The 5′-end of bovine thyroglobulin mRNA encodes a hormonogenic peptide

Luc Mercken, Marie Jeanne Simons and Gilbert Vassart

Insitut de Recherche Interdisciplinaire, Faculté de Médicine, Université Libre de Bruxelles, Campus Erasme, Bât C, Route de Lennik 808, 1070 Bruxelles, Belgium

Received 4 October 1982

The sequence of 370 bases at the 5′-end of bovine thyroglobulin mRNA has been determined. A 41 base untranslated segment was found preceeding the ATG initiator codon. It is followed by an open reading frame providing the first data on thyroglobulin primary structure. Analysis of the amino acid sequence demonstrated the presence of an 18 residue hydrophobic segment representing a putative signal peptide. Comparison of the amino terminal sequence of thyroglobulin with that of peptides known to contain thyroid hormones [7,8] demonstrated that the first tyrosine in native thyroglobulin is mainly found as thyroxine in the mature iodinated protein [8]. Our results clearly identify the amino-terminal region of thyroglobulin as an important hormonogenic domain of the protein.

Thyroid hormone *Protein sequence* *cDNA cloning*

1. INTRODUCTION

Thyroid hormones belong to a specific class of amino acids, the iodothyronines, the biosynthesis of which involves an iodinated protein precursor, thyroglobulin (Tg). Tg is a dimeric glycoprotein (M_r 2 × 330000) synthesized in large amount by the thyroid gland from the translation of an 8 kilobase mRNA [1]. Amongst the 120 tyrosine residues of Tg, about 40 become iodinated and a maximum of 8–12 [2,3] of the resulting iodotyrosines may couple into thyroxine (T_4) or triiodothyronine (T_3) which remain part of the polypeptide backbone of the protein. It is only after the complete lysosomal hydrolysis of Tg that active thyroid hormones are released from the gland. In terms of overall yield, thyroid hormone biosynthesis presents as a wasteful phenomenon. However, the spatial requirements for the efficient coupling of iodotyrosines must be particularly well met in specific domains of Tg, since thyroid hormone formation occurs even on molecules containing as few as 5 atoms of iodine [4]. In agreement with this view, thyroxine-rich peptides have been isolated from Tg [3,5,6] and some have been sequenced [7,8].

In two instances [5,6] these peptides were only found to be attached to Tg by disulfide bonds. This opened the possibilities that either they could be encoded by a separate mRNA [6] or that thyroid hormone formation would involve the rupture of peptide bonds [5]. Here, we have determined the sequence of 350 nucleotides at the 5′-end of bovine Tg mRNA. One of the hormogenic peptides [6] was readily identified within the single open reading frame of the mRNA.

2. MATERIALS AND METHODS

2.1. *cDNA sequencing*

A recombinant plasmid [9] containing a 2.6 kilobasepair (kb) insert corresponding to the 5′-end region of Tg structural gene was used as starting material.

The 2.6 kb clone was restricted by a 5-fold excess of restriction endonuclease. The fragments were either 3′-labeled by filling-in their extremities with the adequate deoxy[^{32}P]ribonucleotide in presence of reverse transcriptase, or by cordycepin in presence of terminal transferase. The labeled fragments were then cut with suitable restriction

Published by Elsevier Biomedical Press
001457593/82/0000—0000/$2.75

endonucleases or strand separated by polyacrylamide gel electrophoresis. The fragments were chemically degraded as in [10].

The 2.6 kb clone was restricted by *Pst*I and the fragments were inserted in the *Pst*I site of M13 mp2Am4/*Pst* (kindly provided by Dr G. Winter). Sequencing was then performed by the chain terminator method [11].

2.2. *mRNA sequencing*

A 60 basepair *Pst*I–*Alu*I fragment was prepared from the 2.6 kb clone. About 10 ng were hybridized to 5 μg bovine thyroglobulin mRNA [12] and elongated as in [13].

3. RESULTS AND DISCUSSION

Except for ~60 bases corresponding to the 5′-end of the mRNA, the complete structural gene of bovine Tg has been cloned into overlapping recombinant plasmids [9]. To determine the amino terminal portion of Tg, we have partially sequenced our 5′-end clone. As no initiator codon could be found within the first 300 nucleotides of the single open reading frame, we have sequenced the uncloned portion of the mRNA by the primer extension method [13]. The sequencing strategy is described in fig. 1.

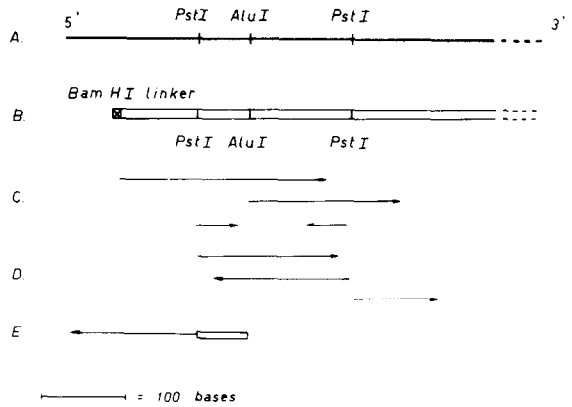

Fig. 1. Strategy for the sequencing of the 5′-end of Tg structural gene. Restriction map of the 5′-end portion of the gene (A) and of the 2.6 kb clone (B) used as substrate for sequencing. The arrows in (C) and (D) represent sequences obtained by the chemical degradation and the chain termination methods, respectively. The uncloned portion of the mRNA was sequenced by a primer extension method (E) as in section 2.

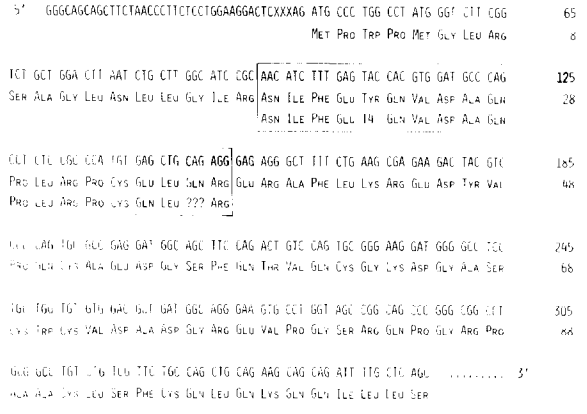

Fig. 2. Primary structure of the amino-terminal portion of bovine thyroglobulin as deduced from the sequence of its structural gene. Amino acids are numbered from the first ATG codon. Correspondence is shown between the sequence from residue 19–37 and that of the hormonogenic peptide identified in [8].

The major characteristics of the resulting sequence are depicted in fig. 2. The first initiator ATG is found ~40 nucleotides downstream from the 5′-terminus. This untranslated sequence contains a triplet of nucleotides which could not be identified by Sanger's method [10], probably because of the secondary structure of the mRNA. It is followed by a stretch of 16 codons specifying a relatively hydrophobic sequence which could represent the signal peptide for secretion [14]. It is noteworthy that a second ATG is found in phase at position 5. According to the current concepts on initiation of translation [15], we favour the first ATG as the true initiator.

Rawitch et al. [8] have determined the sequence of a thyroxine-rich peptide of 19 amino acids isolated from bovine Tg. The peptide was obtained by trypsin digestion of a M_r 10000 polypeptide linked to Tg by disulfide bonds. It contains ~30% of the thyroxine present in mature Tg [6]. An almost identical peptide was found between position 19 and 37 of the sequence deduced from the cloned cDNA (fig. 2). It is difficult to determine whether the minor discrepancy (Gln instead of Glu) is due to imprecision of the protein sequence or represents a true genetic difference between bovine stocks. A glutamine is read without ambiguity at position 36 which corresponds to an in-

determination in Rawitch's peptide. From the comparison of the two sequences, a hormonogenic function may thus be assigned to the first tyrosine residue of Tg at position 23 of the polypeptide chain. Another tyrosine residue is found at position 47. Further protein sequence data will determine whether this is involved in the hormonogenic coupling reaction.

Our results clearly identify the amino-terminal portion of native Tg as one of the important hormonogenic domains of the molecule. Interestingly, this sequence shares no detectable homology with the hormonogenic peptides isolated in [7]. Although these were obtained from a different species, this suggests that the various hormone forming sites along the Tg molecule do not consist of the plain repetition of a single domain. It is noteworthy that the thyroxine residue corresponding to the hormonogenic tyrosine identified in the present study is contained in a structure sharing no peptide bond with the mature iodinated protein [6]. This finding gives support to the concept that thyroid hormone formation could involve the cleavage of peptide bonds [16].

ACKNOWLEDGEMENTS

We thank Professor J.E. Dumont for his continuous support and interest, Dr. A.B. Rawitch for having made his data available before publication and Mrs D. Leemans for the preparation of the manuscript. This work was supported by NIH grant AM 21732 and FRSM grant 3.4334.80 to G.V., by a Grant Action Concertée to J.E. Dumont and by ARBD asbl. L. M. is fellow of the IRSIA Foundation and G.V. is 'Maître de Recherche' at the Belgian FNRS.

REFERENCES

[1] Van Herle, A., Vassart, G. and Dumont, J.E. (1979) New Engl. J. Med. 301, 239–249; 307–314.
[2] Sorimachi, K. and Ui, N. (1974) Biochim. Biophys. Acta 342, 30–40.
[3] Marriq, C., Arnaud, C., Rolland, M. and Lissitzky, S. (1980) Eur. J. Biochem. 111, 33–47.
[4] Rolland, M., Montfort, M.F. and Lissitzky, S. (1973) Biochim. Biophys. Acta 303, 338–347.
[5] Dunn, J.T., Dunn, A.D., Heppner, G. and Kim, P.S. (1981) J. Biol. Chem. 256, 942–947.
[6] Chernoff, S.B. and Rawitch, A.B. (1981) J. Biol. Chem. 256, 9425–9430.
[7] Marriq, C., Rolland, M. and Lissitzky, S. (1982) EMBO J. 1, 397–401.
[8] Rawitch, A.B., Chernoff, S.B. and Hamilton, J.W. (1982) Fed. Proc. 41, 1178.
[9] Christophe, D., Mercken, L., Brocas, H., Pohl, V. and Vassart, G. (1982) Eur. J. Biochem. 122, 461–469.
[10] Maxam, A. and Gilbert, W. (1980) Methods Enzymol. 65, 499–560.
[11] Sanger, F., Nicklen, S. and Coulson, A.R. (1977) Proc. Natl. Acad. Sci. USA 74, 5463–5467.
[12] Vassart, G., Verstreken, L. and Dinsart, C. (1977) FEBS Lett. 79, 15–18.
[13] Smith, A.J.H. (1980) Methods Enzymol. 65, 560–580.
[14] Alvino, C., Tassi, V., Polistina, C., Di Lauro, R. and Bonatti, S. (1982) Eur. J. Biochem. 125, 15–19.
[15] Kozak, M. (1980) Cell 22, 7–8.
[16] Dunn, J., Kim, P.S. and Dunn, A.D. (1982) J. Biol. Chem. 257, 88–94.