Vol. 137, No. 1, 1986 BIOCHEMICAL AND BIOPHYSICAL RESEARCH COMMUNICATIONS
May 29, 1986 Pages 142-148

ANALYSIS OF SEQUENCE AND STRUCTURE HOMOLOGIES BETWEEN THYROGLOBULIN AND ACETYLCHOLINESTERASE : POSSIBLE FUNCTIONAL AND CLINICAL SIGNIFICANCE

Stéphane Swillens[1], Marian Ludgate[1], Luc Mercken[1,*], Jacques E. Dumont[1] and Gilbert Vassart[1,2]

[1]Institut de Recherche Interdisciplinaire and [2]Service de Chimie
Université Libre de Bruxelles, Campus Erasme
route de Lennik 808, B-1070 Brussels BELGIUM

Received April 7, 1986

SUMMARY : The homology between thyroglobulin and acetylcholinesterase (1) has been analyzed in detail. It contains 28.3% identical amino acids and extends over 544 residues, involving more than 90% of the acetylcholinesterase molecule and the C-terminal portion of thyroglobulin. The hydropathy profiles of the homologous regions have been determined and compared. Their striking resemblance suggests that both proteins adopt a similar three dimensional structure and militates for some common property. As thyroglobulin and acetylcholinesterase are known to interact with cell membranes, we suggest that the acetylcholinesterase-like domain of thyroglobulin is involved in the binding. These observations demonstrate that thyroglobulin has evolved from the condensation of a duplicated copy of the acetylcholinesterase gene with an archaic thyroglobulin gene encoding the major hormonogenic domain. The extensive homology in hydropathy profiles suggests that the two proteins may share antigenic determinants. If this were the case, it would provide a rationale for the demonstration of immunoreactive thyroglobulin in neurons (2) and the pathogenesis of Grave's ophtalmopathy.

A significant sequence homology has recently been reported between ACHE from Torpedo californica (1) and bovine Tg (3). As these proteins do not share any known common function, this finding came as a surprise. Indeed, ACHE is an essential enzyme of the nervous system playing a key role in cholinergic synapses and motor endplates while Tg is synthesized only in the thyroid of chordates where it serves as the precursor of thyroid hormones (see (4) for a review). ACHE exists as an asymmetric form associated with structural proteins and as a globular dimer (2 x 68000 dalton) (5,6). Whether these forms are encoded by separate genes is still a debated question (1,7). Tg is a dimeric protein made of identical 2750 residue protomers (3,4). It is

*Present address: Unité de Recherches INSERM U-99, Hopital Henri Mondor, F94010 Créteil, France.

Abbreviations: Tg, thyroglobulin; ACHE, acetylcholinesterase.

0006-291X/86 $1.50

encoded by one of the largest gene unit (about 200 kb) described to date (8-10).

In the present study, we have analyzed in detail the sequence homology between the two proteins. From comparison of hydropathy profiles, we present evidence that they share extensive structural properties. As an important characteristic of both proteins is to interact with cell membrane, we hypothesize that the AChE-like domain of Tg is responsible for the membrane binding and may play a role in membrane signalling, e.g. in the endocytosis of the protein by the thyrocyte. In addition, the similarity in hydropathy profiles suggests that both proteins may share antigenic determinants. We hypothesize that this characteristic may provide an explanation for the pathogenesis of Grave's ophtalmopathy.

MATERIALS AND METHODS

Methods are described in the figure legends.

RESULTS

Fig. 1 displays the best alignment between Tg and AChE sequences as obtained by the computer procedure of Lipman and Pearson (11). It is striking that the homology involves most of AChE sequence and the whole carboxyl terminal portion of Tg. The homology, long of 544 residues, contains 28.3 % identical residues and 15 regions totalling 34 deleted residues (19 and 15 in Tg and AChE respectively). The optimized homology score is 727 (11). When the relevant segment of Tg is compared to 100 random sequences containing the same amino acid composition as AChE, the average optimized score is 40.4 ± 7.65 (s.d.). The score for the Tg-AChE comparison is thus situated at 90 s.d. from the mean which is considered highly significant (11). When the extent of local homology is plotted along the protein sequences using scores computed according to the PAM250 matrix (12) over 7 residue pairs (upper panel of fig. 2), it is obvious that the homology involves the entire AChE sequence, except its termini and four small segments centered at positions 119, 335,

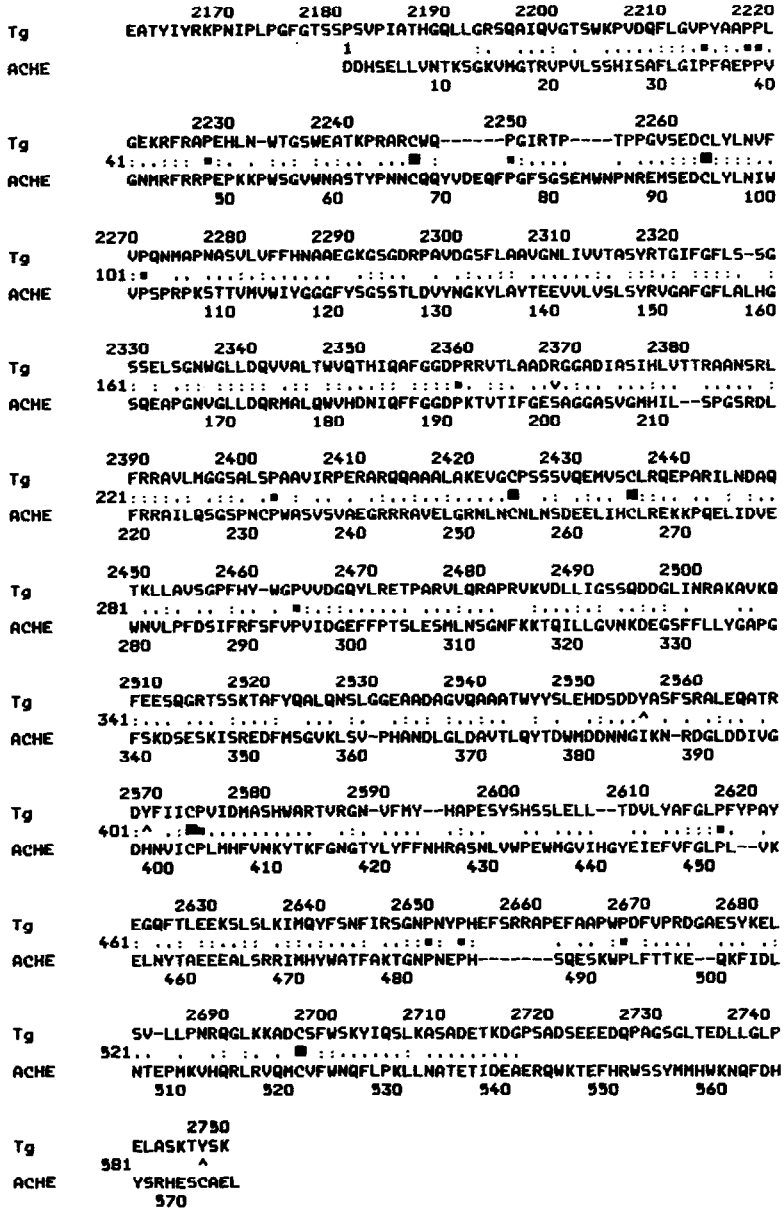

Fig. 1.
Alignment of Tg and ACHE as provided by the computer program FASTP (11). The homology extends from position 13 to position 556 of the alignment, which correspond to positions 2193 and 2717 for Tg and to positions 13 and 541 for ACHE. Colons and dots respectively refer to pairs of identical residues and residue pairs characterized by a conservative replacement (positive score in the PAM250 matrix (12)). Homologous prolines and cysteines are indicated, as well as the residues implicated in the function of the proteins.

437 and 500. Most regions of high homology contain homologous proline or cysteine residues. As already mentionned by Schumacher et al. (1), 6 cysteine residues are located at homologous positions. However, key residues involved

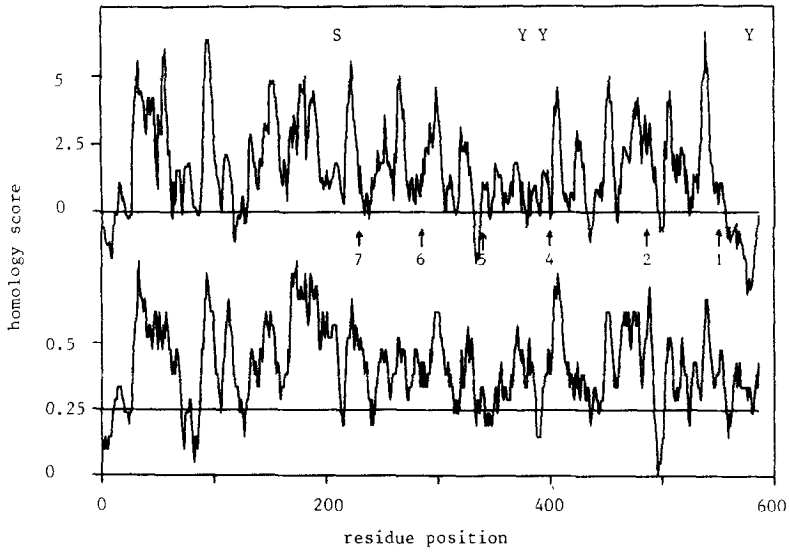

Fig. 2.
Local homology profiles obtained with the protein sequences (upper panel) and the nucleic acid sequences (lower panel). For the protein sequence, the mean score over 7 residues is computed using the PAM250 matrix. Residue deletions have been scored to 0. The horizontal bar (ordinate 0) represents the lower score limit for conservative residue replacement. Positions of the residues involved in the function of the proteins (serine for AChE and tyrosine for Tg) and of the intron-exon junctions are indicated. For the nucleic acid sequences, the mean score over 21 bases is computed considering a score of 1 for a match and a score of 0 for a mismatch or a base deletion. The horizontal bar represents the lower limit for a significant local homology.

in the specialized functions of each protein are not conserved : the three hormonogenic tyrosines identified in this region of Tg (3) have no counterpart in AChE, and the serine within the active center of AChE (7) is absent in the corresponding segment of Tg. Moreover, the environment of these residues is poorly conserved (see fig. 1). With the exception of the second intron located at position 486 of the alignment, the available positions of introns in rat Tg sequence (13) tend to coincide with regions of lower homology in agreement with the theory of Craik et al.(14).

Nucleic acid sequences have been compared on the basis of the protein alignment (lower panel of fig. 2). Using scores of +1 and 0 for match and mismatch respectively and a gap penalty of -1, a score of 630 was obtained, placing the homology at 20 s.d. from the mean obtained with random sequences according to the method of Reich et al.(15). The distribution of base sequence homologies along the cDNA is very similar to that observed with

amino acids (fig. 2). The correlation between the two profiles is characterized by a coefficient of 0.75.

Obviously, the high sequence homology is not associated with the conservation of the specific protein functions. On the other hand, the environment of key residues implicated in the protein structure seems to be conserved. In order to substantiate this observation, the hydropathy profiles of both proteins were determined using the consensus scale of Eisenberg (16). As illustrated on fig. 3, the two proteins display a striking similarity in their hydropathy profiles. The correlation coefficient computed between both profiles is 0.55 which is considered highly significant (17). Interestingly, the peaks of homology correspond, on average, to hydrophobic regions. Because proline and cysteine residues have no marked hydrophobic character, this apparent correlation requires the presence of homologous hydrophobic residues in the near vicinity. Only two regions of high homology are essentially hydrophilic, namely a part of the first homology peak (around position 45) and the whole peak located at position 222. In both cases, the hydrophilic character is mainly due to the presence of two homologous arginine residues.

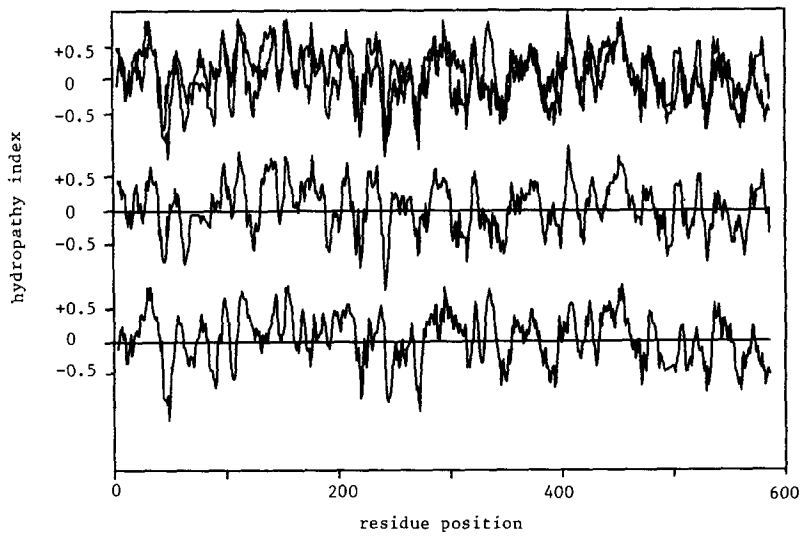

Fig. 3.
Hydropathy profiles are obtained on the basis of the consensus scale of Eisenberg (16). The index refers to an average score over 7 residues. Bottom : ACHE profile; middle : Tg profile; top : the two profiles are superimposed.

DISCUSSION

Considering that comparison of hydropathy profiles is reportedly able to detect three dimensional structural homologies in families of protein containing few identical residues (17), the present results are strong indication that the C-terminal domain of Tg and ACHE share more than a common ancestor. They most probably adopt similar overall structure.

As stated by Schumacher et al.(1), at first sight, Tg and ACHE seem to "have in common only the functional property of being secreted from specialized cells". Considering the surprisingly high sequence and structural homologies between the two proteins, we would favor the concept that some common selective pressure has been exerted on both genes since the duplication event. ACHE is known to interact with membranes without being an integral membrane protein (6,18). Similarily, Tg has been shown to bind to the apical membrane of the thyrocyte. It has been proposed that this binding plays an essential role in the selectivity of the endocytotic process leading to thyroid hormone secretion (19). From its vital role in the transmission of nerve influx, ACHE function most probably antedates the evolution of functional Tg. We suggest that the Tg gene has evolved from the shuffling of two preexisting gene units. An archaic Tg, on one hand, consisting mostly of tandemly repeated domains and containing the major hormonogenic site near the amino terminus (3,4,9), together with a genuine copy of the ACHE gene. We hypothesize that this newly acquired C-terminal domain is involved in the interaction of Tg with the apical membrane of the thyrocyte. It is tempting to speculate that the minor hormonogenic sites present in this domain might play the role of sensors, e.g. signalling whether the iodinated Tg is ready for endocytosis.

The homology of hydropathy profiles and, in particular the existence of similar hydrophilic domains (positions 45 and 222), suggest that Tg and ACHE might share common antigenic determinants. This may provide a rationale to the demonstration of "immunoreactive Tg" in neural cells of annelids (2) and, more importantly, to the pathogenesis of Grave's ophtalmopathy. Indeed,

autoimmune processes involving Tg-related antigenes have been implicated in this disease (20,21). It is conceivable that humoral and/or cellular immunity reactions elicited by Tg epitopes would recognize AChE on the highly innervated and thyroid hormone stimulated ocular muscles and provoke the disease. If this hypothesis is correct, other neuromuscular junctions could be involved, although less severely.

ACKNOWLEDGEMENTS

We thank Michel Georges for fruitful discussions. This study has benefitted from grants of the following organisms : Actions Concertées du Ministère de la Politique Scientifique, NIH (21732)and ARBD asbl.

REFERENCES

1. Schumacher, M., Camp, S., Maulet, Y., Newton, M., MacPhee-Quigley, K., Taylor, S.S., Friedmann, T., and Taylor, P. (1986) Nature 319, 407-409.
2. Marcheggiano, A., Iannoni, A., and Davoli, C. (1985) Cell Tissue Res. 241, 429-433.
3. Mercken, L., Simons, M.J., Swillens, S., Massaer, M., and Vassart, G. (1985) Nature 316, 647-651.
4. Vassart, G., Bacolla, A., Brocas, H., Christophe, D., de Martynoff, G., Leriche, A., Mercken, L., Parma, J., Pohl, V., Targovnik, H., and Van Heuverswyn, B. (1985) Mol. Cell. Endocrin. 40, 89-97.
5. Lee, S.L., and Taylor, P. (1982) J. Biol.Chem. 257, 12292-12301.
6. Lee, S.L., Camp, S.J., and Taylor, P. (1982) J. Biol. Chem. 257, 12302-12309.
7. MacPhee-Quigley, K., Taylor, P., and Taylor, S. (1985) J. Biol. Chem. 260, 12185-12189.
8. Van Ommen, G.J.B., Arnberg, A.C., Baas, F., Brocas, H., Sterk, A., Tegelaers, W.H., Vassart, G., and De Vijlder, J.J.M. (1983) Nucleic Acids Res. 11, 2273-2285.
9. Musti, A.M., Avvedimento, E.V., Polistina, C., Ursini, V.M., Obici, S., Nitsch, L., Cocozza, S., and Di Lauro, R. (1986) Proc. Natl. Acad. Sci. USA 83, 323-327.
10. de Martynoff, G. (1984) Ann. Endoc. (Paris) 45, 63.
11. Lipman, D.J., and Pearson, W.R. (1985) Science 227, 1435-1441.
12. Dayhoff, M. (1978) Atlas of Protein Sequence and Structure 5 suppl. 3, National Biomedical Research Foundation, Silver Spring.
13. Avvedimento, V.E., Musti, A.M., Obici, S., Cocozza, S., and Di Lauro, R. (1984) Nucleic Acids Res. 12, 3461-3473.
14. Craik, C.S., Rutter, W.J., and Fletterick, R. (1983) Science 200, 1125-1129.
15. Reich, J.G., Drabsch, H., and Daumler, A. (1984) Nucleic Acids Res. 13, 5529-5543.
16. Eisenberg, D. (1984) Ann. Rev. Biochem. 53, 595-623.
17. Sweet, R.M., and Eisenberg, D. (1983) J. Mol. Biol. 171, 479-488.
18. Futerman, A.H., Fiorini, R.M., Roth, E., Low, M.G., and Silman, I. (1985) Biochem. J. 226, 369-377.
19. Vandenhove-Vandenbroucke, M.F. (1980) The Thyroid Gland, pp. 61-79, Raven Press, New-York.
20. Mullin, B.R., Levinson, R.E., Friedman, A., Henson, D.E., Winand, R.J., and Kohn, L.D. (1976) Endocrinology 100, 351-366.
21. Konishi, J., Herman, M.M., and Kriss, J.P. (1974) Endocrinology 95, 437-440.