*Proc. Natl. Acad. Sci. USA*
Vol. 84, pp. 2813–2817, May 1987
Evolution

# Evolutionary origin of cholinergic macromolecules and thyroglobulin

(choline acetyltransferase/acetylcholinesterase/acetylcholine receptor/gene recruitment/evolution)

NOZOMU MORI*†, NOBUYUKI ITOH‡§, AND PAUL M. SALVATERRA‡

Divisions of *Biology and ‡Neurosciences, Beckman Research Institute of the City of Hope, 1450 East Duarte Road, Duarte, CA 91010

*Contributed by Susumu Ohno, November 6, 1986*

ABSTRACT **We have compared the amino acid sequences of proteins that are involved in acetylcholine (AcCho) metabolism and cholinergic neurotransmission: choline acetyltransferase (ChoAcTase), acetylcholinesterase (AcChoEase), and a neuronal α subunit of nicotinic AcCho receptor (AcChoR). A comparison of *Drosophila* ChoAcTase and rat neuronal α subunit of AcChoR shows a limited segmental type homology, which may suggest a similar acetylcholine binding site in the two proteins evolving by convergence. We note a global homology of 21–44% identity between *Drosophila* ChoAcTase and *Torpedo* AcChoEase. Six homologous segments of 40–60 amino acids cover 38% and 54% of the sequences, raising the possibility of a common evolutionary origin. We also note that mammalian thyroglobulin (TG), the precursor for thyroid hormones, contains an AcChoEase-like sequence at its carboxyl end. This homology raises the possibility that the gene for TG has evolved by gene fusion or condensation (i.e., recruiting a preexisting redundant copy of a gene for AcChoEase during vertebrate evolution). Our results demonstrate that the record of evolutionary history for nervous system proteins can be read across the boundaries of separation between vertebrates and invertebrates. They also provide molecular evidence for the common evolutionary origins of the nervous and endocrine systems in vertebrates—both evolving to make intercellular communication possible.**

Chemical neurotransmission requires gene products to act in a coordinated manner during the synthesis, degradation, and reception of a neurotransmitter. All of these gene products should have some structural homology in their amino acid sequences, allowing them to bind and process a common transmitter. In addition they should have unique sequences to carry out their individual biochemical tasks. With the application of molecular cloning techniques to neurobiologically important macromolecules, it is now possible to directly compare the amino acid sequences of neurotransmitter biosynthetic and degradative enzymes (1–4) and receptors (5–10). Such comparisons not only identify homologous regions within related proteins but also provide evidence for their evolutionary origins.

Acetylcholine (AcCho), the first chemical neurotransmitter identified, has a universal distribution in animal nervous systems and the proteins with which it interacts are well characterized. The nicotinic receptor for AcCho (AcChoR) in vertebrate muscle and fish electric organs has been cloned from a variety of species (see refs. 11–14 for recent reviews). In addition, a nicotinic α-subunit-type AcChoR has been isolated from a rat pheochromocytoma cell line, PC12, and is thought to represent a neuronal type AcChoR (9). Recently, the catabolic enzyme for AcCho (acetylcholinesterase, AcChoEase) has been cloned from *Torpedo californica* (4), and we have isolated and sequenced a cDNA clone for the anabolic enzyme choline acetyltransferase (ChoAcTase) of *Drosophila melanogaster* (3). Even though we only know the sequences for these macromolecules in different species of divergent evolutionary phylogeny, a comparison of their sequences reveals several interesting features. In addition, we describe a detailed analysis of the surprising homology between *Torpedo* AcChoEase and rat thyroglobulin (TG).

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

## METHODS

The amino acid sequence of *Drosophila* ChoAcTase (728 residues) has been deduced from the sequence of a cDNA clone, pCha-2 (3), and partially confirmed by microsequencing several tryptic peptides isolated from purified ChoAcTase (15). Homologous sequences reported in this paper were identified by visual inspection of the deduced amino acid sequences from published cDNA sequences. Final alignments were optimized by using the Wilbur and Lipman algorithm (16) (i.e., the ALIGN program available through BIONET). Sequence homology searches were performed on BIONET by using the IFIND program and searching the European Molecular Biology Laboratory (EMBL) and GenBank DNA sequence data bases[¶] and the National Biomedical Research Foundation protein sequence data base.[‖]

## RESULTS AND DISCUSSION

A comparison of the amino acid or nucleic acid sequence of *Drosophila* ChoAcTase with the data bases revealed neither significant local nor global homology. The best homology to the ChoAcTase cDNA sequence was found for epidermal growth factor precursor cDNA (17, 18) (47%), but most of the matches were out of reading frame with respect to ChoAcTase. The following analysis was performed on sequences not yet represented in the data bases.

**Homologous Domains in *Drosophila* ChoAcTase and *Torpedo* AcChoEase.** There is a striking global homology between *Drosophila* ChoAcTase and *Torpedo* AcChoEase. Alignment of six polypeptide segments along the length of the two

Abbreviations: AcCho, acetylcholine; AcChoEase, acetylcholinesterase; AcChoR, acetylcholine receptor; ChoAcTase, choline acetyltransferase; TG, thyroglobulin.

†Present address: Division of Biology, California Institute of Technology, Pasadena, CA 91125.

§Present address: Department of Biological Chemistry, Faculty of Pharmaceutical Sciences, Kyoto University, Kyoto 606, Japan.

¶European Molecular Biology Laboratory (1986) EMBL Nucleotide Sequence Data Library (Eur. Mol. Biol. Lab., Heidelberg, F.R.G.), Tape Release 7.0; and National Institutes of Health (1986) Genetic Sequence Databank: GenBank (Research Systems Div., Bolt, Beranek, and Newman, Cambridge, MA), Tape Release 40.0.

‖Protein Identification Resource (1985) Protein Sequence Database (Natl. Biomed. Res. Found., Washington, DC), Release 7.0.

polypeptide chains includes 274- and 319-amino acid residues, which cover 38% and 54% of the sequences and show 19–39% identity, including only one or two gaps in three sequences (segments 2, 3, and 5) and no gaps in the others (segments 1, 4, and 6) (Fig. 1). The degree of homology ranges from 27% to 55% when homologous amino acid replacement (19) is considered for calculation. This homology level is similar to that found between homeobox sequences and the yeast mating-type regulatory proteins (20). However, it should be noted that, in the present comparison, the homology starts at the amino-terminal region and spans almost the entire coding region of the two proteins.

Although the statistical significance of such homologies does not always imply secondary or tertiary structural similarity in proteins (21), it often acts as a guide to a possible evolutionary relationship (22, 23). In this regard, the active-site hexapeptide in *Torpedo* AcChoEase (residues 198–203) is completely replaced in *Drosophila* ChoAcTase (residues 172–177). Since the active site for ChoAcTase is not known, we cannot assess the importance of this observed divergence for catalytic function. It seems likely, however, that these two proteins share a common evolutionary origin. It is interesting in this regard that AcChoEase, like ChoAcTase, has been shown to catalyze acylation of choline (24). Although *Torpedo* AcChoEase is homologous to other esterases in the active site region (25), no significant homology extending throughout the sequences was found with other serine hydrolases (4). This implies that the origin of AcChoEase may be different from the pancreatic proteases.

**Segmental Homology Between ChoAcTase and Neuronal Nicotinic AcChoR α Subunit.** Although the sequences are short and segmental, five homologous sequences could be seen between *Drosophila* ChoAcTase and rat neuronal nicotinic AcChoR α subunit (Fig. 2). The five regions in AcChoR (segments 1–5) are all thought to be in the extracellular portion of this integral membrane protein (5–9, 26). Three of them (segments 2, 4, and 5) are located in the vicinity of the presumed AcCho binding site, which forms two loops connected by disulfide bonds between Cys-192 and Cys-193 and between Cys-128 and Cys-142 (27, 28). The other two regions (segments 1 and 3) are in the amino-terminal signal peptide,

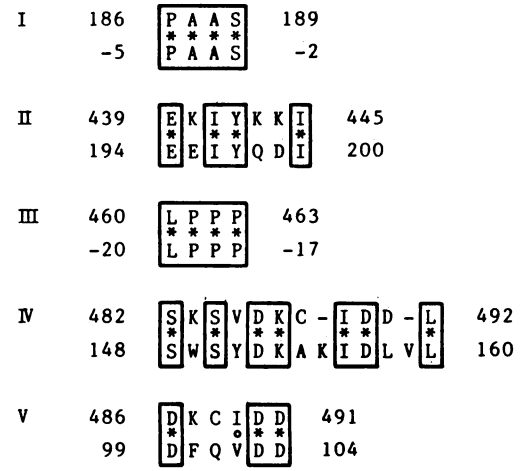


FIG. 2. Comparison of homologous segments in *Drosophila* ChoAcTase (upper line of each row) and rat neuronal nicotinic AcChoR α subunit (9) (lower line of each row).

which is posttranslationally processed in the course of AcChoR subunit assembly (29, 30). Four of these homologous sequences in ChoAcTase (segments 2–5) cluster in a central region around residues 439–492, where four cysteine residues are present at positions 423, 428, 488, and 510 (3) (Fig. 3). If two disulfide bonds are formed between Cys-423 and Cys-428 and between Cys-488 and Cys-510, they bring the four homologous sequences with AcChoR (segments 2–5) into close proximity with this region of ChoAcTase. Segments 2, 4, and 5 contain negatively charged amino acids, which could be possible candidates for binding to positively charged choline. The homologous segments of ChoAcTase and neuronal AcChoR are not always conserved in muscle nicotinic AcChoR. In addition, none of these five mutually homologous sequences could be found in *Torpedo* AcChoEase. Consequently, the segmental homologies between *Drosophila* ChoAcTase and rat neuronal AcChoR α subunit seem to reflect convergent rather than divergent evolution.

**Homology Between *Torpedo* AcChoEase and the Carboxyl-Terminal Hormonogenic Domain of Rat TG.** Schumacher *et*

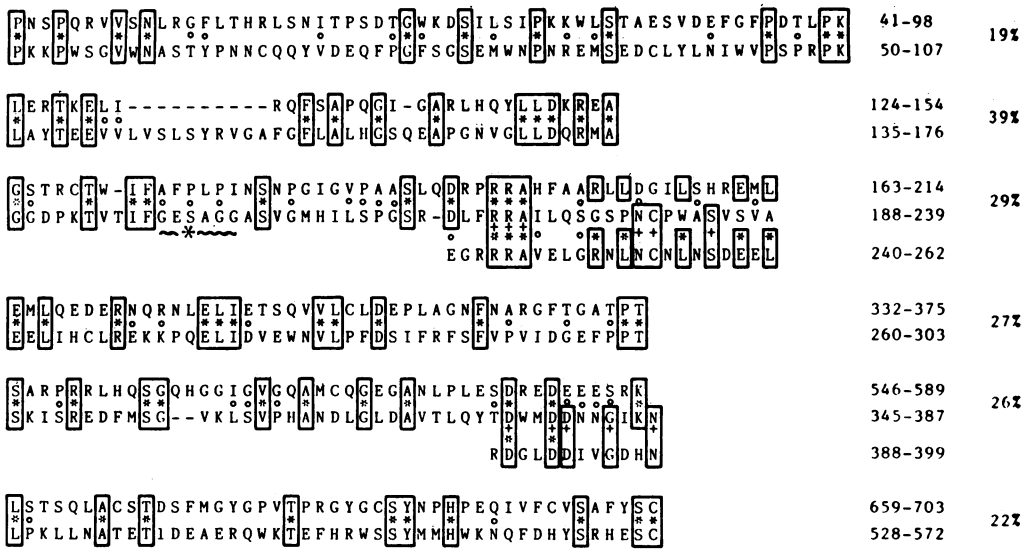


FIG. 1. Homologous amino acid sequences of *Drosophila* ChoAcTase (3) (upper line of each row) and *Torpedo* AcChoEase (4) (lower line of each row). The six specific regions within the proteins were selected to exhibit the greatest homology. Identities are marked by asterisks and boxed. Intrasequence identities in the AcChoEase sequence are indicated by a + and boxed. Open circles indicate homologous amino acid replacements (19). Numbering on the right denotes the amino acid positions and percent identities, where gap regions were not considered for the calculation. The active-site residue (Ser-200) for AcChoEase is marked by a large asterisk between two thick wavy lines. Amino acids are designated by standard one letter abbreviations.

*al.* (4) noted an unexpected homology between *Torpedo* AcChoEase and bovine TG (31). They reported that 546- and 540-residue segments of the *Torpedo* AcChoEase and bovine TG share 28% identity, with five gaps in the alignment, and that the positions of six of the eight cysteines in the homologous segments are conserved. Swillens *et al.* (32) extended the analysis, including the hydropathy profiles, to confirm a similar three-dimensional structure in the homologous regions of the two proteins. They proposed that the homologous regions are involved in interaction with cell membranes, although neither protein is very hydrophobic. In addition, recent biochemical studies of AcChoEase indicate that the enzyme is attached to certain locations in membranes through a covalently linked phosphatidylinositol at the carboxyl terminus of the protein molecule (33, 34). This attachment method has also been suggested for several other surface glycoproteins (reviewed in refs. 35 and 36).

Recently the complete organization of the rat TG gene has been reported (37). The TG gene spans >170 kilobases (kb) and is distributed in 42 exons and 41 introns. As the exon/intron boundaries in the 3′ half of the gene (38) as well as the sequence of 967 amino acids at the carboxyl-terminal end of rat TG (39) were already known, we compared the best alignments between AcChoEase of *Torpedo californica* and TG of the rat (Fig. 4).

In contrast to the previous analyses with bovine TG (5, 32), our alignment shows homology for the entire sequence of *Torpedo* AcChoEase, which contains the amino-terminal signal peptide as well as the carboxyl-terminal portion and the whole carboxyl-terminal unique region of rat TG (residues 2169–2750), which begins just after the repetitive domain composed by the type III motif (31). Our alignment shows 28% identical residues. As already noted (4, 31), the positions of six of the eight cysteines in the homologous segments are conserved. In addition, there is a region that shares high homology (35–48%) (residues 123–225 in AcChoEase and residues 2292–2395 in TG corresponding to exons 35 and 36), while the active-site Ser-200 of AcChoEase and two of the three hormonogenic tyrosines (at positions 2555 and 2569) of rat TG are not conserved. The homologous structure of the carboxyl-terminal end of TG with AcChoEase may suggest that TG, like AcChoEase, may also attach to cellular membranes via a phosphatidylinositol hydrophobic anchor.

**TG Gene and Presumable AcChoEase Gene Organization in Vertebrates.** We have also analyzed the homologous region between AcChoEase and TG by comparing AcChoEase with the organization of exon/intron boundaries in the rat TG gene (Fig. 5). The coding information for rat TG is distributed in 42 exons, most of which correlate with repetitive structural domains previously defined by protein sequences deduced from human, rat, and bovine TG cDNA (31, 39–41). The 5′-end half of this gene, including 32 exons and spanning >80 kb, encodes highly repetitive domains (38). This portion of TG seems to have arisen basically as a consequence of sequential duplication and recombination of three types of cysteine-rich motifs (31). In contrast, the last 580 residues at the carboxyl terminus are unique, showing no repetitive sequences, being poor in cysteine, and containing a cluster of tyrosines of which three are hormonogenic (31). It is in this region that the homology with AcChoEase exists. It should be noted that the homology starts immediately after the end of the repetitious domain and ends at the carboxyl terminus of TG (see Fig. 5). The boundary of the repetitious domain and the AcChoEase-like unique domain (i.e., exon 32/33) is interrupted by intron 32 with a length of 3 kb (38). The size of this intron is relatively long compared to other introns in the 5′-end half of the rat TG gene (37, 38). Thus, the contiguous region of homology is encoded by the last 10 exons (exons 33–42) of the 3′ end of the gene.

In general, the positions at which introns interrupt homologous genes in vertebrates are conserved. For example, genes for two functionally different proteins, low-density lipoprotein receptor and epidermal growth factor precursor (42, 43) and the homologous blood coagulation proteases (44, 45) have conserved their relative intron positions. Thus, it is reasonable to predict that the AcChoEase gene in higher vertebrates will be found to be interrupted by introns located in similar positions to those found in the rat TG gene. In several cases during the evolution of homologous genes, intron loss (46) and intron sliding (47) have been postulated to account for seemingly anomalous positions of introns. Even if these mechanisms were operating during evolution of

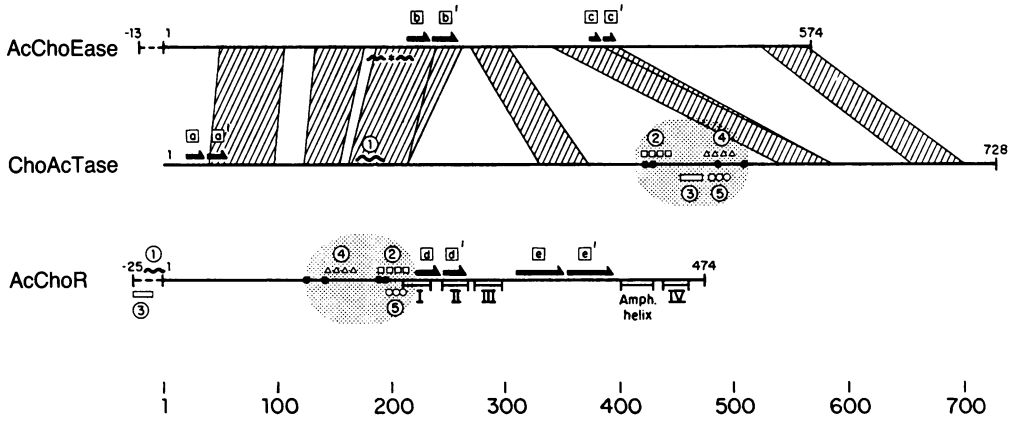


FIG. 3. Schematic representation of structural homologies within and among *Drosophila* ChoAcTase, *Torpedo* AcChoEase, and rat neuronal nicotinic AcChoR α subunit. Each sequence is represented as a line, while a dashed line indicates the signal sequence or the sequence only included in precursors. The amino-terminal position of the *Drosophila* ChoAcTase has not yet been identified; therefore, position 1 only indicates the furthest upstream residue so far sequenced (3). Homologous regions between ChoAcTase and AcChoEase are indicated by hatching between the two sequences (see Fig. 1). The active-site peptide for AcChoEase is marked as in Fig. 1. The circled numbers denote the five homologous segments between ChoAcTase and AcChoR α subunit, distinguished by different symbols at approximate positions along each sequence (see Fig. 2). Thick arrows indicate internally duplicated sequences (data not shown). The four extracellular cysteine residues characteristic of AcChoR α subunit (9) are marked by closed circles. Also shown by closed circles are the four cysteine residues in ChoAcTase. These residues are located in the vicinity of the homologous clustered segments (2–5) when comparing ChoAcTase with the AcChoR α subunit. The two regions enclosed by a stippled oval include all four cysteine residues and the homologous segments of ChoAcTase and the AcChoR α subunit. The transmembrane regions of AcChoR (I–IV) and the amphipathic helix (26) are indicated. Amino acid residues are numbered below the lines.

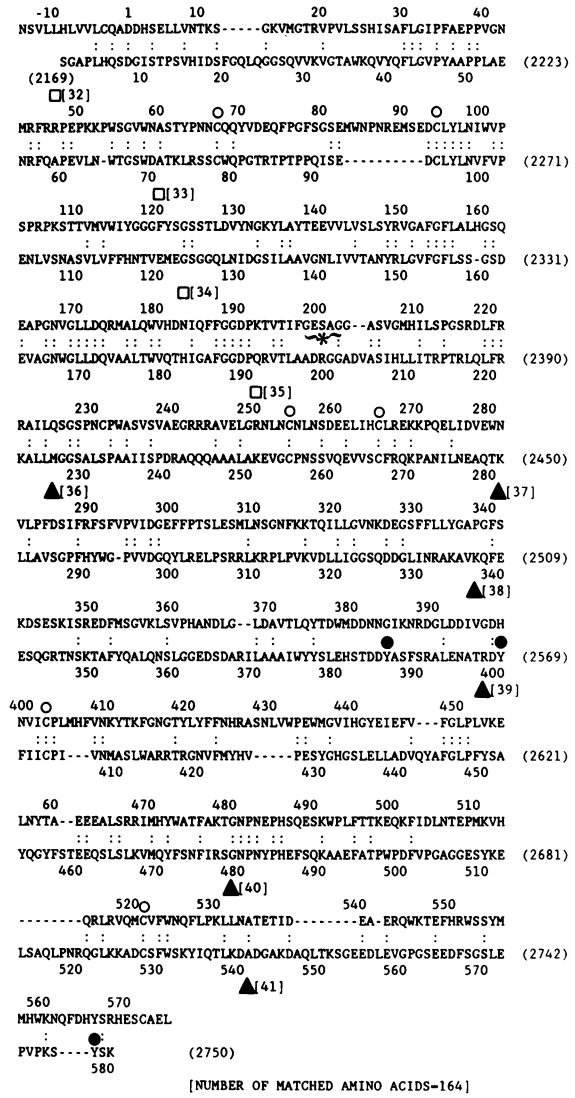

FIG. 4. Optimal alignment of amino acid sequences for *Torpedo* AcChoEase (upper line of each row) and the homologous region of rat TG (i.e., the carboxyl terminus) (lower line of each row). The sequences were aligned by using the ALIGN Program with K-tuple = 1, window size = 20, and gap penalty = 2. Changing these parameters yields several other possible alignments (i.e., for K-tuple/gap penalty of 1/1, 2/1, 1/2, or 2/2 results in alignments with 183, 159, 159, and 128 matched amino acids and 24, 33, 12, and 11 gaps, respectively (i.e., 22–32% homology). In the alignment shown here, there are 164 matched amino acids and 7 gaps for each sequence (i.e., 28% homology). Identical residues are marked by colons between the sequences. The numbers along the sequences refer to residue positions. The alignment arbitrarily starts with the serine residue immediately following the end of the last repetitive domain (the type III a3 motif), since only 967 amino acids are known for rat TG. This serine corresponds to position 2169 in bovine TG, where the total sequence is known. The right side of each row of the TG sequence shows the corresponding positions of amino acids in bovine TG in parentheses. ○, Six homologous cysteine residues; ●, hormonogenic tyrosine residues; ▲, definitive intron positions for TG; □, less certain intron positions for TG with the number in brackets. The active-site peptide of AcChoEase is indicated as in Fig. 1.

the AcChoEase and TG genes, the coding region of the AcChoEase gene in vertebrates should still be basically distributed in a similar manner to that observed in the homologous gene segment coding for the carboxyl-terminal domain of rat TG.

However, no position of the 11 gaps introduced for alignment in Fig. 4 corresponds to the 10 introns (introns 32–41) in the rat TG gene. Other possible alignments using different parameters (data not shown) also do not produce corresponding gaps and introns. This may imply that many mutations, including substitution, insertion, and deletion, have accumulated in a redundant copy of the AcChoEase gene, before its fusion to the 80-kb-long 5′-end half of the present TG gene.

**Concluding Remarks.** We have identified amino acid sequence homologies among *Drosophila* ChoAcTase, *Torpedo* AcChoEase, rat neuronal nicotinic AcChoR $\alpha$ subunit, and rat TG and discussed their possible evolutionary relationships (summarized in Figs. 3 and 5). Although the three proteins involved in AcCho metabolism, ChoAcTase, AcChoEase, and neuronal AcChoR, are all localized to cholinergic synapses, they are found in different intracellular compartments and have different functional roles during cholinergic neurotransmission. The expression of these three proteins is thought to be regulated by the formation of cholinergic synapses during nervous system development and through the long history of evolution of neuronal systems. The traces of evolution could be detected in the amino acid sequences of these three important neurobiological proteins. The rat neuronal AcChoR $\alpha$ subunit and *Drosophila* ChoAcTase share several interesting similarities that may have arisen by convergent evolution. In contrast, the global sequence homology between *Drosophila* ChoAcTase and *Torpedo* AcChoEase suggests that their genes have evolved divergently from a common ancestral gene at the time when animals began to develop a primitive nervous system. Perhaps this implies that genes for anabolic and catabolic enzymes in other metabolic pathways may also have evolved from a common ancestral gene. In addition to these homologies, the homology found between AcChoEase and TG was striking and surprising. Apparently the origin of ChoAcTase and AcChoEase is older than TG. They seem to have been differentiated about 800 million years ago, when vertebrates and invertebrates were evolving separately. Thyroid organ is considered to have developed in the fishes only 400–450 million years ago. Consequently, it is during the 350–400 million years in the Precambrian period, after the separation of vertebrates from invertebrates and before the emergence of the fishes (48), that a redundant copy of the AcChoEase gene may have accumulated many mutations and fused to generate a functional TG gene. We cannot answer the question of why an AcChoEase-like sequence was incorporated into TG, but the homology between the two amino acid sequences tells us that it was. Since the gene fusion in the TG gene is a more recent event than the divergence of AcChoEase and ChoAcTase genes, we predict that some vertebrate AcChoEase gene sequences will show much stronger homology to the TG gene than that observed here with Torpedo AcChoEase. It should be noted that we find significant homology between ChoAcTase, AcChoEase, and TG despite the considerable evolutionary distance among *Drosophila*, an insect, *Torpedo*, a marine ray, and rat and bovine mammals. As more sequences become available for these proteins in other species, it may be possible to confirm the generality of our hypothesis that genes for ChoAcTase, AcChoEase, and even TG diverged from a common ancestral gene and reorganized through genomic condensation and exonic shuffling, as has been proposed for many genes encoding vertebrate proteins (49–51), at the time of formation of nervous systems and then endocrine systems in animals.

We thank Drs. S. Ohno and T. Okada for advice and encouragement; Drs. J. Ono, K. Ikeda, and J. Singer-Sam for discussions; and Dr. A. Riggs for support. This work was supported by National Institutes of Health Grant NS19482.

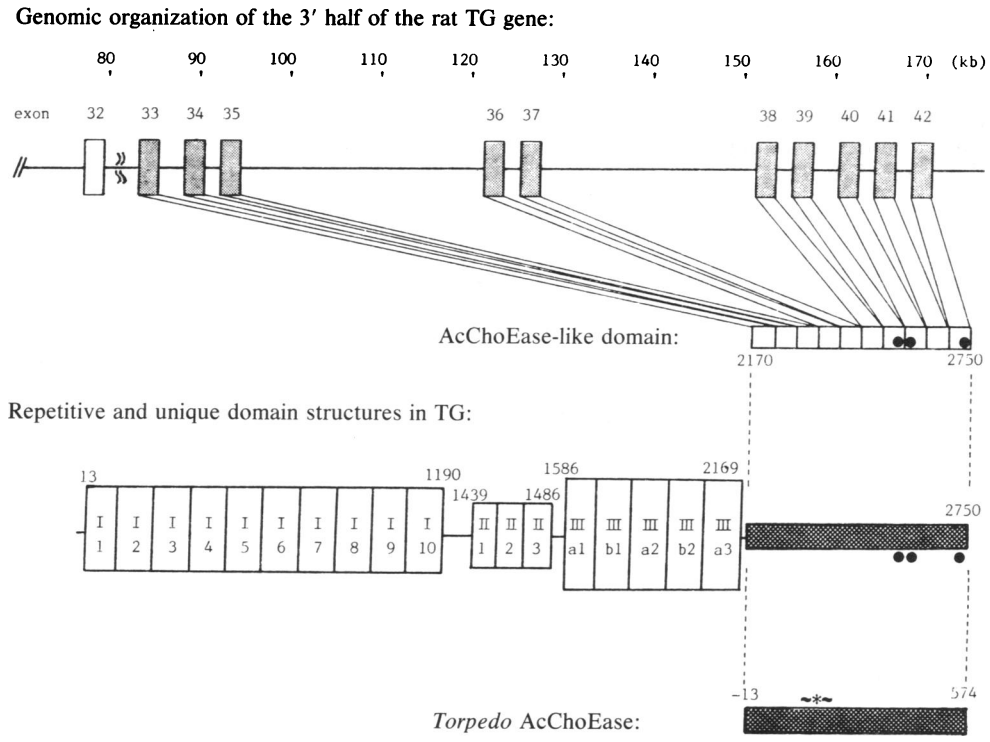


FIG. 5. Schematic representation of homologous domains of AcChoEase and TG (*Lower*) and the genomic organization of part of the rat TG gene encoding the AcChoEase-like domain (*Upper*). (*Lower*) The three types of repetitive domains are represented as I, II, and III as described (31). The residue number for the start and end of each repetitive domain in bovine TG is indicated. The unique carboxyl-terminal TG domain showing homology with AcChoEase and AcChoEase is crosshatched. (*Upper*) The 3′-end half (80–170 kb downstream from the first exon) of the rat TG gene is shown as described (37). Exons are indicated by boxes, and introns, by a continuous line; the exons that encode the AcChoEase-like domain (exon 33–42) are stippled. The wavy lines between exons 32 and 33 indicate the division of the genomic region encoding the repetitive and the unique AcChoEase-like domain in rat TG. The positions of the three hormonogenic tyrosine residues in TG and the active site residues in AcChoEase are indicated as in Fig. 4.


1. Chikaraishi, D. M., Brilliant, M. H. & Lewis, E. J. (1983) *Cold Spring Harbor Symp. Quant. Biol.* **48,** 309–318.
2. Grima, B., Lamouroux, A., Blamot, F., Biguet, N. F. & Mallet, J. (1985) *Proc. Natl. Acad. Sci. USA* **82,** 617–621.
3. Itoh, N., Slemmon, J. R., Crawford, G. D., Morita, E., Itakura, K., Hawke, D., Shively, J. E., Williamson, R. & Salvaterra, P. M. (1986) *Proc. Natl. Acad. Sci. USA* **83,** 4081–4085.
4. Schumacher, M., Camp, S., Maulet, Y., Newton, M., MacPhee-Quigley, K., Taylor, S. S., Friedmann, T. & Taylor, P. (1986) *Nature (London)* **319,** 407–409.
5. Noda, M., Takahashi, H., Tanabe, T., Toyosato, M., Furutami, Y., Hirose, T., Asai, M., Inayama, S., Miyata, T. & Numa, S. (1982) *Nature (London)* **299,** 793–797.
6. Noda, M., Takahashi, H., Tanabe, T., Toyosato, M., Kikyotani, S., Furutani, Y., Hirose, T., Takashima, H., Inayama, S., Miyata, T. & Numa, S. (1983) *Nature (London)* **302,** 528–532.
7. Numa, S., Noda, M., Takahashi, H., Tanabe, T., Toyosato, M., Furutani, Y. & Kikyotani, S. (1983) *Cold Spring Harbor Symp. Quant. Biol.* **48,** 57–69.
8. Boulter, J., Luyten, W., Evans, K., Mason, P., Ballivet, M., Goldman, D., Stengelin, S., Martin, G., Heinemann, S. & Patrick, J. (1985) *Neuroscience* **5,** 2545–2552.
9. Boulter, J., Evans, K., Goldman, D., Martin, G., Treco, D., Heinemann, S. & Patrick, J. (1986) *Nature (London)* **319,** 368–374.
10. Dixon, R. A. F., Kobilka, B. K., Strader, D. J., Benovic, J. L., Dohlman, H. G., Frielle, T., Bolanowski, M. A., Bennett, C. D., Rands, E., Diehl, R. E., Mumford, R. A., Slater, E. E., Sigal, I. S., Caron, M. G., Lefkowitz, R. J. & Strader, C. D. (1986) *Nature (London)* **321,** 75–79.
11. Tůcek, S. (1983) in *Handbook of Neurochemistry*, ed. Lajatha, A. (Plenum, New York), Vol. 4, pp. 219–249.
12. Conti-Tronconi, B. M. & Raftery, M. (1982) *Annu. Rev. Biochem.* **51,** 491–530.
13. Stroud, R. M. & Finer-Moore, J. (1985) *Annu. Rev. Cell Biol.* **1,** 317–351.
14. Massoulie, J. & Bon, S. (1982) *Annu. Rev. Neurosci.* **5,** 57–106.
15. Slemmon, J. R., Salvaterra, P. M., Crawford, G. D. & Roberts, E. (1982) *J. Biol. Chem.* **257,** 3847–3852.
16. Wilbur, W. J. & Lipman, D. J. (1983) *Proc. Natl. Acad. Sci. USA* **80,** 726–730.
17. Gray, A., Cull, T. J. & Ullrich, A. (1983) *Nature (London)* **303,** 722–725.
18. Scott, J., Urdea, M., Quiroga, M., Sanchez-Pescador, R., Fong, N., Selby, M., Rutter, W. J. & Bell, G. I. (1983) *Science* **221,** 236–240.
19. Dayhoff, M. O., Barker, W. C. & Hunt, L. T. (1983) *Methods Enzymol.* **91,** 524–545.
20. Shepherd, J. C., McGinnis, W., Carrasco, A. E., DeRobertes, E. M. & Gehring, W. J. (1984) *Nature (London)* **310,** 70–71.
21. Kabsch, W. & Sander, C. (1984) *Proc. Natl. Acad. Sci. USA* **81,** 1075–1078.
22. Doolittle, R. F. (1981) *Science* **214,** 149–159.
23. Doolittle, R. F. (1985) *Sci. Am.* **253,** 88–99.
24. Hestrin, S. (1949) *J. Biol. Chem.* **180,** 879–881.
25. MacPhee-Quigley, K., Taylor, P. & Taylor, S. S. (1985) *J. Biol. Chem.* **260,** 12185–12189.
26. Finer-Moore, J. & Stroud, R. M. (1984) *Proc. Natl. Acad. Sci. USA* **81,** 155–159.
27. Kao, P. N., Cworle, A. J., Kaldany, R. R. J., Silver, M. L., Wideman, J., Stein, S. & Karlin, A. (1984) *J. Biol. Chem.* **259,** 11662–11665.
28. Kao, P. N. & Karlin, A. (1986) *J. Biol. Chem.* **261,** 8085–8088.
29. Anderson, D. J. & Blobel, G. (1981) *Proc. Natl. Acad. Sci. USA* **78,** 5598–5602.
30. Anderson, D. J. & Blobel, G. (1983) *Cold Spring Harbor Symp. Quant. Biol.* **48,** 125–134.
31. Mercken, L., Simons, M. J., Swillens, S., Massaer, M. & Vassart, G. (1985) *Nature (London)* **316,** 647–650.
32. Swillens, S., Landgate, M., Mercken, L., Dumont, J. E. & Vassart, G. (1986) *Biochem. Biophys. Res. Commun.* **137,** 142–148.
33. Futerman, A. H., Lowq, M. G., Michaelson, D. M. & Silman, I. (1985) *J. Neurochem.* **45,** 1487–1494.
34. Haas, R., Brandt, P. T., Knight, J. & Rosenberry, T. L. (1986) *Biochemistry* **25,** 3098–3105.
35. Kolata, B. (1985) *Science* **229,** 550.
36. Low, M. G., Ferguson, A. J., Futerman, A. H. & Silman, I. (1986) *Trends Biochem. Sci.* **11,** 212–215.
37. Musti, A. M., Avvedimento, E. V., Polistina, C., Ursini, V. M., Obici, S., Nitsch, L., Cocozza, S. & Di Lauro, R. (1986) *Proc. Natl. Acad. Sci. USA* **83,** 323–327.
38. Avvedimento, V. E., Musti, A. M., Obici, S., Cocozza, C. & DiLauro, R. (1984) *Nucleic Acids Res.* **12,** 3461–3472.
39. DiLauro, R., Obici, S., Condtiffe, D., Ursisni, M., Musti, A., Moscatelli, C. & Avvedimento, V. E. (1985) *Eur. J. Biochem.* **148,** 7–11.
40. Mercken, L., Simons, M. J., Demartynoff, G., Swillens, S. & Vassart, G. (1985) *Eur. J. Biochem.* **147,** 49–64.
41. Malthiery, Y. & Lissitzky, S. (1985) *Eur. J. Biochem.* **147,** 53–58.
42. Sudhoff, T. C., Golstein, J. L., Brown, M. S. & Russell, D. W. (1985) *Science* **228,** 815–822.
43. Sudhof, T. C., Russell, D. W., Goldstein, J. L., Brown, M. S., Sanchez-Pescador, R. & Bell, G. I. (1985) *Science* **228,** 893–895.
44. Ny, T., Elgh, F. & Lund, B. (1984) *Proc. Natl. Acad. Sci. USA* **81,** 5355–5359.
45. Nagamine, Y., Pearson, D., Altus, M. & Reich, E. (1984) *Nucleic Acids Res.* **12,** 9525–9541.
46. Perler, R., Efstratiadis, A., Lomedico, P., Gilbere, W., Kolodner, R. & Dodgson, J. (1980) *Cell* **20,** 555–566.
47. Craik, C. S., Rutter, W. J. & Filetterick, R. (1983) *Science* **220,** 1125–1129.
48. McLaughlin, P. J. & Dayhoff, M. O. (1972) in *Atlas of Protein Sequence and Structure*, ed. Dayhoff, M. O. (Natl. Biomed. Res. Found., Washington, DC), Vol. 5, pp. 47–52.
49. Gilbert, W. (1978) *Nature (London)* **271,** 501.
50. Gilbert, W. (1985) *Science* **228,** 823–824.
51. Doolittle, R. F. (1985) *Trends Biochem. Sci.* **10,** 223–237.