The EMBO Journal vol.6 no.7 pp.1865-1873, 1987

cDNA sequences of *Torpedo marmorata* acetylcholinesterase: primary structure of the precursor of a catalytic subunit; existence of multiple 5'-untranslated regions

Jean-Louis Sikorav, Eric Krejci[1] and Jean Massoulié

Laboratoire de Neurobiologie, Ecole Normale Supérieure, 46, rue d'Ulm, 75230 Paris Cedex 05, France

[1]Present address: Laboratoire de Génétique et de Biologie Cellulaire, Case 907, 13288 Marseille Cedex 9, France

Communicated by J.P.Thiéry

cDNA clones coding for a catalytic subunit of acetylcholinesterase were isolated from cDNA libraries constructed from *Torpedo marmorata* electric organ. The nucleotide sequence of the cloned cDNAs codes for a 599-amino acid precursor containing a 24-amino acid signal peptide. This primary structure has been compared with the sequences of *Torpedo californica* and *Drosophila melanogasta* acetylcholinesterases, and with that of human butyrylcholinesterase. Genomic blot experiments carried out with cDNA restriction fragments used as hybridization probes are in agreement with the existence of a single gene coding for the different catalytic subunits of *Torpedo* acetylcholinesterase. Unexpectedly, we observed multiple 5'-untranslated regions, which may contain several initiation codons.

Key words: acetylcholinesterase/*Torpedo marmorata*/cDNA sequence/multiple 5'-untranslated regions

Introduction

Acetylcholinesterase (AChE, EC 3.1.1.7) is of great physiological importance since it is responsible for the rapid hydrolysis of the neurotransmitter, acetylcholine, at cholinergic synapses. Vertebrates also possess another type of cholinesterase, butyrylcholinesterase (BuChE, EC 3.1.1.8). AChE and BuChE differ in their specificity towards different choline esters, and in their sensitivity to some inhibitors. The catalytic properties of cholinesterases have been reviewed by Rosenberry (1975). These enzymes are serine hydrolases. Their active site presents an esteratic subsite in which the active serine is assumed to be engaged in a charge-relay system similar to that described in serine proteases, and an anionic subsite, which binds quaternary ammonium residues. Labeling experiments of *Electrophorus* AChE with the photoactivatable reagent DDF (Kieffer *et al.*, 1986) allowed the identification of a peptide corresponding to this anionic subsite, Gly-Ser-X-Phe (X = undetermined residue labeled by DDF).

Both AChE and BuChE present a number of molecular forms, which may be classified as asymmetric (A) and globular (G) forms (Massoulié and Bon, 1982; Rosenberry, 1985). The asymmetric forms are complex molecules in which tetrameric assemblies of catalytic subunits are attached by disulfide bonds to the strands of a triple-helical collagen-like 'tail'. Amphiphilic globular forms have been shown to possess a C-terminal-linked glycolipid (for a review, see Silman and Futerman, 1987).

Torpedo electric organs constitute a most favourable material to study the structure of both asymmetric and globular amphiphilic AChE. The catalytic subunits of the two types of forms are equivalent in their catalytic activity and in their immunological reactivity. Their primary structures show few differences (Doctor *et al.*, 1983). They possess identical N-terminal sequences, but a different mol. wt (MacPhee-Quigley *et al.*, 1985; Bon *et al.*, 1986).

The diversity of AChE forms may be generated at different levels: multiplicity of genes, post-transcriptional modification of mRNA precursors and post-translational processes. In order to elucidate the respective roles of these mechanisms, we have undertaken an analysis of *Torpedo* AChE at the genetic level. We first identified a cDNA clone, λAChE1, which carries a 160-bp insert coding for a fragment of AChE, by immunological characterization of the recombinant protein (Sikorav *et al.*, 1985). Using this cDNA probe, we demonstrated the existence of at least three mRNAs for *T. marmorata* AChE (5.5 kb, 10.5 kb, 14.5 kb), in electric organs and electric lobes. These results are consistent with the observation that an electric organ mRNA preparation directed *in vitro* the synthesis of several polypeptide precursors of AChE catalytic subunits (Sikorav *et al.*, 1984). Similar observations were made for *T. californica* (Schumacher *et al.*, 1986).

The primary structure of a catalytic subunit of *T. californica* AChE has been deduced from a cDNA sequence (Schumacher *et al.*, 1986). We report the primary structure of the precursor of a catalytic subunit of *T. marmorata* AChE. We show that AChE mRNAs probably derive from a single gene, and possess multiple 5'-untranslated regions. This multiplicity is not *a priori* related to the polymorphism of the native enzyme, and a possible correlation with mRNAs of different sizes remains to be explored.

Results

The insert of λAChE1 (Sikorav *et al.*, 1985) was used as a probe to isolate the clone pAChE2, which contains a 2.5-kb insert. The sequence of pAChE2 was determined as indicated in Figure 1, and found to contain a single large open reading frame (nucleotides 482-2122). This open reading frame contains the sequence of the λAChE1 insert, as well as the active site serine. The pAChE2 clone, however, lacked the N-terminal sequence of *Torpedo* AChE (MacPhee-Quigley *et al.*, 1985; Bon *et al.*, 1986). A comparison with the cDNA sequence of *T. californica* AChE indicates an extensive homology from nucleotides 490-2502 of pAChE2. No significant homology was found upstream of nucleotide 490.

This suggested that the 5' region of the insert did not correspond to the coding sequence. We therefore performed an S1 mapping experiment, which showed that the totality of the pAChE2 insert is not represented in mRNAs (Figure 2). This does not preclude the possibility that the non-protected region might represent an intronic sequence, absent from mature mRNAs. The protected region corresponded only to the region of homology with *T. californica* AChE cDNA.

This was confirmed by primer extension of an oligonucleotide, ON2, complementary to nucleotides 490-510 of pAChE2. We

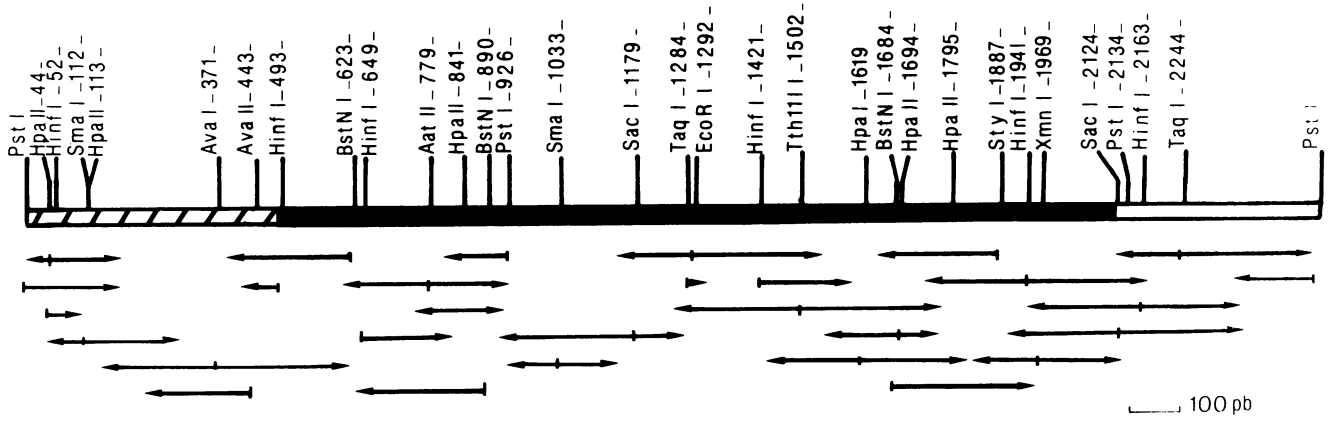

Fig. 1. Sequencing strategy for the cDNA insert of pAChE2. The direction and extent of sequence determinations are shown by horizontal arrows under each clone used. The poly(dG)−poly(dC) tails generated by the cDNA cloning procedure are not included in the restriction map. The open reading frame is shown in black. The open bar represents the 3′-untranslated region, and the stippled bar represents the region which is not protected in S1 mapping (see text). The insert of AChE1, subcloned in the *Eco*RI site of pUC8, was sequenced from both ends. It corresponds to nucleotides 1130−1288 of the pAChE2 insert.

thus determined a 150 nucleotide long sequence (not shown), corresponding to the expected amino acid N-terminal sequence of *T. marmorata* AChE (Bon *et al.*, 1986).

In order to determine the cDNA sequence of the N-terminal region, we analyzed several clones obtained from a restricted library (see Materials and methods). We thus reconstructed one type of mRNA, including the complete coding sequence (Figure 3). The sequence contains the expected N-terminal amino acid sequence of a mature catalytic subunit of *T. marmorata* AChE (Bon *et al.*, 1986), including the two differences oberved with *T. californica*, at positions 3 and 20 (Bon *et al.*, 1986, and Figure 4). Three in-frame termination codons are found upstream of the open reading frame (at positions −96/−94, −93/−91 and −84/−82). Assuming that translation starts at the first initiation codon of the main open reading frame, we find a 24 amino acid long hydrophobic sequence preceding the mature sequence. There are three other initiation codons in the 5′-untranslated region (at positions −132/−130, −93/−91 and −56/−54), which initiate open reading frames containing respectively 12, 22 and nine amino acids (cf. Figure 7).

Genomic blots, evidence for a single AChE gene

Total genomic DNA was digested wtih *Pst*I, *Eco*RI, *Bam*HI and *Sac*I. The fragments were used in Southern blot experiments, hybridizing with a *Pst*I−*Pst*I restriction fragment (nucleotides 926−2134 of pAChE2). Single restriction fragments are detected in the case of *Pst*I, *Bam*HI and *Sac*I (Figure 5, lanes 1, 3 and 4) and three restriction fragments in the case of *Eco*RI (Figure 5, lane 2). The existence of several *Eco*RI restriction fragments is consistent with the fact that the *Pst*I−*Pst*I probe contains an internal *Eco*RI site (at position 1292 of pAChE2). A smaller probe, which also contained the *Eco*RI site (an *Ava*I−*Ava*I restriction fragment, nucleotides 1033−1440 of pAChE2) hybridizes to only two of these three bands (not shown). This suggests the existence of another *Eco*RI site within the gene, probably in an intron.

Analysis of the 5′-untranslated region

The primer extension analyses show the existence of multiple extended products, some of which appear as doublets (Figure 6A). Their sizes are ~290, 330 and 430 nucleotides (for bands 1a and 1b, 2a and 2b and 3, respectively). They may either correspond to mRNAs differing in the length of their 5′-untranslated regions, or to premature terminations of elongation.

S1 mapping analysis in fact established the existence of multiple 5′-untranslated regions (Figure 6B). The cDNA part of the probe, which contained a 5′-untranslated region and the N-terminal part of the coding sequence, was protected in its totality, yielding a 362 nucleotide long fragment, together with two shorter fragments, 200 and 237 nucleotides long.

These experiments clearly imply the existence of at least three types of 5′-untranslated regions. We have indeed isolated one clone (λ5′AChE11), which lacks the nucleotide sequence comprised between the two extremities indicated by the S1 mapping experiment (Figure 7). Yet another type of 5′-untranslated region was observed in *T. californica* (Figure 7, Schumacher, 1987).

Discussion

Different approaches have been used to isolate AChE genes: a purely genetic approach allowed the characterization of the *Drosophila* AChE gene (Hall and Spierer, 1986); Taylor and his collaborators used oligonucleotide probes derived from peptidic sequences (Schumacher *et al.*, 1986) to clone *T. californica* AChE. We have used antibodies to identify a cDNA clone from an expression library (Sikorav *et al.*, 1985). This clone, which encodes amino acids 245−298 of the mature AChE catalytic subunit enabled us to obtain the complete structure of a precursor of a catalytic subunit of AChE by the successive screening of two cDNA libraries.

Interpretation of the structure of pAChE2

In a first screening, we isolated a cDNA clone, pAChE2, whose 5′ end presents an unexpected structure, lacking the N-terminal sequence of AChE, deduced from protein sequencing (MacPhee-Quigley *et al.*, 1985; Bon *et al.*, 1986). The 490-nucleotide sequence which precedes the coding region has the following properties: (i) it is not protected in S1 mapping experiments; (ii) it contains a 160 nucleotide long direct repeat (nucleotides 58−215, 84% homologous to nucleotides 315−468); (iii) a probe derived from this region (restriction fragment *Ava*I−*Ava*I, nucleotides 112−371) hybridizes to repeated sequences of *Torpedo* genomic DNA, and these sequences are expressed in *Torpedo* liver, electric organ and electric lobe, according to Northern blot analysis (data not shown).

Interestingly, the beginning of the coding region of pAChE2 (amino acid 32 of mature AChE) corresponds to the beginning of the homology which is observed between AChE and the C-

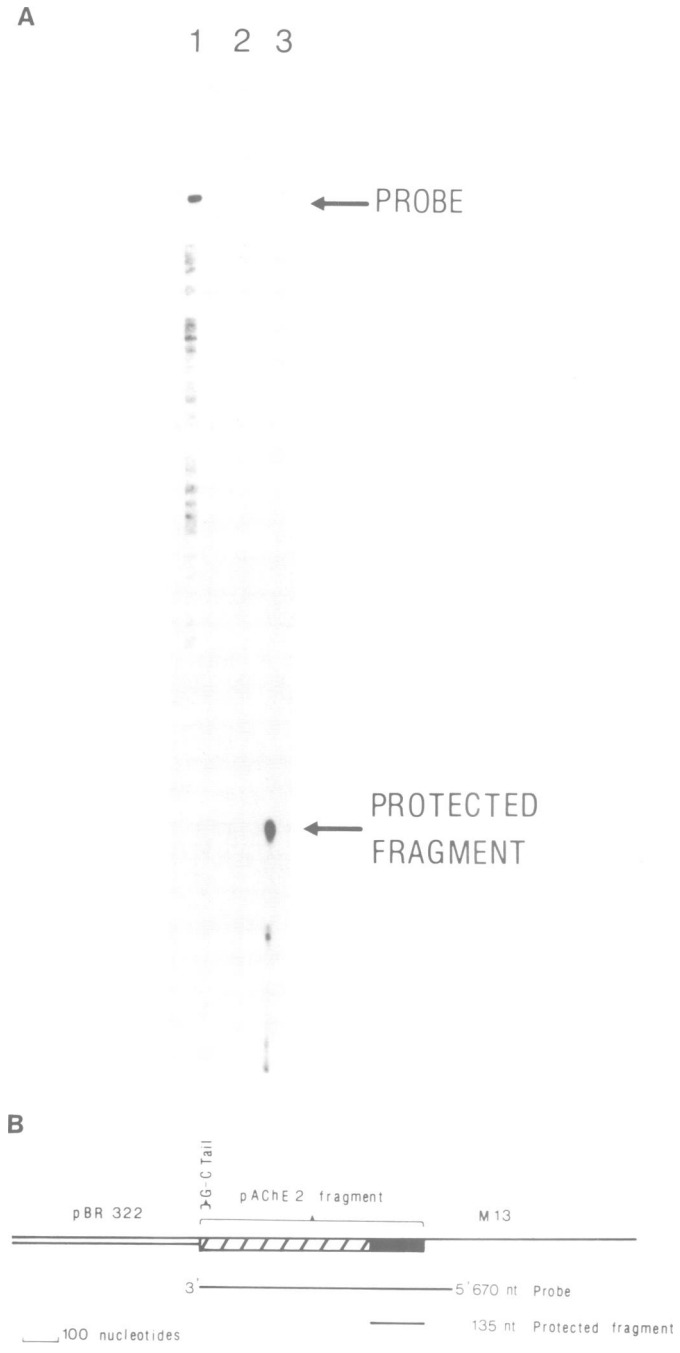

Fig. 2. S1 mapping nuclease analysis of pAChE2 5′ region, mapped with a uniformly labeled single-stranded probe. The plasmid pAChE2 was digested with the restriction enzyme *Bst*NI, and the extremities of the restriction fragments were filled by the Klenow fragment of DNA polymerase I. A 1600 nucleotide long fragment (coordinate 2636 of pBR322 to nucleotide 623 of pAChE2) was isolated and cloned at the *Sma*I site of M13mp10. A phage containing the insert in the relevant orientation was characterized by dideoxynucleotide sequencing, and used to synthesize a complementary probe with an M13 universal primer. A probe corresponding to a naturally occurring termination of the polymerase (in the GC pairs flanking the cDNA insert) was isolated and used for S1 mapping (**A**) Electric organ poly(A)$^+$ RNA (5 μg) (**lane 3**) and wheat germ RNA (**lane 2**) were hybridized to 2 × 10^5 c.p.m. of the probe, then digested with 50 units of S1 nuclease. The arrows indicate the position of the intact probe and of the protected fragment. **Lane 1** undigested probe, 10^4 c.p.m. The mol. wt standards used were a sequencing ladder, and an end-labeled 1-kb ladder (BRL) (not shown). (**B**) Structure of the probe and of the protected fragment. The full bar represents the protein coding region of pAChE2; the stippled bar represents the region which is not protected in S1 mapping. The 5′−3′ polarity of the probe is indicated.

terminal domain of thyroglobulin (Schumacher *et al.*, 1986; Swillens *et al.*, 1986). Since thyroglobulin has probably evolved by gene fusion of an ancestral protein with a portion of AChE, this position is likely to correspond to the limit of two exons.

Taken together, these observations favour the hypothesis of an intronic nature of the 5′ end of pAChE2. The putative intronic sequence is terminated by CG, instead of the usual consensus termination AG. An example of an intron ending in the same manner has however been recently reported for one of the transcripts of *Drosophila Per* gene (Citri *et al.*, 1987). An analysis of the corresponding genomic structure will be necessary to elucidate this question.

A single gene codes for the multiple molecular forms of Torpedo AChE

The multiplicity of AChE catalytic subunits, corresponding to asymmetric and globular forms, and the multiplicity of mRNAs, identified by AChE cDNA probes, raise the question of the possible existence of multiple genes for the enzyme. To investigate this question, we performed Southern blot analyses with a cDNA probe containing a large portion of AChE coding sequence (400 amino acids, including the sequence surrounding the active site serine). If the different subunits were derived from distinct genes, these genes would be expected to hybridize equally well with the probe, since the primary sequences are largely identical (particularly around the active site serine). The observed pattern is in fact consistent with the existence of a single gene (Figure 5). In our high stringency conditions, we avoided the detection of related genes, such as that of butyrylcholinesterase (Toutant *et al.*, 1985).

Comparison of T. marmorata and T. californica AChEs

The sequence determined for *T. marmorata* codes for a 599 amino acid long precursor, containing a 24 amino acid long signal peptide, i.e. three amino acids longer than the signal peptide of *T. californica* AChE.

The amino acid sequences obtained here for the mature catalytic subunit of *T. marmorata* AChE, and that previously determined for *T. californica* (Schumacher *et al.*, 1986) are identical, except at 13 positions (Figure 4). This strong homology implies that these two sequences correspond to the same type of catalytic subunit, probably that of the asymmetric forms, because of the presence of the C-terminal sequence determined for this structure (Schumacher *et al.*, 1986).

Comparison of Torpedo AChE with structurally related proteins

A comparison of *Torpedo* AChE with *Drosophila* AChE and human BuChE leads to several observations.

(i) All the cysteine residues involved in internal disulfide bridges in *Torpedo* (MacPhee-Quigley *et al.*, 1986) are conserved. The cysteine residue which is located near the end of *Torpedo* AChE, and is involved in interchain disulfide linkage, is absent in *Drosophila* AChE.

(ii) The C-terminal portion of *Drosophila* AChE diverges from the other AChE sequences known so far. Interestingly, the beginning of the divergence approximately coincides with the position of the last intron in the coding sequence of *Drosophila* AChE (P.Fournier, personal communication), and of thyroglobulin (Avvedimento *et al.*, 1984). The divergent sequence (underlined in Figure 4) is thus encoded by a separate exon. It is largely hydrophobic (Hall and Spierer, 1986), and it is possibly exchanged in part for a glycolipid anchor (Silman and Futerman, 1987; Cross, 1987), as suggested by the fact that *Drosophila* AChE is a dimeric, phosphatidylinositol-anchored enzyme (Gnagey *et al.*, 1987).

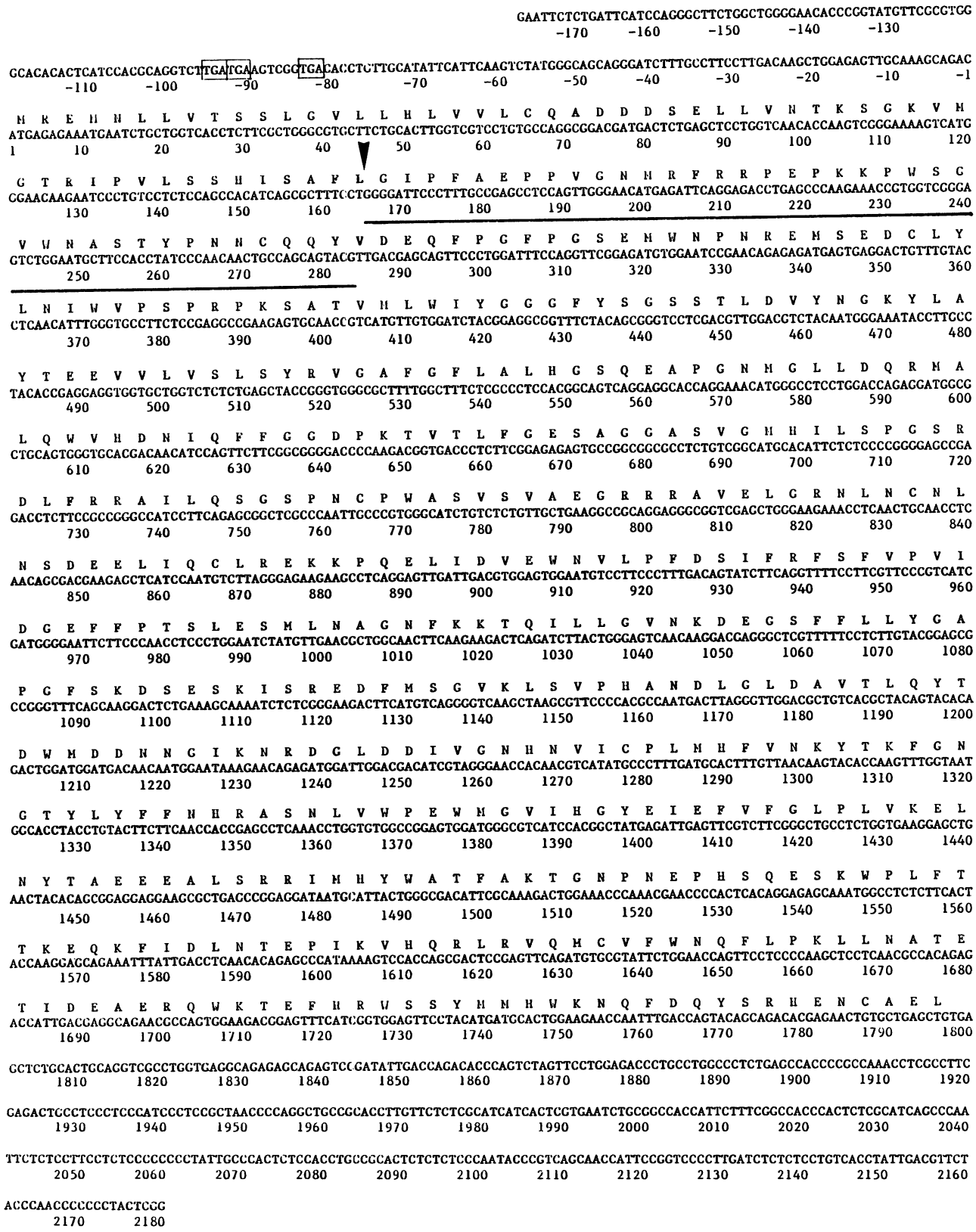

Fig. 3. Primary structure of an AChE mRNA, deduced from pAChE2 and cDNAs isolated from a restricted library. The nucleotide sequence is numbered in the 5′ to 3′ direction, beginning at the first nucleotide of the ATG initation codon. The predicted amino acid sequence is shown above the corresponding nucleotide sequence. The *Eco*RI fragments of six independent cDNA clones, isolated from the λgt10 library (λ5′AChE3, λ5′AChE4, λ5′AChE7, λ5′AChE8, λ5′AChE10 and λ5′AChE12) were subcloned in both orientations in M13mp18 and sequenced using an M13 universal primer or the oligonucleotide ON2, when in the relevant orientation. A common 460-nucleotide sequence was established and combined with that of pAChE2 to establish the structure of an mRNA coding for AChE. From nucleotide 166 to the end, the sequence corresponds to the 3′ part of pAChE2. Underlining indicates sequences common to pAChE2 and clones from the restricted DNA library (166−284). The arrow indicates the 5′ extremity of the *bona fide* coding region of pAChE2, corresponding to nucleotide 491 of the pAChE2 insert. In-frame termination codons preceding the open reading frame are boxed.

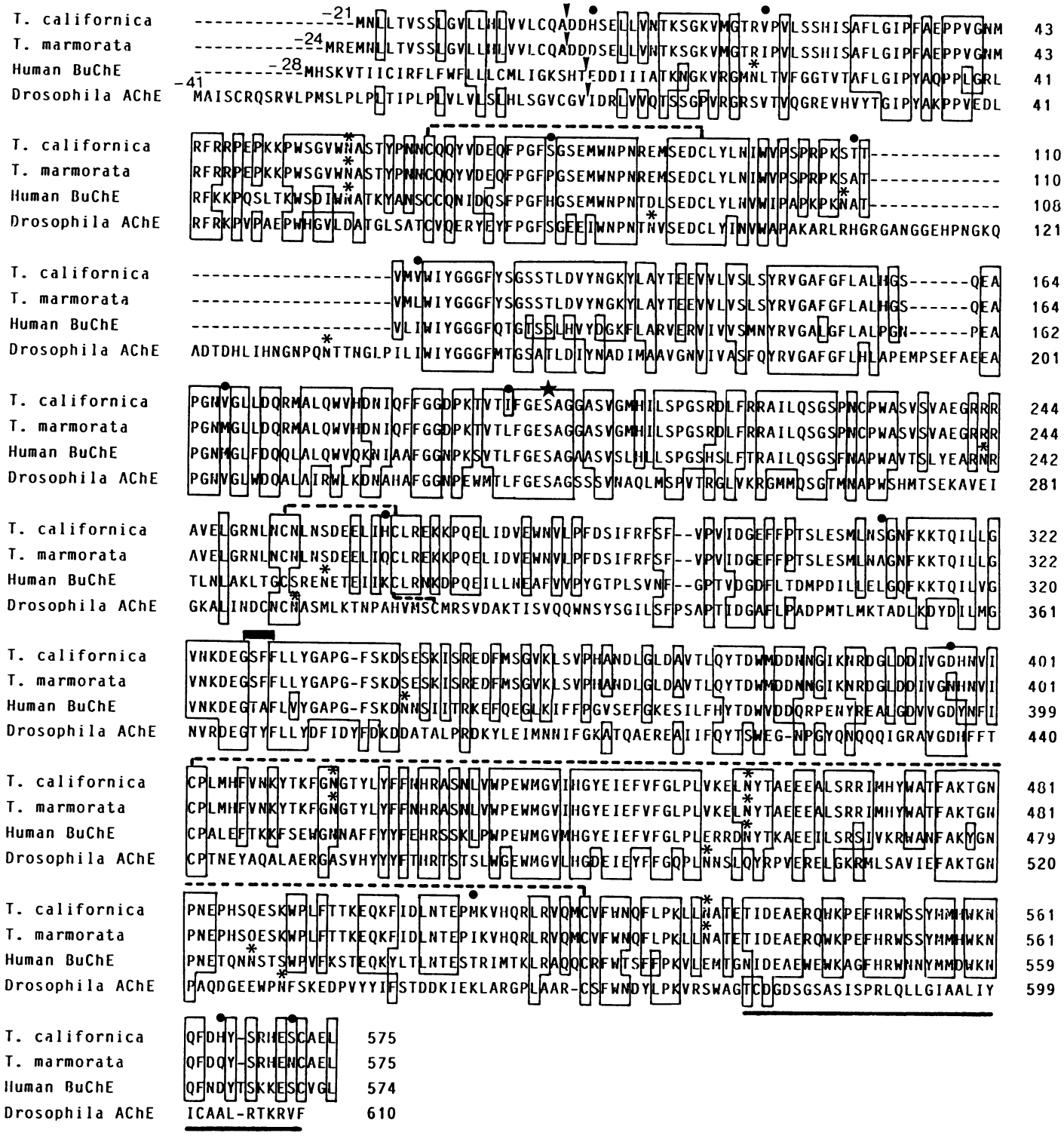

Fig. 4. Comparison of the primary structures of cholinesterases. The amino acid sequences of AChE from *T. californica* (Schumacher *et al.*, 1986; Schumacher, 1987). *T. marmorata* and *Drosophila melanogaster* (Hall and Spierer, 1986) were deduced from cDNA sequences. The structure of human serum BuChE was determined chemically (Lockridge *et al.*, 1987) and from cDNA sequences (Prody *et al.*, 1987). The arrows indicate the N-terminal amino acids of the mature proteins, from N-terminal sequences determined in the case of *T. californica* (MacPhee-Quigley *et al.*, 1986), *T. marmorata* (Bon *et al.*, 1986), and *Drosophila* (T.Rosenberry, personal communication). Identical residues are framed. The star indicates the serine residue of the active site (position 200 in *Torpedo*); the thick line (328–331) indicates a peptide which appears to be involved in the anionic subsite, because of its homology to an *Electrophorus* AChE labeled peptide (Kieffer *et al.*, 1986). Dots indicate divergences between the two *Torpedo* species. Asterisks indicate potential N-glycosylation sites in the case of *Torpedo* and *Drosophila* AChE, and observed glycosylated asparagines in the case of human BuChE. Dashed lines indicate the disulfide loops, as determined for *T. californica* AChE (MacPhee-Quigley *et al.*, 1986). Cysteine 231 was found to be free and cysteine 572 was found to be engaged in the formation of dimers. We underlined the region of divergence between *Drosophila* AChE and the other cholinesterases, starting at the position of the last intron in the coding sequence (D.Fournier, personal communication).

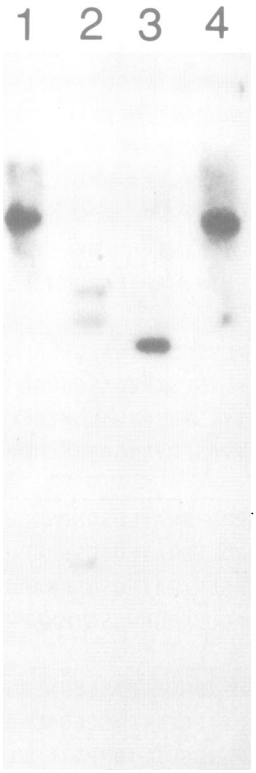

Fig. 5. Genomic blot of *Torpedo* DNA with an AChE probe. *Torpedo* liver DNA (10 μg) was digested to completion with *Bam*HI (**lane 1**), *Eco*RI (**lane 2**), *Pst*I (**lane 3**) and *Sac*I (**lane 4**). Hybridization, after electrophoresis and transfer, was done with a *Pst*I−*Pst*I fragment of the pAChE2 insert. Mol. wts were estimated from DNA and 1-kb ladder (BRL). The protected fragments were ~10 kb (*Bam*HI), 1.6, 4.8 and 5.6 kb (*Eco*RI), 4.3 kb (*Pst*I) and 9 kb (*Sac*I).

In preliminary S1 mapping experiments, we observed the existence of two types of AChE mRNAs, whose divergence is precisely located at the site of the putative exon−exon boundary. They probably correspond respectively to glycolipid-anchored globular forms and asymmetric collagen-tailed forms.

(iii) Vertebrate cholinesterases are more closely related to each other than to *Drosophila* AChE (204 conserved residues between *Torpedo* and *Drosophila* AChE, 192 between human BuChE and *Drosophila* AChE and 326 between *Torpedo* AChE and human BuChE). This suggests that the divergence between vertebrate AChE and BuChE occurred in the deuterostomian lineage, perhaps simultaneously with the emergence of vertebrates. Similarly, the C-terminal portion of thyroglobulin (Mercken *et al.*, 1985; Schumacher *et al.*, 1986; Swillens *et al.*, 1986) is more closely related to vertebrate cholinesterase than to *Drosophila* AChE since: (a) it shares 155 identical amino acids with *Torpedo* AChE, 147 with human BuChE, and only 117 with *Drosophila* AChE; and (b) the 30 amino acid long insertion found in *Drosophila* AChE is found neither in vertebrate ChE nor in vertebrate thyroglobulin. This relationship implies that this C-terminal domain of thyroglobulin appeared after the divergence between insects and vertebrates, and has therefore probably evolved by fusion of a cholinesterase or cholinesterase-like gene.

The remarkable maintenance of a strong homology, and particularly of the tertiary structure of the protein (conservation of disulfide bonds, and similarity of hydropathy profiles), together with the loss of the esteratic activity, raises the question of its

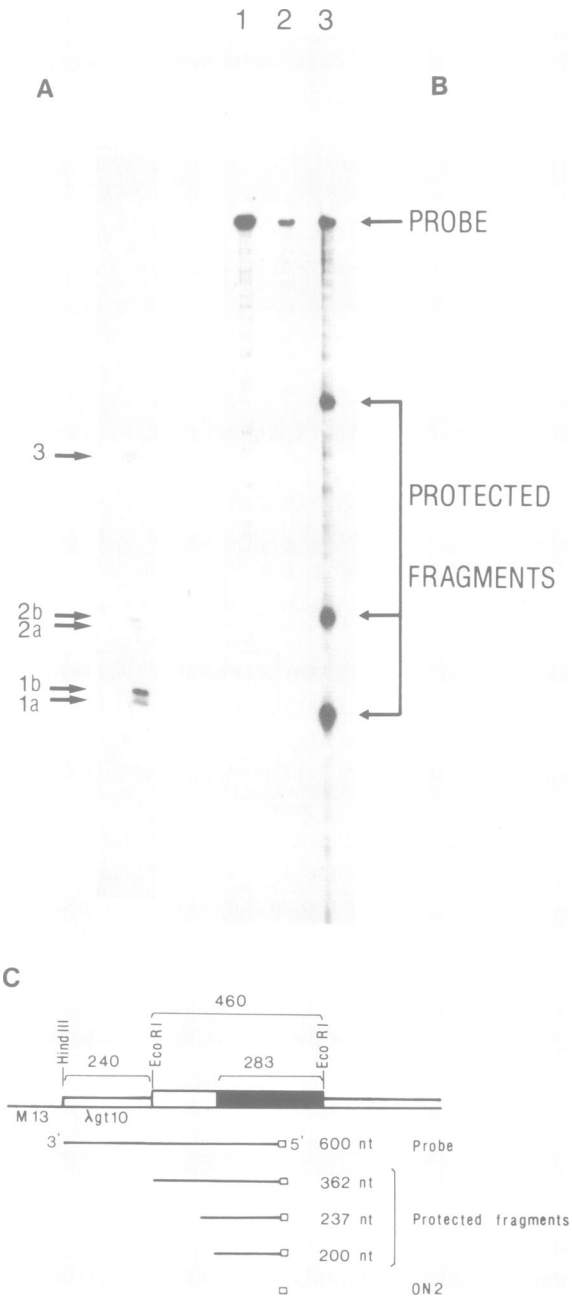

Fig. 6. Multiple 5′-untranslated regions. (**A**) Primer extension analysis of the AChE mRNAs. The oligonucleotide ON2 was hybridized to poly(A)$^+$ electric organ RNA and elongated with AMV reverse transcriptase. The extended products were analyzed by electrophoresis on a 4% sequencing gel. The sizes of the extended products were deduced from the end-labeled 1-kb ladder (BRL); 290 (1a, b), 330 (2a, b) and 430 nucleotides (3). (**B**) S1 mapping analysis of the 5′-untranslated region of AChE mRNAs. The phage λ5′AChE3 was digested with restriction enzymes *Hin*dIII and *Bgl*II (at coordinates 32.47 and 33.61 of λgt10) (cf. Huyhn *et al.*, 1985), producing a fragment that contains the totality of the *Eco*RI insert. M13mp19 was digested with *Hin*dIII and *Bam*HI, in order to subclone this fragment in the relevant orientation. A complementary probe was obtained, using oligonucleotide ON2 as a primer, and *Hin*dIII to cleave the extended product. This probe was used for S1 mapping. The mol. wts were estimated from a sequencing ladder (not shown); **lane 1**: undigested probe (10^4 c.p.m.); **lane 2**: wheat germ tRNA; **lane 3**: electric organ poly(A)$^+$ RNA (5 μg). Samples for **lanes 2** and **3** were digested with S1 nuclease. (**C**) Diagram of the structure of the probe and of the protected fragments. The closed bar represents the protein coding region, and the open bar, the 5′-untranslated region. The 5′−3′ polarity of the probe is indicated.

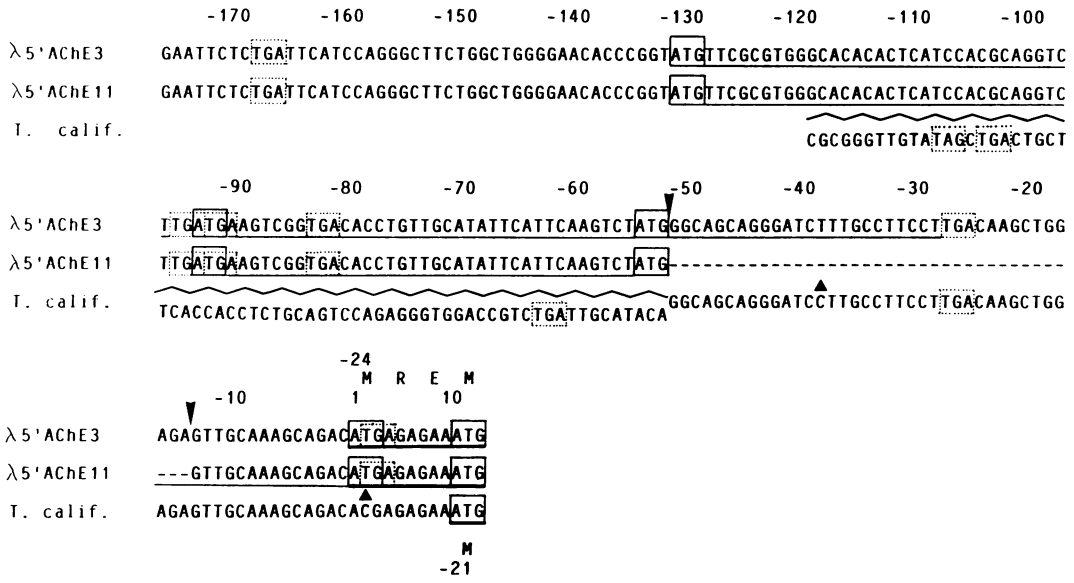

Fig. 7. Structure of the 5'-untranslated regions of AChE mRNAs. The structure of the 5'-untranslated regions obtained for six cDNA clones (Figure 3) was compared to that of clone λ5'AChE11, and a sequence obtained for *T. californica* (Schumacher, 1987). The nucleotides are numbered as in Figure 3, and the amino acids as in Figure 4. Arrows indicate the extremities of the S1-protected fragments (cf. Figure 6). We introduced insertions, indicated by dashes, to maximize the homologies between the three sequences. Triangles indicate differences between the 5'-untranslated sequence of *T. californica* and *T. marmorata* AChE. Upstream of position −52, the sequences of the two species are not homologous; this part of the alignment, indicated by a zig-zag line, is therefore arbitrary. Initiation and stop codons are framed in full line and dashed line, respectively. The open reading frames defined by the initiation and stop codons are underlined.

functional significance. It has been proposed that this structure is involved in membrane binding (Swillens *et al.*, 1986). However, this binding probably does not result from the attachment of a glycolipid anchor, since thyroglobulin is not endowed with an adequate C-terminal sequence. It is therefore tempting to suggest that the maintenance of the structure is necessary for protein−protein interactions, such as those existing between non-disulfide linked catalytic subunits of AChE in polymeric forms (Massoulié and Bon, 1982).

Active sites of ChEs

The residues involved in the charge-relay system (Rosenberry, 1975) are likely to be conserved between all cholinesterases. Only histidines 423 and 438 are conserved, and might therefore be involved in the catalytic mechanism.

A Gly-Ser-Phe-Phe sequence (residues 328−331), is likely to represent the equivalent of the tetrapeptide (Gly-Ser-X-Phe), which has been shown to exist in the anionic subsite of *Electrophorus* AChE (Kieffer *et al.*, 1986). Interestingly, the tetrapeptide is located in a hydrophobic region (12 amino acids) and contiguous to two negatively charged residues (at positions 326−327), which may interact with the quaternary ammonium of acetylcholine. X corresponds to phenylalanine in *Torpedo* and to tryptophan in *Drosophila*. Human BuChE contains an alanine at this position.

Existence of multiple 5'-untranslated regions

The 5'-untranslated region of the *Torpedo* AChE gene is characterized by two unusual features: it presents several variants, and may contain multiple initiation codons.

In *Drosophila*, the 5'-untranslated region is exceptionally long (up to 1000 nucleotides). Extension primer analysis (Figure 6A) suggests that it is probably shorter in *Torpedo* AChE mRNAs, and that the existence of the large AChE-mRNAs detected in Northern blots does not result from large 5'-untranslated regions. We cannot, however, rule out the possibility that longer reverse transcripts are not detected in this experiment.

The existence of distinct structures in the *Torpedo* 5'-untranslated region, indicated by cDNA clones, was confirmed by S1 mapping and by cDNA cloning (Figures 6B and 7). A 38 nucleotide long sequence is missing in the cDNA clone λ5'AChE11, and the limits of this sequence appear to coincide with extremities of protected fragments obtained by S1 mapping. Two of the extended products differ by ~40 nucleotides in their size (1a, b and 2a, b). This is likely to be due to the alternative splicing of this 38-nucleotide exon. In addition, the structure of a 5'-untranslated region observed for *T. californica* (Schumacher, 1987), is also consistent with the S1 mapping (Figure 7). λ5'AChE3 accounts for the 362-nucleotide protected species, λ5'AChE11 probably corresponds to the 200-nucleotide fragment and the sequence reported for *T. californica* to the 237-nucleotide fragment. The fact that the extension products are longer than the proposed corresponding cDNAs implies the existence of an additional 5' region, possibly that determined in the *T. californica* cDNA. The possible correlation between these multiple 5'-untranslated regions and the multiple mRNAs remains to be explored.

Some of the 5'-untranslated sequences contain upstream initiation codons. Interestingly, the existence of short open reading frames upstream of the coding sequence has been reported for several receptor mRNAs (β-adrenergic receptor, oestrogen receptor and muscarinic receptor) (Kubo *et al.*, 1986), as well as in *Drosophila* AChE (Hall and Spierer, 1986). The existence of two types of mRNAs, differing in the presence of upstream initiation codons, has been reported for the enzyme HMG-CoA reductase, which is involved in cholesterol biosynthesis, and is subject to a feed-back regulation of expression (Reynolds *et al.*, 1985). It has been suggested that the role of these multiple regions is to permit the control of the expression of HMG-CoA reductase at the translational level, in agreement with experimental evidence obtained in the case of the regulatory protein GCN4,

involved in the biosynthesis of amino acids in yeast (Mueller and Hinnebusch, 1986).

Our genomic blot experiments strongly suggest that all forms of AChE catalytic subunits derive from a single gene in *Torpedo*. Our current understanding of the structure of asymmetric and globular forms implies that their difference lies in the C-terminal region, and points to a differential splicing in the corresponding coding sequence, which has not yet been documented. We show here that differential splicing operates also in the 5′-untranslated region, and this does not seem related to the polymorphism of the enzyme.

Materials and methods

General methods

The preparation of DNA, digestion by restriction enzymes, electrophoresis and nucleic acid blotting were carried out by standard procedures (Maniatis *et al.*, 1982). Labeled nucleotides [α-^{32}P]dXTP (3000 Ci/mmol), [γ-^{32}P]ATP (5000 Ci/mmol), [α-^{32}P]ddATP (5000 Ci/mmol) and [α-^{35}S]thio-dXTP (400 Ci/mmol) were obtained from Amersham.

Construction and screening of cDNA libraries

A first cDNA library was constructed from *T. marmorata* electric organ poly(A)$^+$ RNA by standard techniques (Auffray *et al.*, 1980; Maniatis *et al.*, 1982); dC-tailed double-stranded cDNA was hybridized to *Pst*I-cleaved, dG-tailed pBR322, and used to transform *Escherichia coli* MC1061 (Hanahan, 1985). 10^4 clones were obtained and screened with the nick-translated *Eco*RI insert of a previously isolated cDNA clone, λAChE1 (Sikorav *et al.*, 1985).

To obtain the 5′-coding sequence missing from pAChE2, a restricted cDNA library was constructed in the following way. A 21-mer primer, ON1, (5′-CGTACTGCTGGCAGTTGTTGGG-3′), complementary to nucleotides 590–610 of the pAChE2 insert, was hybridized to 100 μg of poly(A)$^+$ RNA from electric organ, and extended wtih 30 units of AMV reverse transcriptase. (Boehringer-Mannheim). The RNAs were degraded with alkali, then the cDNAs were made double-stranded with the Klenow fragment of DNA polymerase I (New England Biolabs), treated with *Eco*RI methylase and S1 nuclease (PL Biochemicals). Phosphorylated *Eco*RI linkers were added to the blunted double-stranded cDNAs, which were then digested with *Eco*RI and inserted into λgt10 (Huyhn *et al.*, 1985). *In vitro* packaging yielded a library of 2×10^5 recombinant phages. The library was screened with ON2, a 21-mer polynucleotide (5′-GGCTCGGCAAAGGGAATCCCC-3′), complementary to nucleotides 490–510 or pAChE2 insert, and with restriction fragments derived from pAChE2.

Nucleotide sequence analysis

The *Eco*RI insert of λAChE1 (Sikorav *et al.*, 1985), was subcloned in pUC8 (Yanisch-Perron *et al.*, 1985), yielding a plasmid designated pAChE1. The nucleotide sequences of pAChE1 and pAChE2 inserts were determined by the chemical degradation of Maxam and Gilbert (1980). Restriction fragments were labeled at their 3′ end using either deoxynucleotidyl-terminal-transferase (BRL) for protruding ends or the Klenow fragment of *E. coli* Pol I for flush or recessed ends.

The clones obtained from the restriction cDNA library were sequenced by the dideoxynucleotide procedure (Sanger *et al.*, 1977).

Genomic blot analysis

DNA was prepared, digested with restriction enzymes, separated in 0.7% agarose gels and transferred to nylon membranes (Hybond N, Amersham) as described by Klarsfeld *et al.* (1984). High specific activity labeled probes ($\sim 2 \times 10^9$ c.p.m./μg), obtained by priming of DNA synthesis with random sequence hexanucleotides, were used for hybridization (Feinberg and Vogelstein, 1983). Filters were washed at 65°C to a stringency of 0.1 × SSC.

Primer extension analysis

The oligonucleotide ON2 was 5′ end-labeled with [γ-^{32}P]ATP using T4 polynucleotide kinase (New England Biolabs) resulting in a specific activity of 6×10^7 c.p.m./10 pmol of primer. 2.5×10^6 c.p.m. of primer was hybridized to 100 μg of poly(A)$^+$ RNA from the electric organ, for 90 min at 60°C in the presence of 0.1 M NaCl. After hybridization, the nucleic acids were recovered by ethanol precipitation and extended with 30 units of AMV reverse transcriptase. This treatment was followed by a degradation of RNA wtih 10 μg of RNase A for 15 min at room temperature, phenol extraction and ethanol precipitation. 1/100 of the sample was analyzed on a 4% sequencing gel, and the rest was subjected to partial chemical degradations (Maxam and Gilbert, 1980), prior to electrophoresis in an 8% sequencing gel.

S1 nuclease mapping

Uniformly labeled, single-stranded probes were synthesized and isolated as described (Kelly *et al.*, 1983). Briefly, a relevant restriction fragment was subcloned in an M13 vector (Yanisch-Perron *et al.*, 1985), and a phage containing the insert in the adequate orientation was characterized by sequence analysis. The phage DNA was hybridized to a complementary oligonucleotide, either M13 universal primer (Amersham) or oligonucleotide ON2. The complementary oligonucleotide was extended by the Klenow fragment of DNA Pol I, in the presence of [α-^{32}P]-dCTP (3000 Ci/mmol). The extended fragments were isolated by electrophoresis on a 4% sequencing gel, and eluted by diffusion for 1 h at 37°C.

Probes were hybridized to poly(A)$^+$ RNA in 50% formamide, 0.5 M NaCl, 40 mM Pipes pH 6.4, 1 mM EDTA, for 18 h at 42°C. After hybridization, samples were reacted for 1 h at 37°C with 50 units of S1 nuclease and analyzed in a 4% sequencing gel, and eluted by diffusion for 1 h at 37°C.

Acknowledgements

We thank Dr François Rougeon for critical reading of the manuscript, Nathalie Duval and Bernard Reimund for their help, as well as Thierry Ciora and Eric Clauser for the synthesis of oligonucleotides. We also thank Drs T.L.Rosenberry, P.Taylor and D.Fournier for communication of their unpublished results. This research was supported by the Centre National de la Recherche Scientifique, the Ecole Normale Supérieure, the Muscular Dystrophy Association of America and the Direction des Etudes et Recherches Techniques. J.-L.S. was a recipient of a Rhône-Poulenc fellowship.

References

Auffray,C., Nageotte,R., Chambraud,B. and Rougeon,F. (1980) *Nucleic Acids Res.*, **8**, 1231–1241.

Avvedimento,V.E., Musti,A.M., Obici,S., Cocozza,S. and Di Lauro,R. (1984) *Nucleic Acids Res.*, **12**, 3461–3472.

Bon,S., Chang,J.-Y. and Strosberg,A.D. (1986) *FEBS Lett.*, **209**, 206–212.

Citri,Y., Colot,H.V., Jacquier,A.C., Qiang Yu, Hall,J.C., Baltimore,D. and Rosbash,M. (1987) *Nature*, **48**, 42–47.

Cross,G.A.M. (1987) *Cell*, **48**, 179–181.

Doctor,B.P., Camp,S., Gentry,M.K., Taylor,S.S. and Taylor,P. (1983) *Proc. Natl. Acad. Sci. USA*, **80**, 5767–5771.

Feinberg,A.P. and Vogelstein,B. (1983) *Anal. Biochem.*, **136**, 6–13.

Gnagey,A.L., Forte,M. and Rosenberry,T.L. (1987) *J. Biol. Chem.*, in press.

Hall,L.M.C. and Spierer,P. (1986) *EMBO J.*, **5**, 2949–2954.

Hanahan,D. (1985) In Glover,D.M. (ed.), *DNA Cloning – A Practical Approach*. IRL Press, Oxford and Washington, DC, Vol. 1, pp. 109–135.

Huynh,T.V., Young,R.A. and Davis,R.W. (1985) In Glover,D.M. (ed.), *DNA Cloning – A Practical Approach*. IRL Press, Oxford and Washington, DC, Vol. 1, pp. 49–78.

Kelly,K., Cochran,B.H., Stiles,C.D. and Leder,P. (1983) *Cell*, **35**, 603–610.

Kieffer,B., Goeldner,M., Hirth,C., Aebersold,R. and Chang,J.Y. (1986) *FEBS Lett.*, **202**, 91–96.

Klarsfeld,A., Devillers-Thiéry,A., Giraudat,J. and Changeux,J.P. (1984) *EMBO J.*, **3**, 35–41.

Kubo,T., Fukuda,K., Mikami,A., Maeda,A., Takahashi,H., Mishina,M., Haga,T., Haga,K., Ichiyama,A., Kangawa,K., Kojima,M., Matsuo,H., Hirose,T. and Numa,S. (1986) *Nature*, **323**, 411–416.

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.

MacPhee-Quigley,K., Taylor,P. and Taylor,S. (1985) *J. Biol. Chem.*, **260**, 12185–12189.

MacPhee-Quigley,K., Vedvick,T., Taylor,P. and Taylor,S.S. (1986) *J. Biol. Chem.*, **261**, 13565–13570.

Maniatis,T., Fritsch,E.F. and Sambrook,J. (1982) *Molecular Cloning. A Laboratory Manual*. Cold Spring Harbor Laboratory Press, NY.

Massoulié,J. and Bon,S. (1982) *Annu. Rev. Neurosci.*, **5**, 57–106.

Maxam,A.M. and Gilbert,W. (1980) *Methods Enzymol.*, **65**, 499–566.

Mercken,L., Simons,M.J., Swillens,S., Massaer,M. and Vassart,G. (1985) *Nature*, **316**, 647–651.

Mueller,P.P. and Hinnebusch,A.G. (1986) *Cell*, **45**, 201–207.

Prody,C.A., Zevin-Sonkin,D., Gnatt,A., Goldberg,O. and Soreq,H. (1987) *Proc. Natl. Acad. Sci. USA*, in press.

Reynolds,G.A., Goldstein,J.L. and Brown,M.S. (1985) *J. Biol. Chem.*, **260**, 10369–10377.

Rosenberry,T.L. (1975) *Adv. Enzymol.*, **143**, 103–218.

Rosenberry,T.L. (1985) In Martonosi,A.N. (ed.), *The Enzymes of Biological Membranes*. Plenum Publishing Corporation, NY, p. 403.
Sanger,F., Nicklen,S. and Coulson,A.R. (1977) *Proc. Natl. Acad. Sci. USA*, **74**, 5463–5467.
Schumacher,M. (1987) PhD Dissertation. University of California, San Diego.
Schumacher,M., Camp,S., Maulet,Y., Newton,M., MacPhee-Quigley,K., Taylor,S.S., Friedmann,T. and Taylor,P. (1986) *Nature*, **319**, 407–409.
Sikorav,J.L., Grassi,J. and Bon,S. (1984) *Eur. J. Biochem.*, **145**, 519–524.
Sikorav,J.L., Vallette,F., Grassi,J. and Massoulié,J. (1985) *FEBS Lett.*, **193**, 159–163.
Silman,I. and Futerman,A.H. (1987) *Biopolymers*, in press.
Swillens,S., Ludgate,M., Mercken,L., Dumont,J.E. and Vassart,G. (1986) *Biochem. Biophys. Res. Commun.*, **137**, 142–148.
Toutant,J.-P., Massoulié,J. and Bon,S. (1985) *J. Neurochem.*, **44**, 580–592.
Yanisch-Perron,C., Vieira,J. and Messing,J. (1985) *Gene*, **33**, 103–119.

Received on March 10, 1987; revised on April 21, 1987