Volume 282, number 1, 17–22 FEBS 09610 April 1991

ADONIS 001457939100319?

Evolutionary origin of thyroglobulin by duplication of esterase gene

Yasumitsu Takagi[1,*], Tsuneo Omura[1] and Mitiko Gō[2]

[1] *Department of Molecular Biology, Graduate School of Medical Science, Kyushu University* and [2] *Department of Biology, Faculty of Science, Nagoya University, Japan*

Received 4 February 1991

Rat liver microsomal carboxyesterase E1 was found to have homology with five esterases and with the COOH-terminal parts of two thyroglobulins. A phylogenetic tree constructed for these proteins shows that this new superfamily has evolved from a common ancestral gene that encoded a carboxyesterase. The tree also shows that the ancestral gene already existed before the divergence of vertebrates and invertebrates and later its duplicated genes gained various kinds of esterase activity. According to the tree, one of the duplicated genes evolved into the COOH-terminal half of thyroglobulin by a gene fusion with a DNA sequence whose evolutionary origin is unknown.

Molecular evolution; Esterase; Thyroglobulin; Gene duplication; Gene fusion

1. INTRODUCTION

Carboxyesterase in mammalian liver hydrolyzes xenobiotics containing ester, thioester or amino bonds. Therefore, its main physiological role is thought to be in drug metabolism and detoxication of many harmful chemicals in our environment [1]. Carboxyesterase E1 of rat liver is one of the carboxyesterases present exclusively in the lumen of the endoplasmic reticulum [2]. To identify the amino acid sequence of rat carboxyesterase E1 responsible for its retention in the endoplasmic reticulum, we determined the primary structure of cDNA clones of this enzyme [3].

The physiological function of carboxyesterase E1 is not clear, but some clues on its function may be obtained from its evolutionary origin deduced from a phylogenetic tree of sequences related to those of carboxyesterase E1. Therefore, we searched for sequences homologous with that of carboxyesterase E1. More than 10 eukaryotic esterases were included in the data base used, but only 5 of them showed significant homology with carboxyesterase E1. Carboxyesterase E1 also showed homology with the COOH-terminal half of thyroglobulin.

Based on the comparisons among the primary structures of these esterases and some serine proteases, Myers et al. proposed that the catalytic mechanism these esterases use, is the same Arg-Asp-Ser or His-Asp-Ser charge relay [4]. In this paper, we propose the evolutionary origin of thyroglobulin and a group of eukaryotic esterases, and show that an ancestral carboxyesterase evolved into various proteins with related or completely different functions by a series of gene duplications and by gene fusion. We also suggest that the COOH-terminal part of thyroglobulin may have had, or still has some other unknown function than being a precursor of the thyroid hormone.

Correspondence address: T. Omura, Department of Molecular Biology, Graduate School of Medical Science, Kyushu University, Fukuoka 812, Japan

**Present address:* Department of Biochemistry, School of Medicine, Keio University, Japan

2. MATERIALS AND METHODS

Seven proteins with homologous sequences to that of rat carboxyesterase E1 were identified in a sequence data base containing EMBL, GENBANK, and NBFR by use of the method of Lipman and Pearson [5]. Alignments for some pairs out of the 8 sequences were obtained and then the full sequences were aligned by inspection of pair-wise alignments introducing the minimum number of insertions and deletions to obtain the maximum number of identical amino acid residues among the 8 proteins (Fig. 1).

The evolutionary distance D between each pair out of the 8 sequences was calculated by Jukes and Cantor's formulation [6]. The standard error of the distance was calculated according to the formulation of Kumura and Ohta [7]. A phylogenetic tree for the 8 proteins was constructed by the NJ method [8] and by a modified UPGMA method [9,10], where no assumption of a constant evolutionary rate was introduced.

3. RESULTS

We aligned the sequences of 8 proteins as described in section 2 (Fig. 1). Table I shows the sequence identities of all pairs of these 8 proteins. The identity between the rat and rabbit carboxyesterases is 65.6%; this value shows that this family has evolved more rapidly than the globin family having about average evolutionary speed among various proteins [11]. As can be seen, the identity varies from 17.5% to 79.1%. An identity of more than 20% is usually required for the statistical significance, i.e. to conclude that two sequences are

derived from a common ancestor [12]. The identities between *Drosophila melanogaster* esterase-6 and all other sequences, except human cholinesterase and *Torpedo californica* acetylcholinesterase, are critical to be concluded as significant. However, the identity between *D. melanogaster* esterase-6 and rat thyroglobulin is 28% between residues 22 and 126 in the numbering system of rat carboxyesterase E1. The functional similarity between those esterases, including *D. melanogaster* esterase-6 suggests that these enzymes were derived from a common ancestor long ago and this family evolved with high speed, so their overall identity is low. The identities of *D. melanogaster* esterase-6 with *T. californica* acetylcholinesterase and human cholinesterase are slightly higher, i.e. 22.8% and 23.7%. The identity matrix (Table I) as well as the functional resemblance of the sequences of these proteins shows that they belong to a same superfamily and that *D. melanogaster* esterase-6 is distantly related to the other sequences.

```
                                                          *50
rat  cest    hpssppvvdttk---------gkvlgkyvs--legftqpvavflgvpfakpplgslrfappepae-pwsfvkntt
rab  cest    hpsappvvdtvk---------gkvlgkfvs--legfaqpvavflgvpfakpplgslrfappqpae-swshvkntt
D.m. est6    sdtddpllvqlpq--------gklrg-r----dngsyys---yesipyaepptgdlrfeapepykqkwsdifdat
hum  Chest   --eddiiiatkn---------gkvrgmnlt--vfggtvt-a-flgipyaqpplgrlrfkkpqslt-kwsdiwnat
T.c. AChest  ---ddhsellvntks------gkvmgtr----vpvlsshisaflgipfaeppvgnmrfrrpepkkp-wsgvwnas
D.m. AChest  --vidrlvvqtss--------gpvrgrs----vtvqgrevhvytgipyakppvedlrfrkpvpaep-whgvldat
bov  Tg      kpniplpgfgtsspsvplathgqllgrsqaiqvgtswkpvdqflgvpyaapplgekrfrapehln--wtgswent
rat  Tg      -gaplhqsdgistpsvhldsfgqlqggsqvvkvgtawkqvyqflgvpyaapplaenrfqapevln--wtgswdat

                                  *100                                        150*
rat  cest    typpmcsqdgvvgklladmlstgkesiplefsedclylniyspadltknsr------------------------
rab  cest    syppmcssdavsghmlselftnrkeniplkfsedclylniytpadltkrgr------------------------
D.m. est6    ktpvaclq---wdqftpgan-------klvgeedcltvsvykpkn-skrnsfp----------------------
hum  Chest   kyansCcq--nidqsfpgfhgsemwnpntdlsedClylnvwipapkpknat------------------------
T.c. AChest  typnnCqqyvde-qf-pgfsgsemwnpnremsedClylniwvpsprpkstt------------------------
D.m. AChest  glsatcvq--eryeyfpgfsgeeiwnpntnvsedclyinvwapakarlrhgrganggehpngkqadtdhlihngn
bov  Tg      kprarcwqpgirtptppgvs------------edclylnvfvpqnmapnas------------------------
rat  Tg      klrsscwqpgtrtptppqis------------edclylnvfvpenlvsnas------------------------

                                                          *200
rat  cest    -------lpvmvwihggglliggasp--ysglalsahenvvvvtiqyrlgiwglfstgde-------hsrgnwah
rab  cest    -------lpvmvwihggglmvggast--ydglalsahenvvvvtiqyrlgiggfgfnidel--------------
D.m. est6    ---------vvahihggafmfgaaw--qnghenvmregkfilvkisyrlgplgfv-stgdr------dlpgnygl
hum  Chest   ---------vliwiygggfqtgtsslhvydgkflarvervivvsmnyrvgalgflalpgnp------eapgnmgl
T.c. AChest  ---------vmvwiygggfysgsstldvyngkylayteevvlvslsyrvgafgflalhgsq------eapgnvgl
D.m. AChest  pqnttnglpiliwiygggfmtgsatldiynadimaavgnvivasfqyrvgafgflhlapempsefaeeapgnvgl
bov  Tg      ---------vlvffhnaaegkgsgdrpavdgsflaavgnlivvtasyrtgifgflssgss-------elsgnwgl
rat  Tg      ---------vlvffhntvemegsggqlnidgsilaavgnlivvtanyrlgvfgflssgsd-------evagnwgl

                                  *250                                        300*
rat  cest    ldqlaalrwvqdnianfggnpdsvtifgesaggvsvsalvlspl--aknlfhraisesgv-vltnldkkntqav
rab  cest    -flvavnrwvqdnianfggdpgsvtifgeSaggqsvsillispl--tknlfhraisesgvallsslfrkn-tksl
D.m. est6    kdqrlalkwikqniasfggepqnvllvghsaggasvhlqmlred--fgqlaraafsfsgnaldpwviqkgargrn
hum  Chest   fdqqlalqwvqkniaafggnpksvtlfgeSagaasvslhllspg--shslftrailqsgsfnapwavtslyearn
T.c. AChest  ldqrmalqwvhdniqffggdpktvtifgeSaggasvgmhilspg--srdlfrrailqsgspncpwasvsvaegrr
D.m. AChest  wdqalairwlkdnahafggnpewmtlfgesagsssvnaqlmspv--trglvkrgmmqsgtmnapwshmtsekave
'bov Tg      ldqvvaltwvqthiqafggdprrvtlaadrggadiasihlvttraansrlfrravlmggsalspaavlrperarq
rat  Tg      ldqvaaltwvqthigafggdpqrvtlaadrggadvasihllitrptrlqlfrkallmggsalspaailspdraqq

                                                          *350
rat  cest    aqm-iatlsgcnntssaamvq----clrqkteaelleltvkldntsms-------------tvldgvvlpktpee
rab  cest    aek-laleagckttsavmvh----clrqkteeelmevtlkmkfmaldlvgdpkentafltvidgvllpkapae
D.m. est6    f--elgrnvgcesaedstslkk---clkskpaselvtavrkflifsyvpfapfspvle---psdapdaiitqdpr
hum  Chest   rtlnlakltgCsrenetelik----Clrnkdpqeillneafvvpygtplsvnfg-------ptvdgdfltdmpdl
T.c. AChest  ravelgrnlnCnlnsdeelih----Clrekkpqelidvewnvlpfdsifrfsfv-------pvidgeffpt-sle
D.m. AChest  lgkallndcncnasmlktnpahvmscmrsvdaktisvqqwnsysgilsfpsa---------ptidgaflpadpm-
bov  Tg      qaaalakevgcpsssvqemvs----clrqeparilndaqtkllavsgpfhywg--------pvvdgqylre-tpa
rat  Tg      qaaalakevgcpnssvqevvs----cfrqkpanilneaqtkllavsgpfhywg--------pvvdgqylre-lps
```

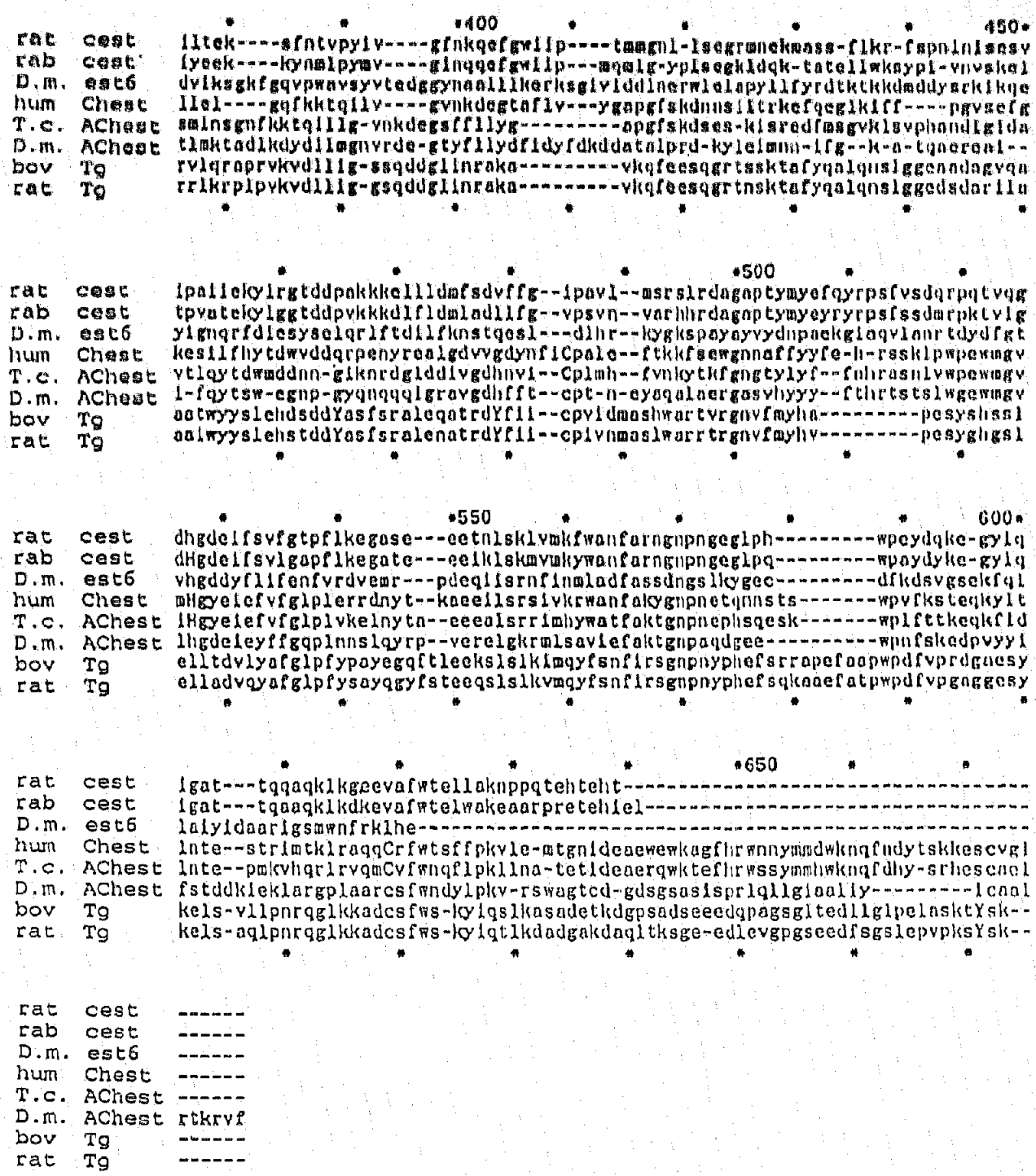

Fig. 1. Alignment of 6 esterases indicated as rat cest, rab cest, D.m. est6, hum Chest, T.c. AChest, D.m. AChest, (rat carboxyesterase E1 [3], rabbit carboxyesterase [22], *Drosophila melanogaster* esterase-6 [13], human cholinesterase [23], *Torpedo californica* [14] and *D. melanogaster* [24] acetylcholinesterases), and COOH-terminal portions of two thyroglobulins indicated as bov Tg and rat Tg (amino acid residues 2168 through 2750 of bovine thyroglobulin [25] and 387 through 967 of rat thyroglobulin [26]. Each sequence of esterase starts at the NH_2-terminal residue of the mature enzyme without a signal sequence. The numbering system of the residues follows the one in rat carboxyesterase E1 [3]. As the cleavage sites of the two esterases from *D. melanogaster* are not known, we predicted the processing points according to the general structure of the signal sequences [27]. Amino acid residues are denoted by single letters in small print except for the following capital letters: C, cysteine residues that participate intramolecular disulfide bonds in *T. californica* acetylcholinesterase and human cholinesterase (positions 81, 110, 311, 326, 485 and 619) [28,29]: Y, tyrosine residues that are iodinated in thyroglobulin (positions 465, 479 and 671) [25,26]: S and H, active site serine and histidine residues of several esterases (positions 255 and 527) [3].

Table 1

Sequence identities (%) for each pair of the 8 proteins compared*

	Rat cest	Rab cest	D.m. est6	Hum Chest	T.c. AChest	D.m. AChest	Bov Tg
Rab cest	65.6						
D.m. est6	18.7	17.5					
Hum Chest	31.0	28.8	22.8				
T.c. AChest	30.2	29.0	23.7	44.3			
D.m. AChest	23.8	22.8	18.6	31.9	35.5		
Bov Tg	23.2	21.3	18.9	25.6	27.6	22.7	
Rat Tg	24.1	22.4	19.2	25.1	25.5	22.6	79.1

*See legend to Fig. 1 for the abbreviations used in this Table.

To see whether the identities were limited to some particular regions, we examined the extents of conservation of amino acid residues at all positions aligned of 7 proteins, excluding esterase-6, which has the lowest identity to carboxyesterase E1. In Fig. 2, however, the positions, where all of the 8 proteins have the same amino acid residues, are indicated by arrows. Positions where at least one gap appeared in the alignment are denoted as having a zero degree of conservation. This diagram reveals that invariant residues are mainly located in the NH_2-terminal halves of the molecules, although regions with a high degree of conservation are found also in their COOH-terminal halves.

A phylogenetic tree of the esterase superfamily was constructed from the evolutionary distances D calculated from the identity matrix (Table 1) as described in section 2. The two different methods used gave essentially the same phylogenetic tree. Therefore, only the tree constructed by the NJ method is shown in Fig. 3. This phylogenetic tree shows that these 8 sequences can be roughly divided into 3 groups (esterase-6,thyroglobulins and other esterases) as a result of gene duplications. The common ancestral gene existed before the divergence of vertebrates and insects, because the both duplications occurred before the divergence of *T. californica* and *D. melanogaster* acetylcholinesterases. Since esterase-6 is a carboxyesterase [13], the ancestral gene had probably the function of a carboxyesterase. One of the groups generated by this gene duplication was the direct ancestor of the gene encoding the COOH-terminal part of thyroglobulin and the ancestral gene of the other esterases than esterase-6. Another duplication leading to choline-, acetylcholine-, and carboxyesterases occurred later. The lineages of choline- and acetylcholine-esterases branched within a very short time period. Therefore their relationship in this tree may not reflect the true evolutionary order.

4. DISCUSSION

Homology between the esterases and thyroglobulins was first discovered by comparison of the primary structures of *T. californica* acetylcholinesterase and

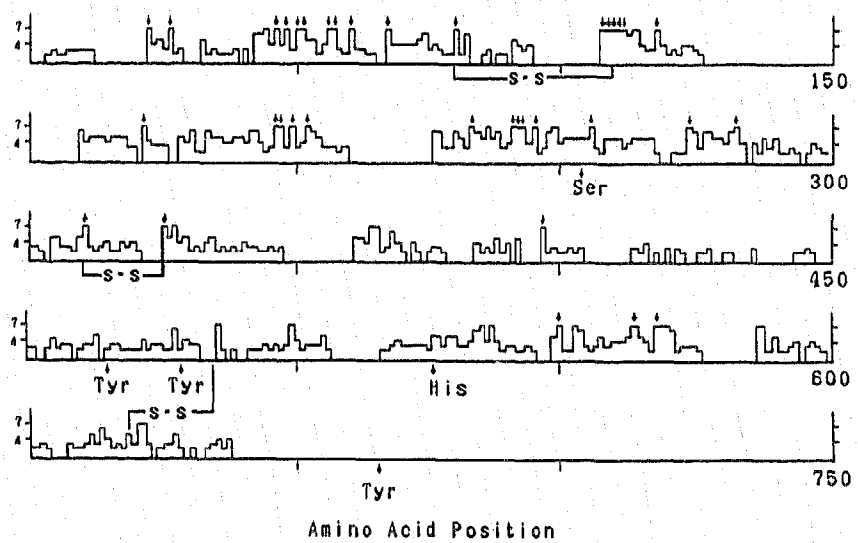

Fig. 2. Degrees of amino acid conservation at each site in 7 proteins excluding esterase-6. Arrows indicate the sites where the residues are conserved in all 8 proteins aligned (Fig. 1). S-S shows an intramolecular disulfide bond as described in the legend to Fig. 1. Tyr, Ser and His are the same as Y, S and H, respectively, explained in the legend to Fig. 1.

bovine thyroglobulin [14], and these two proteins were suggested to have a common evolutionary origin [15,16].

Swillens et al. [15] detected the sequence homology between the COOH-terminal portions of bovine thyroglobulin and *T. californica* acetylcholinesterase and analyzed their hydropathy profiles. Through these observations, they suggested a similar three-dimensional structure for these sequences. But in the COOH-terminal halves of the proteins, sequence identities are lower than those in the NH_2-terminal halves (Fig. 2). It remains to be investigated whether these 8 proteins have a common conformation in the regions showing low identity. Mori et al. [16] analyzed the homology between rat thyroglobulin and *T. californica* acetylcholinesterase and suggested that the gene encoding the COOH-terminal part of the thyroglobulin had evolved from a redundant copy of the acetylcholinesterase gene. The relationship shown in Fig. 3 indicates, however, that both acetylcholinesterase and thyroglobulin evolved from a common ancestral gene that encoded carboxyesterase.

Comparative endocrinological studies have shown that the most primitive animal that uses a thyroglobulin-like, high molecular weight iodoprotein as a precursor of thyroid hormone is the lamprey [17,18]. The lamprey belongs to the agnatha, the oldest class of vertebrates. It is also known that acetylcholinesterase is much more widely distributed in organisms than thyroglobulin [19]. On the other hand, the present study showed that the ancestral gene of the COOH-terminal part of thyroglobulin evolved by a gene duplication from an esterase-like ancestor, and

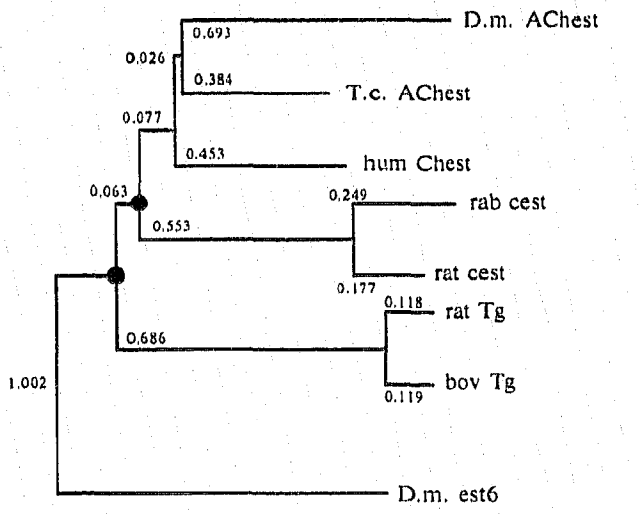

Fig. 3. Phylogenetic tree for the 6 esterases and the COOH-portion of the two thyroglobulins constructed by the NJ method. The 8 sequences are indicated as described in the legend to Fig. 1. The values next to a branch estimate the branch length in numbers of amino acid substitutions per site. At least two possible gene duplication sites are shown by the filled circles.

that the gene of the acetylcholinesterase evolved later by another gene duplication. Our findings also indicate that at least the COOH-terminal part of the thyroglobulin originated before the divergence of vertebrates and invertebrates. Later, this gene fused with the ancestral gene encoding the rest of the present thyroglobulin.

The fact that the lamprey appeared at least 100 million years after the divergence of vertebrates and invertebrates indicates that the COOH-terminal part of thyroglobulin may have had some physiological function(s) other than as a precursor of thyroid homrone. However, as we constructed the phylogenetic tree without assuming a constant evolutionary rate, we can not estimate the time of gene fusion from the tree. A peptide corresponding to the COOH-terminal part of thyroglobulin may be expressed in modern eukaryotes lower than the lamprey.

The hagfish, a salt-water agnatha, produces a much smaller iodoprotein than that of the lamprey as a precursor of thyroid hormone. Recently, Ohmiya et al. [20] isolated a iodoprotein of hagfish thyroid and showed that it differs from thyroglobulin. This structural difference between the iodoproteins of hagfish and lamprey supports our hypothesis that the ancestor of the COOH-terminal part of thyroglobulin, which might have had some other unknown function, was recruited as a precursor of thyroid hormone at almost the same time as the lamprey, a fresh-water agnatha diverged from hagfish, a salt-water agnatha. However, since no structural information is available about the precursor of thyroid hormone of lamprey, there is a possibility that this precursor has a different ancestor from that of higher vertebrate thyroglobulin in spite of their similar biochemical properties. An example of a protein having a similar function but different ancestors is observed in microbial ribonucleases and mammalian ones [21]. Determination of the primary structures of the thyroid hormone precursors both of hagfish and lamprey, and comparison with those of higher vertebrates will provide further clues to understand the origin of thyroid hormones.

Acknowledgements: The authors thank Professor Y. Kondo and Dr. F. Tajima for valuable discussions and critical reading of this manuscript. They also thank Professor T. Gojobori and Dr. N. Saito for supplying a computer program for making an evolutionary tree and Dr. K. Fukami-Kobayashi for drawing a phylogenetic tree. This work was supported in part by a Grant-in-Aid for Scientific Research and a Grant-in-Aid for Special Project Research from the Ministry of Education, Science and Culture of Japan.

REFERENCES

[1] Satoh, T. (1987) Rev. Biochem. Toxicol. 8, 155–181.
[2] Harano, T., Miyata, T., Lee, S., Aoyagi, H. and Omura, T. (1988) J. Biochem. 103, 149–155.
[3] Takagi, Y., Morohashi, K., Kawabata, S., Gō, M. and Omura, T. (1988) J. Biochem. 104, 801–806.

[4] Myers, M., Richimond, R.C. and Oakeshott, J.G. (1988) Mol. Biol. Evol. 5, 113–119.
[5] Lipman, D.J. and Pearson, W.R. (1985) Science 227, 1435–1441.
[6] Jukes, T.H. and Cantor, C.R. (1969) in: Mammalian Protein Metabolism (Munro, H.N., ed.) pp. 21–132, Academic Press, New York.
[7] Kimura, M. and Ohta, T. (1972) J. Mol. Evol. 2, 87–90.
[8] Saitou, N. and Nei, M. (1987) Mol. Biol. Evol. 4, 406–425.
[9] Tajima, F. and Nei, M. (1984) J. Mol. Evol. 1, 269–285.
[10] Lee, W.H. (1981) Proc. Natl. Acad. Sci. USA 78, 1085–1089.
[11] Wilson, A.C., Carlson, S.S. and White, T.J. (1977) Annu. Rev. Biochem. 46, 573–639.
[12] Doolittle, R.F. (1981) Science 214, 149–159.
[13] Oakeshott, J.G., Collet, C., Phillis, R.W., Nielsen, K.M., Russell, R.J., Chambers, G.K., Ross, V. and Richmond, R.C. (1987) Proc. Natl. Acad. Sci. USA 84, 3359–3363.
[14] Schumacher, M., Camp, S., Maulet, Y., Newton, M., MacPhee-Quigley, K., Taylor, S.S., Friedmann, T. and Taylor, P. (1986) Nature 319, 407–409.
[15] Swillens, S., Ludgate, M., Mercken, L., Dumont, J.E. and Vassart, G. (1986) Biochem. Biophys. Res. Commun. 137, 142–148.
[16] Mori, N., Itoh, N. and Salvaterra, P.M. (1987) Proc. Natl. Acad. Sci. USA 84, 2813–2817.
[17] Suzuki, S. and Kondo, Y. (1973) Gen. Comp. Endocrinol. 21, 451–460.
[18] Aloj, S., Salvatore, G. and Roche, J. (1967) J. Biol. Chem. 242, 810–3814.
[19] Prosser, C.L. and Brown, Jr., F.A. (1962) In: Comparative Animal Physiology, 2nd edn., pp. 607–613, Saunders, Philadelphia.
[20] Ohmiya, Y., Suzuki, S. and Kondo, Y. (1989) Eur. J. Biochem. 182, 11–18.
[21] Nakamura, K.T., Iwahashi, K., Yamamoto, Y., Iitaka, Y., Yoshida, N. and Mitui, Y. (1982) Nature 299, 564–566.
[22] Korza, G. and Ozols, J. (1988) J. Biol. Chem. 263, 3486–3495.
[23] Lockridge, O., Bartels, C.F., Vaughan, T.A., Wong, C., Norton, S.E. and Johnson, L.L. (1987) J. Biol. Chem. 262, 549–557.
[24] Hall, L.M.C. and Spierer, P. (1986) EMBO J. 5, 2949–2954.
[25] Mercken, L., Simons, M., Swillens, S., Massaer, M. and Vassart, G. (1985) Nature 316, 647–651.
[26] Di Lauro, R., Obiel, S., Condliffe, E., Ursini, M.V., Mutti, A., Moscatelli, C. and Avvedimento, V.E. (1985) Eur. J. Biochem. 148, 7–11.
[27] Perlman, D. and Halvorson, H.O. (1983) J. Mol. Biol. 167, 391–409.
[28] MacPhee-Quigley, K., Vedvick, T.S., Taylor, P. and Taylor, S.S. (1986) J. Biol. Chem. 261, 13565–13570.
[29] Lockridge, O., Adkins, S. and LaDue, N. (1987) J. Biol. Chem. 262, 12945–12952.