (Ala 539) was G in clone Z3 but was A in clone Z35. Nucleotide 1914 was A in clone Z3 but was G in clone Z35.

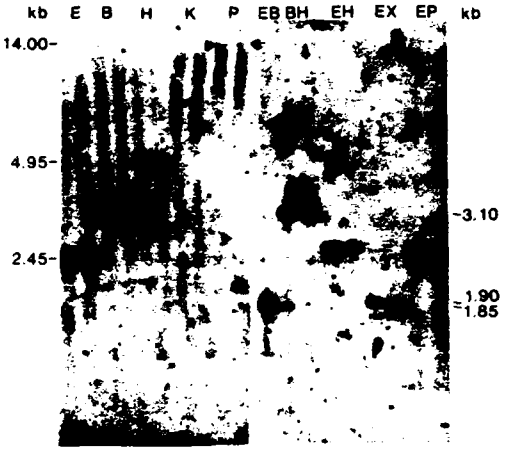


FIG. 4 Hybridization of genomic blots with ^{32}P-oligolabeled cDNA clone OH57. Each lane contained 25 μg of human placental genomic DNA digested with *Eco*RI (E), *Bam*HI (B), *Hin*dIII (H), *Kpn* I (K), *Pst* I (P), or combinations of these (EX, *Eco*RI plus *Xba* I). Electrophoresis in 1.2% agarose was followed by DNA transfer to nitrocellulose filter. Hybridization was at 68°C for 24 hr with probe (4 ng/ml) in 6× SSC/0.25% dry milk. The blots were washed in 2× SSC/0.1% $NaDodSO_4$ at 68°C (3 times, 1 hr each) and exposed for 20 hr with intensifying screen.

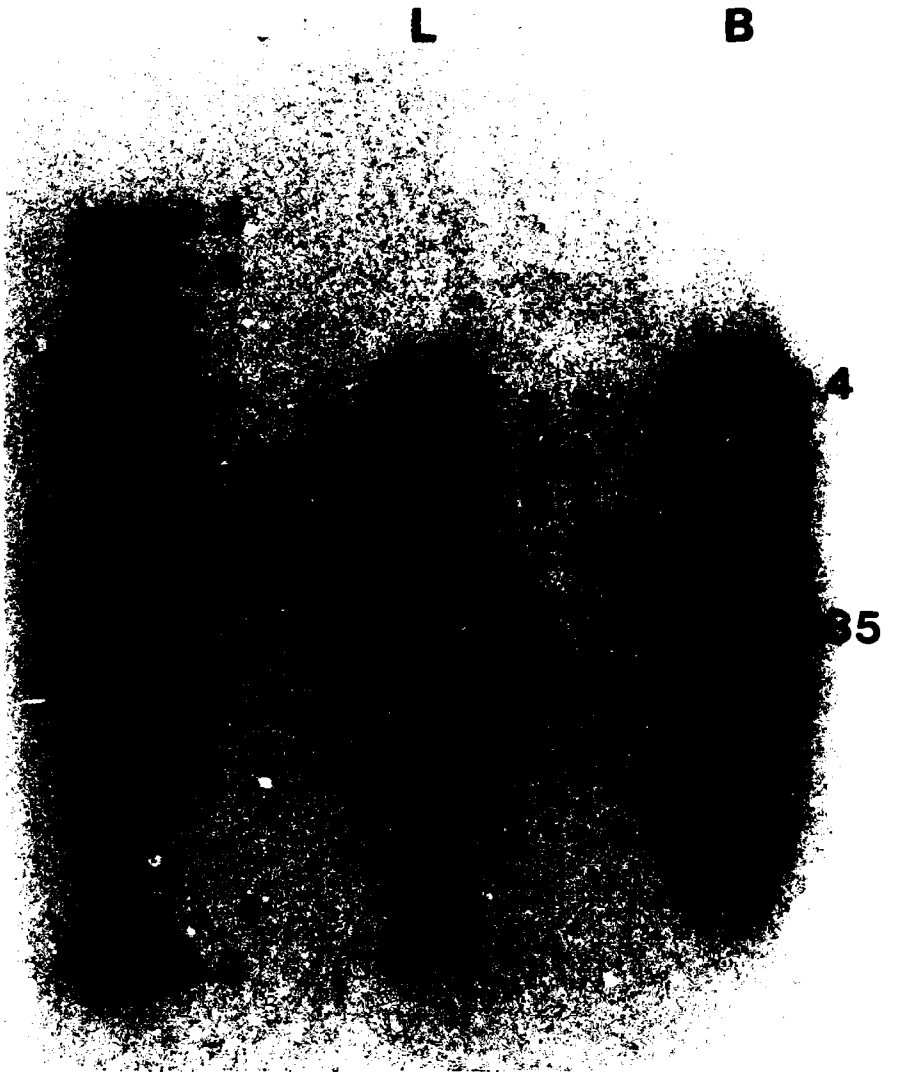


Fig. 5. Northern blot. Poly A(+) RNA, 1 ug in each lane, from human liver (lane L) and human brain (lane B) was applied to a 1.0 % agarose gel containing glyoxal. After transfer to Gene Screen Membrane, the RNA was hybridized with a 2.8 kb riboprobe representing full-length cholinesterase. Faint bands at 4.2, 3.4, 2.8, 2.45, 1.45, and 1.0 kb are likely to be cholinesterase RNA. The same set of faint bands is present in liver and brain RNA. Dark bands at 4.4 and 1.85 kb are the 28 S and 18 S ribosomal RNA. Radiolabeled RNA molecular weight markers are shown in the left lane.

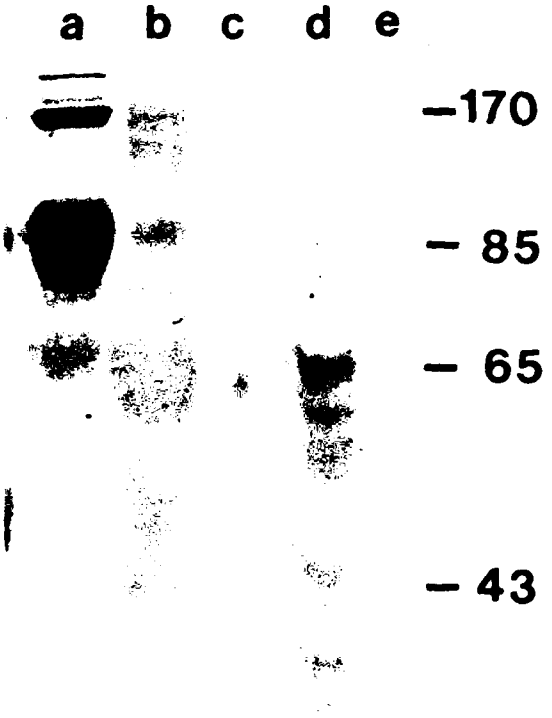


Fig. 6. Detection of cholinesterase expression on Western blot. Samples were first subjected to SDS-PAGE, then transferred to nitrocellulose membrane and hybridized with cholinesterase antibody. Lane a contains highly purified human serum cholinesterase. Lane b contains human albumin. Lanes c, d, and e are bacterial cell lysates. Lane c is the cell lysate from a control experiment in which JM109 cells were transfected with plasmid vector devoid of human DNA. Lane d is the cell lysate from JM109 cells transfected with pBCHE and induced for one h with IPTG. Lane e is the same as d except that no IPTG inducer was added. Cholinesterase antibodies recognized several bands in lane d; the band at 65 kdaltons may be intact cholinesterase protein, while the lower molecular weight bands may be proteolytic degradation fragments.

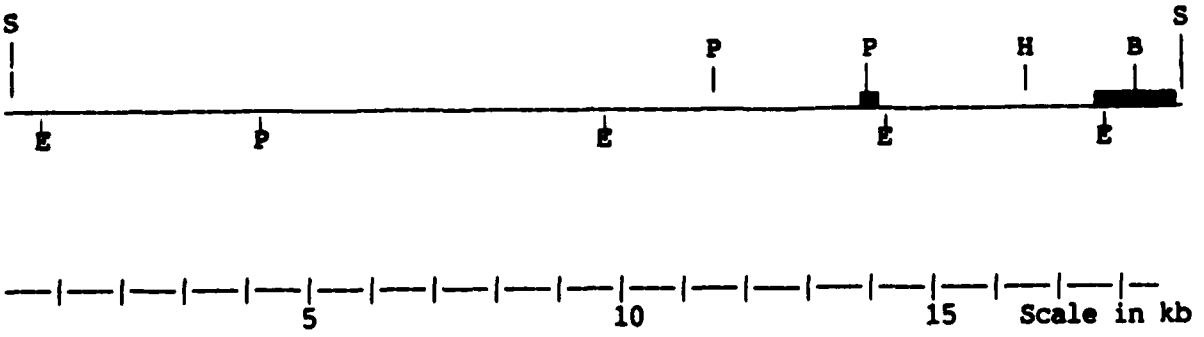
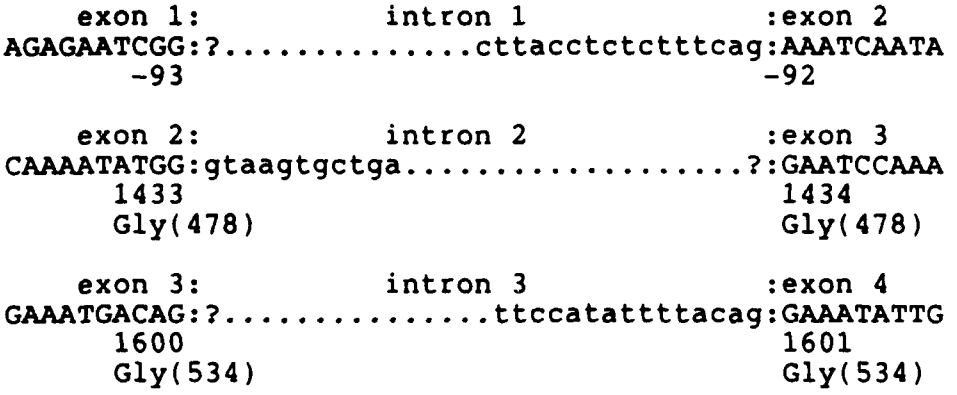
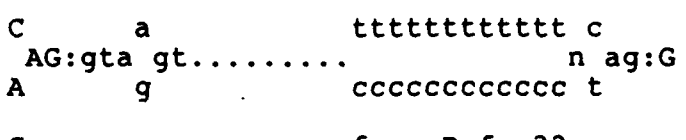


Fig. 7. Restriction map of genomic clone P117N. The heavy bars are exons 1 and 2. Restriction sites are Eco RI (E), Pvu II (P), Hind III (H), Bam HI (B), and Sal I (S). Sal I marks the boundary of the human DNA and the phage vector. There are additional Eco RI, Pvu II, and Hind III sites within the clone that have not yet been mapped.





Consensus sequence from Ref. 22.

Fig. 8. Intron/exon boundaries of cholinesterase. Exon sequences are in capital letters. Intron sequences are in lowercase letters. Numbers below the DNA sequence are the nucleotide numbers. The boundary for intron 3 is from sequencing an unprocessed cDNA clone.

Fig. 9. Partial nucleotide sequence of the cholinesterase gene.

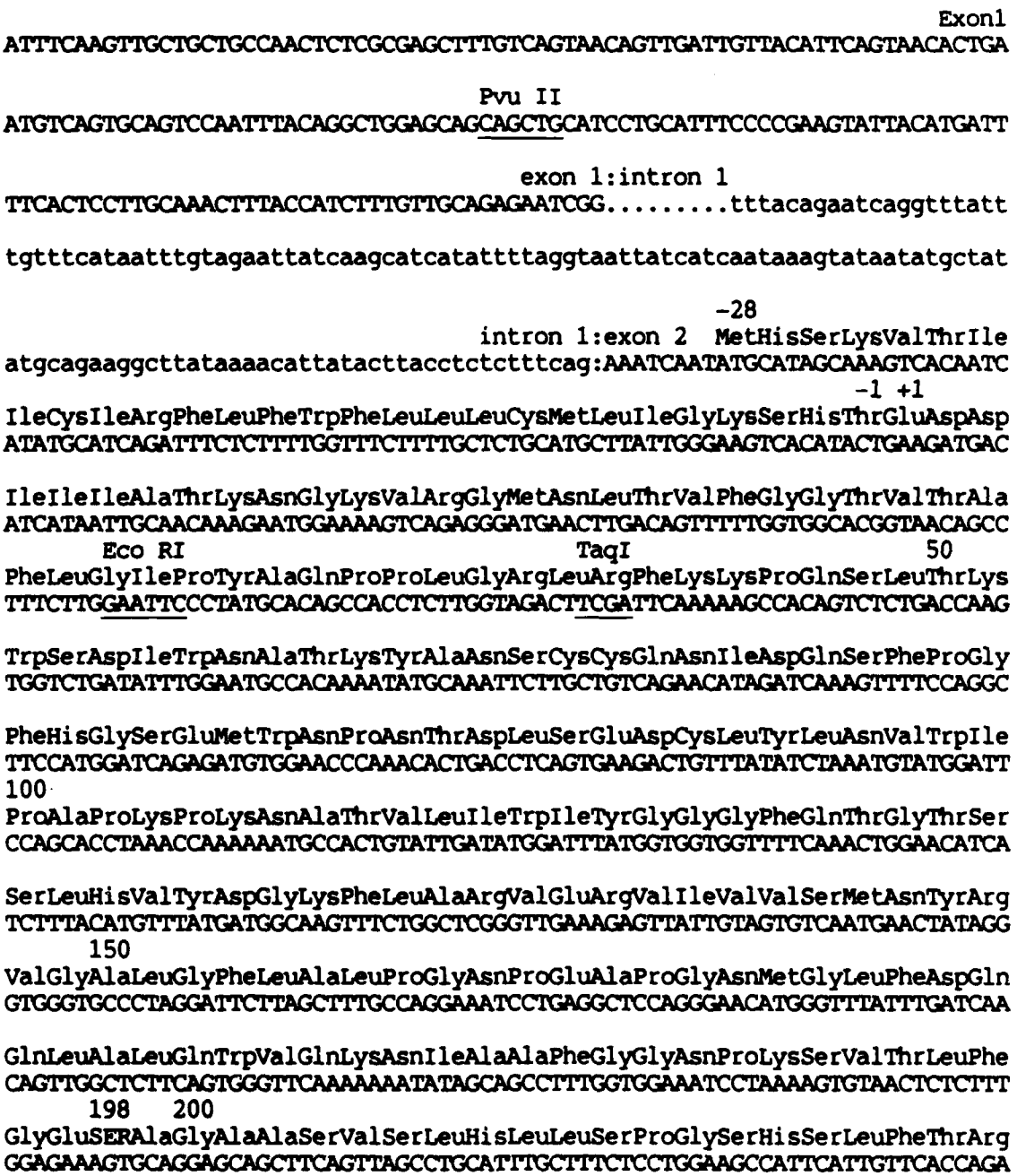
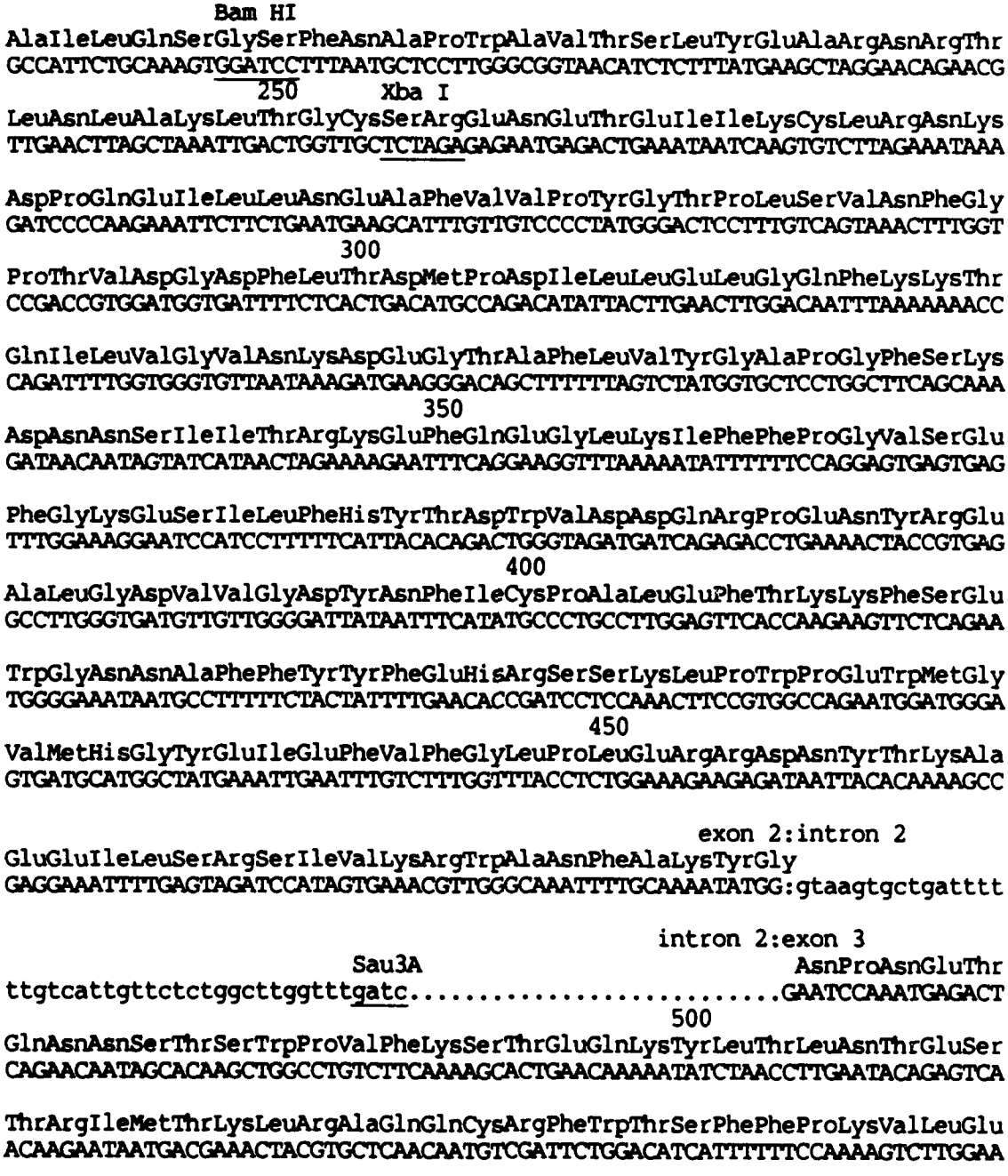


Fig. 9. Partial nucleotide sequence of the cholinesterase gene.



```
   exon 3:intron 3
MetThrGly
ATGACAG.....ctgtactgtgtagttagagaaaatgg

cttttgtattcgaaattatttttcagttaatgaaacagataaaaattttgattaatacaacttattccatat

intron3:exon4                                                    550
         AsnIleAspGluAlaGluTrpGluTrpLysAlaGlyPheHisArgTrpAsnAsnTyrMetMet
tttacagGAAATATTGATGAAGCAGAATGGGAGTGGAAAGCAGGATTCCATCGCTGGAACAATTACATGATG

                                                                 574
AspTrpLysAsnGlnPheAsnAspTyrThrSerLysLysGluSerCysValGlyLeu***
GACTGGAAAAATCAATTTAACGATTACACTAGCAAGAAAGAAAGTTGTGTGGGTCTCTAATTAATAGATTTA

CCCTTTATAGAACATATTTTCCTTTAGATCAAGGCAAAAATATCAGGAGCTTTTTTACACACCTACTAAAAA

AGTTATTATGTAGCTGAAACAAAAATGCCAGAAGGATAATATTGATTCCTCACATCTTTAACTTAGTATTTT

ACCTAGCATTTCAAAACCCAAATGGCTAGAACATGTTTAATTAAATTTCACAATATAAAGTTCTACAGTTAA

TTATGTGCATATTAAAACAATGGCCTGGTTCAATTTCTTTCTTTCCTTaataaaTTTAAGTTTTTTCCCCCC

AAAATTATCAGTGCTCTGCTTTTAGTCACGTGTATTTTCATTACCACTCGTAAAAAGGTATCTTTTTTAAAT

GAATTAAATATTGAAACACTGTACACCATAGTTTACAATATTATGTTTCCTAATTAAAATAAGAATTGAATG

TCAATATGAGATATTAAAATAAGCACAGAAAATCAAAAAAAAAAA
```

Fig. 9. Partial nucleotide sequence of the cholinesterase gene. The first 67 nucleotides were obtained from a genomic clone and not a cDNA clone, so it is unknown whether they belong to exon 1. A total of four exons and three introns is identified above. Intron sequences are in lowercase letters. The mature protein begins at Glu +1 and ends at Leu 574. The signal peptide starts at Met -28. The active site serine is Ser 198. Unique Pvu II, Eco RI, Bam HI, and Xba I restriction sites are underlined, as well as a Taq I and Sau 3A site that were particularly useful for mapping.

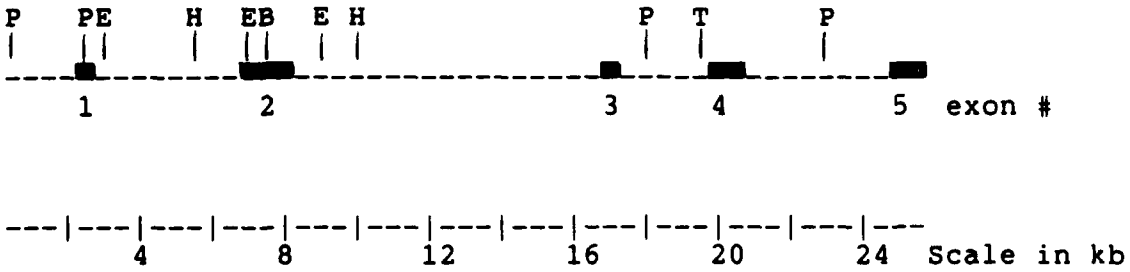


Fig. 10. Tentative restriction map of the cholinesterase gene. The five exons are shown as solid rectangles. The lengths of introns 2 and 3 are tentative. The sizes are: exon 1, 121 bp; intron 1, 4.0 kb; exon 2, 1525 bp; intron 2, 8.5 kb; exon 3, 167 bp; intron 3, 2.3 kb; exon 4, 604 bp. Intron 4 and exon 5 are speculative. Restriction sites are Pvu II (P), Hind III (H), Eco RI (E), Bam HI (B), and Taq I (T).

LITERATURE CITED

1. Lockridge, O. Annual Summary Report. Isolation of genomic clone for human cholinesterase. U.S. Army Medical Research and Development Command, Contract No. DAMD17-86-C-6037, December 1986.

2. Lockridge, O., Bartels, C. F., Vaughan, T. A., Wong, C. K., Norton, S. E. and Johnson, L. L. (1987) J. Biol. Chem. 262, 549-557. Complete amino acid sequence of human serum cholinesterase.

3. McTiernan, C., Adkins, S., Chatonnet, A., Vaughan, T. A., Bartels, C. F., Kott, M., Rosenberry, T. L., La Du, B. N. and Lockridge, O. (1987) Proc. Natl. Acad. Sci. USA 84, 6682-6686. Brain cDNA clone for human cholinesterase.

4. Massoulie, J. and Bon, S. (1982) Annu. Rev. Neurosci. 5, 57-106. The molecular forms of cholinesterase and acetylcholinesterase in vertebrates.

5. Vigny, M., Gisiger, V. and Massoulie, J. (1978) Proc. Natl. Acad. Sci. USA 75, 2588-2592. Nonspecific cholinesterase and acetylcholinesterase in rat tissues: molecular forms, structural and catalytic properties, and significance of the two enzyme systems.

6. Brimijoin, S., Mintz, K. P. and Alley, M. C. (1983) Mol. Pharmacol. 24, 513-520. Production and characterization of separate monoclonal antibodies to human acetylcholinesterase and butyrylcholinesterase.

7. Hodgkin, W. E., Giblett, E. R., Levine, H., Bauer, W. and Motulsky, A. G. (1965) J. Clin. Invest. 44, 486-493. Complete pseudocholinesterase deficiency: genetic and immunologic characterization.

8. Schumacher, M., Camp, S., Maulet, Y., Newton, M., MacPhee-Quigley, K., Taylor, S. S., Friedmann, T. and Taylor, P. (1986) Nature 319, 407-409. Primary structure of Torpedo californica acetylcholinesterase deduced from its cDNA sequence.

9. MacPhee-Quigley, K., Vedvick, T. S., Taylor, P. and Taylor, S. S. (1986) J. Biol. Chem. 261, 13565-13570. Profile of the disulfide bonds in acetylcholinesterase. 10. Lockridge, O., Adkins, S. and La Du, B. N. (1987) J. Biol. Chem. 262, 12945-12952. Location of disulfide bonds within the sequence of human serum cholinesterase.

11. Kamholz, J., DeFerra, F., Puckett, C. and Lazzarini, R. (1986) Proc. Natl. Acad. Sci. USA 83, 4962-4966. Identification of three forms of human myelin basic protein by cDNA cloning.

12. Benton, W. D. and Davis, R. W. (1977) Science 196, 180-182. Screening lambda gt recombinant clones by hybridization to single plaques in situ.

13. Maniatis, T., Fritsch, E. F. and Sambrook, J. (1982) Molecular Cloning, A Laboratory Manual. Cold Spring Harbor, New York.

14. Johnson, D. A., Gautsch, J. W., Sportsman, J. R. and Elder, J. H. (1984) Gene Anal. Techn. 1, 3-8. Improved technique utilizing nonfat dry milk for analysis of proteins and nucleic acids transferred to nitrocellulose.

15. Feinberg, A. P. and Vogelstein, B. (1984) Anal. Biochem. 137, 266-267. Addendum. A technique for radiclabeling DNA restriction endonuclease fragments to high specific activity.

16. Sanger, F., Coulson, A. R., Barrell, B. J., Smith, A. J. H. and Roe, B. A. (1980) J. Mol. Biol. 143, 161-178. Cloning in single-stranded bacteriophage as an aid to rapid DNA sequencing.

17. Yanisch-Perron, C., Vieira, J. and Messing, J. (1985) Gene 33, 103-119. Improved M13 phage cloning vectors and host strains: nucleotide sequences of the M13mp18 and pUC19 vectors.

18. Harris, H., Hopkinson, D. A. and Robson, E. B. (1962) Nature 196, 1296-1298. Two-dimensional electrophoresis of pseudocholinesterase components in normal human serum.

19. Schenborn, E. T. and Mierendorf, R. C. (1985) Nucleic Acids Res. 13, 6223-6236. A novel transcription property of SP6 and T7RNA polymerases: dependence on template structure.

20. Thomas, P. S. (1983) Methods Enzymol. 100, 255-266. Hybridization of denatured RNA transferred or dotted to nitrocellulose paper.

21. Frischauf, A.-M., Lehrach, H., Poustka, A. and Murray, N. (1983) J. Mol. Biol. 170, 827-842. Lambda replacement vectors carrying polylinker sequences. 22. Padgett, R. A., Grabowski, P. J., Konarska, M. M., Seiler, S. and Sharp, P. A. (1986) Annu. Rev. Biochem. 55, 1119-1150. Splicing of messenger RNA precursors.

23. Berger, J., Garattini, E., Hua, J.-C. and Udenfriend, S. (1987) Proc. Natl. Acad. Sci. USA 84, 695-698. Cloning and sequencing of human alkaline phosphatase cDNA.

24. Sikorav, J. L., Krejci, E. and Massoulie, J. (1987) EMBO J. 6, 1865-1873. cDNA sequences of Torpedo marmorata acetylcholinesterase: primary structure of the precursor of a catalytic subunit; existence of multiple 5'-untranslated regions.

25. Brown, S. S., Kalow, W., Pilz, W., Whittaker, M. and Woronick, C. L. (1981) Adv. Clin. Chem. 22, 1-123. The plasma cholinesterases: a new perspective.

26. Khoury, G. F., Brill, J., Walts, L. and Busuttil, R. W. (1987) Anesthesiology 67, 273-274. Atypical serum cholinesterase eliminated by orthotopic liver transplantation.

27. Prody, C. A., Zevin-Sonkin, D., Gnatt, A., Goldberg, O. and Soreq, H. (1987) Proc. Natl. Acad. Sci. USA 84, 3555-3559. Isolation and characterization of full-length cDNA clones coding for cholinesterase from fetal human tissues.

28. Zakut, H., Matzkel, A., Schejter, E., Avni, A. and Soreq, H. (1985) J. Neurochem. 45, 382-389. Polymorphism of acetylcholinesterase in discrete regions of the developing human fetal brain.

29. Atack, J. R., Perry, E. K., Bonham, J. R., Candy, J. M. and Perry, R. H. (1986) J. Neurochem. 47, 263-277. Molecular forms of acetylcholinesterase and butyrylcholinesterase in the aged human central nervous system.

30. Haas, R. and Rosenberry, T. L. (1985) Anal. Biochem. 148, 154-162. Quantitative identification of N-terminal amino acids in proteins by radiolabeled reductive methylation and amino acid analysis: application to human erythrocyte acetylcholinesterase.

31. Hall, L. M. C. and Spierer, P. (1986) EMBO J. 5, 2949-2954. The Ace locus of Drosophila melanogaster: structural gene for acetylcholinesterase with an unusual 5' leader.

32. Kozak, M. (1983) Microbiol. Rev. 47, 1-45. Comparison of initiation of protein synthesis in procaryotes, eucaryotes, and organelles. 33. Von Heijne, G. (1985) J. Mol. Biol. 184, 99-105. Signal sequences. The limits of variation.

34. Leff, S. E., Rosenfeld, M. G. and Evans, R. M. (1986) Annu. Rev. Biochem. 55, 1091-1117. Complex transcriptional

units: diversity in gene expression by alternative RNA processing.

35. Birnstiel, M. L., Busslinger, M. and Strub, K. (1985) Cell 41, 349-359. Transcription termination and 3' processing: the end is in site.

36. McLauchlan, J., Gaffney, D., Whitton, J. L. and Clements, J. B. (1985) Nucleic Acids Res. 13, 1347-1368. The consensus sequence YGTGTTYY located downstream from the AATAAA signal is required for efficient formation of mRNA 3' termini.

37. Powell, J. S., Berkner, K. L., Lebo, R. V. and Adamson, J. W. (1986) Proc. Natl. Acad. Sci. USA 83, 6465-6469. Human erythropoietin gene: high level expression in stably transfected mammalian cells and chromosome localization.

38. Massoulie, J., Sikorav, J-L., Krejci, E., and Reimund, B. (1987, Israel) International workshop on structural and functional aspects of the cholinergic synapse. p. 36. Alternative splicing generates different mRNAs for the catalytic subunits of globular and asymmetric forms of Torpedo acetylcholinesterase.

39. Malthiery, Y. and Lissitzky, S. (1987) Eur. J. Biochem. 165, 491-498. Primary structure of human thyroglobulin deduced from the sequence of its 8448-base complementary DNA.

40. Avvedimento, V. E., Musti, A. M., Obici, S., Cocozza, S. and Di Lauro, R. (1984) Nucleic Acids Res. 12, 3461-3472. Structural organization of the 3' half of the rat thyroglobulin gene.

41. Mori, N., Itoh, N. and Salvaterra, P. M. (1987) Proc. Natl. Acad. Sci. USA 84, 2813-2817. Evolutionary origin of cholinergic macromolecules and thyroglobulin.

42. Baas, F., van Ommen, G-J. B., Bikker, H., Arnberg, A. C. and de Vijlder, J. J. M. (1986) Nucleic Acids Res. 14, 5171-5186. The human thyroglobulin gene is over 300 kb long and contains introns of up to 64 kb.

DISTRIBUTION LIST

1 copy: Commander
US Army Medical Research and Development Command
ATTN: SGRD-RMI-S
Fort Detrick
Frederick, MD 21701-5012

5 copies: Commander
US Army Medical Research and Development Command
ATTN: SGRD-PLE
Fort Detrick
Frederick, MD 21701-5012

2 copies: Administrator
Defense Technical Information Center
ATTN: DTIC-DDA
Cameron Station
Alexandria, VA 22314

1 copy: Commandant
Academy of Health Sciences, US Army
ATTN: AHS-CDM
Fort Sam Houston, TX 78234

1 copy: Dean, School of Medicine
Uniformed Services University of the Health Sciences
4301 Jones Bridge Road
Bethesda, MD 20014