Vol. 119, No. 1, 1984 BIOCHEMICAL AND BIOPHYSICAL RESEARCH COMMUNICATIONS
February 29, 1984 Pages 335-342

The Structure of a Naturally Occurring 10K Polypeptide Derived from the Amino Terminus of Bovine Thyroglobulin

Allen B. Rawitch, Luc Mercken,* James W. Hamilton and Gilbert Vassart*

Department of Biochemistry, University of Kansas Medical Center, Kansas City, KS 66103

*Institut de Recherche Interdisciplinaire (IRIBHN) Hôpital Erasme, Route de Lennick, 808, 1070-Bruxelles, Belgique

Received January 12, 1984

SUMMARY—A combination of data derived from peptide sequencing and nucleic acid sequencing of cloned cDNA fragments has been used to define the complete amino acid sequence of a 10,000 M.W., thyroxine containing polypeptide derived from bovine thyroglobulin. This fragment, TG-F, which was obtained following reduction and alkylation, has been placed at the amino terminus of the parent protein with hormone located at residue 5 in the primary sequence of the thyroglobulin molecule. The carboxyl terminal sequence of this fragment -Cys-Gln-Leu-Gln is found on the N-terminal side of a lys residue, suggesting that the peptide bond cleavage which occurs to produce this 80 residue fragment from the parent (330K) thyroglobulin chain is a gln-lys. In addition, the amino acid sequence of this 10K fragment contains:

1. No sequence which would be a substrate for glycosylation and no carbohydrate.
2. Several repeated homologous amino acid sequences.
3. A striking number of β-bends predicted from Chou-Fasman analyses, particularly near its carboxyl terminus.

Thyroglobulin, the polypeptide precursor of the iodothyronine hormones, thyroxine and triiodothyronine, is a large and complex protein which undergoes extensive post translational modification. The mature protein contains both iodinated amino acids (1) and a large number, 25-30, of covalently bound oligosaccharide units (2, 3). While its polypeptide structure has been the subject of substantial efforts to define subunit sizes and structure-function relationships (4-6), these studies have been difficult due to the very large size of the protein.

In spite of this large size, several investigators have reported very specific or restricted domains within thyroglobulin, which appear to be sites of hormone production (7-10). Moreover, the amino acid sequence of two thyroxine containing polypeptide fragments has been recently described (11, 12).

A different approach has involved a study of the structure of the nucleic acids encoding thyroglobulin (13-16). These studies have shown that the nascent thyroglobulin

0006-291X/84 $1.50

molecule is composed of two identical polypeptide chains of 330,000 molecular weight. Recently, Christophe et al. (17) have successfully cloned the cDNA coding for bovine thyroglobulin, with the exception of approximately 60 bases at the 5' terminus of the message, into four overlapping recombinant plasmids. The sequence of the missing bases at the 5' end was obtained by the primer extension method following hybridization of a restricted cDNA fragment with the 33S mRNA, and a total of approximately 350 bases from the 5' end of the message has been sequenced (18).

Significantly, a base sequence coding for one of the recently described thyroxine containing polypeptide sequences (11) has been located near the 5' terminus of the mRNA which codes for bovine thyroglobulin. In this report, we describe the location of additional peptide sequences within the structure of the amino terminal fragment of thyroglobulin and their corresponding coding within the 5' region of the thyroglobulin mRNA. In addition, we have determined the carboxyl terminal sequence of the naturally occurring 10K hormonogenic peptide previously described (7) and fixed its location within the message, thus defining the amino acid sequence at one of the cleavage points in the nascent thyroglobulin polypeptide which gives rise to the smaller polypeptide fragments seen in isolated thyroglobulin. The primary structure of this fragment has been analyzed by the method of Chou and Fasman to predict secondary structure. The sequence was also treated according to Sankoff (19) in order to evaluate internal homology within the polypeptide.

MATERIALS AND METHODS

Preparation of a Thyroxine-enriched Fragment of Bovine Thyroglobulin—The thyroxine-enriched polypeptide was prepared from bovine thyroglobulin as previously described (7). Bovine thyroglobulin was reduced and alkylated, and the resulting carboxymethylated polypeptides were fractionated on a column of Sepharose CL-4B (Pharmacia) in 6 M urea. The small, iodine-enriched fraction obtained was subsequently chromatographed on a column of Bio-Gel P-60 (Bio-Rad) in 0.1% NH_4HCO_3 buffer and the purified fragment was obtained. The pure, iodine-enriched fragment was stored at -20°C following lyophilization.

Tryptic Digests—Protein samples were dissolved in 1% NH_4HCO_3. L-1-Tosylamido-2-phenylethyl chloromethyl ketone-treated trypsin (Worthington), corresponding to 5% by weight of the protein sample, was dissolved in 0.001 N HCl. Half of the trypsin solution was added to the sample, and the digest was allowed to proceed at 37°C. After 3 h, the remaining trypsin was added. The digest was continued for an additional 3 h, and the trypsin was inactivated by lowering the pH to 5 with glacial acetic acid and freezing at -20°C.

Automated NH_2-terminal Sequencing—Automated Edman (20) degradations were performed using a Beckman 890C protein-peptide sequenator equipped with a cold trap

with the 0.1 M Quadrol program (Beckman 121078). The butyl chloride fractions were converted to PTH derivatives by treatment with HCl for 10 min at 80°C. The fractions were then extracted twice with ethyl acetate and dried. PTH derivatives were identified by HPLC using a methanol-sodium acetate gradient on a reverse phase C_{18} column (21) and back-hydrolysis in HCl followed by amino acid analysis (22).

Carboxypeptidase Digestion—Samples were digested at 37°C with a mixture of equal amounts of carboxypeptidases A and B in 1% NH_4HCO_3 and at an enzyme to substrate molar ratio of 1:50. Aliquots equivalent to 5 nmol were removed after various time intervals, frozen, and lyophilized. The aliquots were then mixed with sample buffer and applied directly to a microbore amino acid analyzer.

Iodine Analysis—Iodine analyses were carried out by the method of Sandell and Kolthoff (23) following a wet ashing procedure using perchloric acid.

Amino Acid Analysis—Analyses were carried out on a Beckman model 121 MB automated amino acid analyzer following hydrolysis in constant boiling (5.7 N) HCl for 22 h at 110°C under vacuum.

HPLC Peptide Mapping—Following digestion of TG-F with trypsin in 1% ammonium bicarbonate, the digests were fractionated using high pressure liquid chromatography. A Beckman liquid chromatograph model 332 with model 420 microprocessor controller and an Altex Ultrasphere C_{18} column (4.6 x 250 mm) were used. A Hitachi 100-40 spectrophotometer was used as the detector. A linear gradient was used with the initial buffer, consisting of 1% ammonium bicarbonate, 5% acetonitrile, while the second component was 100% acetonitrile. The flow rate was 1 ml/min. The resulting profiles were monitored at 230, 325 and 350 nm. Fractions were collected allowing for the delay between appearance in the detector and outflow from the system, and the solvent was removed under a stream of nitrogen.

cDNA Sequencing—The sequencing of the portion of the cDNA corresponding to the 5' terminus of the thyroglobulin has been described elsewhere (18).

RESULTS

Amino acid sequences established from peptides in the 10K, naturally occurring fragment derived from bovine thyroglobulin (TG-F) are given in Table I. Figure 1 shows the amino acid sequence predicted from the partial cDNA structure established by Mercken et al. (18) with those residues confirmed by peptide sequencing procedures

TABLE I

	Sequence Obtained	Procedure
1.	Asn-Ile-Phe-Glu-T4-Gln-Val-Asp-Ala-Gln-Pro-Leu-Arg-Pro-Cys-Glu-Leu-Gln-Arg*	Automated Sequenator with back-hydrolysis plus carboxypeptidase time release
2.	Asp-Gly-Ala-Ser-Cys-Trp-Cys-Val-Asp-Ala-Asp-Gly-Arg**	Automated Sequenator with back-hydrolysis
3.	- - - -Phe-Cys-Gln-Leu-Gln***	Carboxypeptidase time release data

*confirmed in bovine, ovine and porcine thyroglobulin, 10K fragments (Rawitch et al., unpublished)

**confirmed in bovine and ovine thyroglobulin, 10K fragments (Rawitch et al., unpublished)

***confirmed in bovine thyroglobulin, 10K fragment only

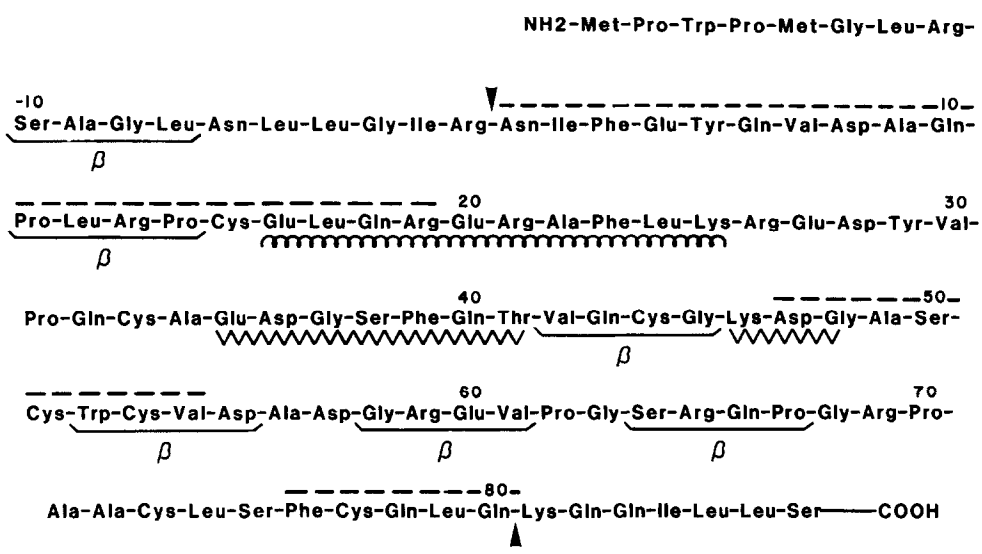

Figure 1. The N-terminal amino acid sequence proposed by Mercken et al. (18) from the cloned cDNA sequence of bovine thyroglobulin. The beginning and end of the TG-F, 10K peptide previously described are indicated by the large arrows. The overscored residues have been independently established by amino acid sequencing of the intact 10K fragment or of tryptic peptides derived from it. The proposed secondary structure is indicated according to Figure 2. —Alpha helix, -Beta structure, - Beta turn.

overscored. Also indicated in Figure 1 (arrows) are the proposed amino acid carboxyl termini of the TG-F, 10K fragment, based on carboxypeptidase data from the intact fragment, the amino acid composition of TG-F (7) and an assumption of an 18 residue signal sequence (18) at the amino terminus of the translation product from thyroglobulin mRNA.

Figure 2 illustrates analyses of the proposed sequence of TG-F by the method of Chou and Fasman (24). As may be seen in this summary plot (Figure 2A), a single region of predicted alpha helix extends from approximately residue 15 to residue 25, and two short stretches of β-structure (∿ 5 residues long) are predicted at residues 35 and 46. On the other hand, the analysis for β-turns (Figure 2B) predicts very high probabilities of a series of β-turns, especially in the C-terminal region of the TG-F structure. These structural features are also summarized in Figure 1. Examination of the proposed TG-F amino acid sequence by casual inspection and by computer analysis according to Sankoff (19) reveals several homologous repeats of amino acid sequence as illustrated in Figure 3. These homologous repeats ranged from 3 to 13 residues in length, with the 13 residue homology extending past the TG-F carboxy terminus by six residues.

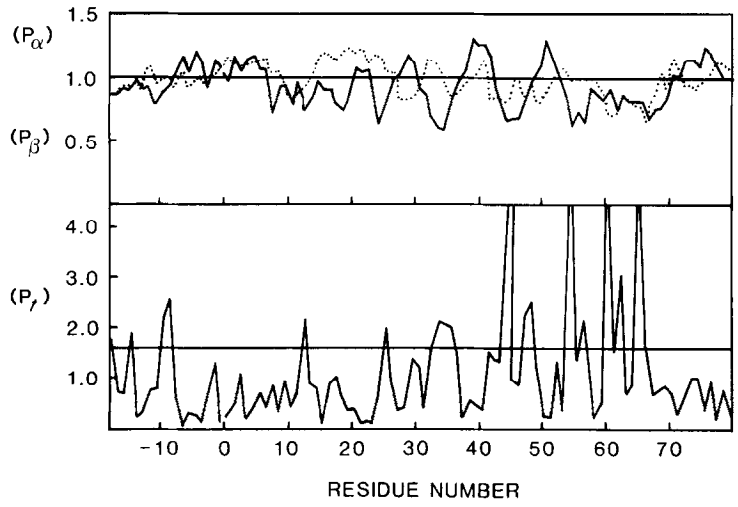

Figure 2. Secondary structure probability plots according to Chou and Fasman (24). (A) Profiles of α-helix probability <P_α> (- - - -) and b structure probability <P_β>(———). (B) A profile of β turn probability <P_t> versus residue number. Minimum cutoff values are indicated by the horizontal lines in each case.

DISCUSSION

The nature of the low molecular weight polypeptides derived from thyroglobulin upon reduction of disulfide bonds has been the subject of several investigations in recent years (7) (8) (10), with considerable divergence of opinion with respect to the role, if any, that these polypeptide fragments might play in the structure and function of this protein

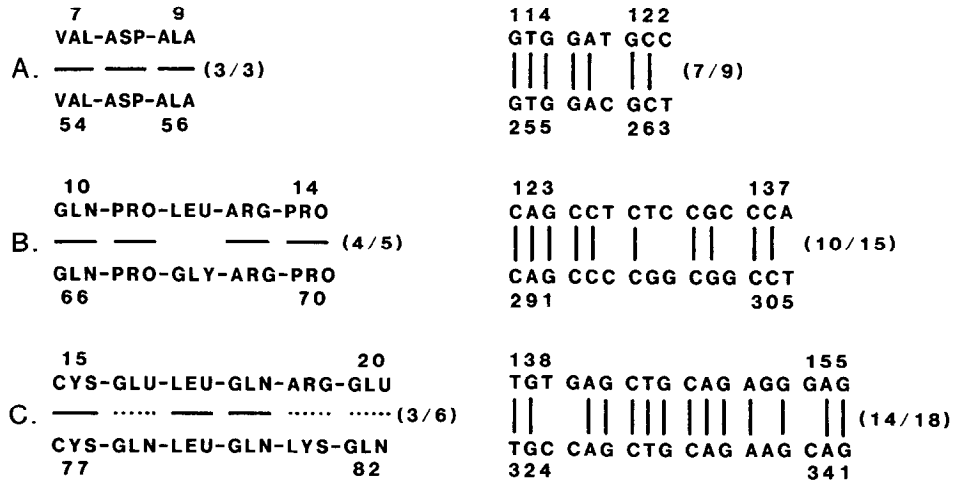

Figure 3. A summary of amino acid and nucleotide sequence homologous repeats in the structure of the TG-F, 10K fragment. The fractions given in parentheses are the fraction of residues found to be identical in the compared regions. The solid lines between amino acid sequences indicate identity, and the broken lines indicate one base substitution.

Chernoff and Rawitch have reported earlier (7) that at least one of these low molecular weight fragments, TG-F, is enriched in both iodine and thyroid hormone. Moreover, the amino acid sequence of a 19 residue, thyroxine containing, tryptic peptide has been established (11) and the coding for this sequence located within the 5' terminal region of the thyroglobulin mRNA (18).

Using combined data from peptide and nucleic acid sequencing studies we have proposed the complete amino acid sequence of the 10K peptide, TG-F (Figure 1). The proposed amino terminal of the fragment is based on: 1) the presence in the mRNA coding of an initial, 18 residue hydrophobic sequence which meets the criteria seen in signal or leader peptides (18); 2) the correspondence between the amino acid composition observed in TG-F (7) and that predicted for the polypeptide beginning 19 residues from the initiation codon and extending 80 residues to the proposed carboxyl terminus; and 3) the observations of several groups in the past which suggest that asx is a predominant amino terminal residue in thyroglobulin (25, 26).

The carboxyl terminal sequence was deduced from the release kinetics of carboxy peptidase experiments which yielded data compatible only with the proposed carboxyl terminus.

The peptides given in Table I not only substantiate the sequence predicted by Mercken et al. (18) for the amino terminal region of the thyroglobulin polypeptide chain, but also confirm the hormonogenic nature of this region of the protein. While the location of the thyroxine at residue position 5 in TG-F identifies one of the two tyrosines which participate in hormone formation (the acceptor tyrosine), the location of the second, or donor, tyrosine, which is iodinated and then contributes a diiodophenol moiety to form hormone, remains undefined. An examination of the amino acid sequence of TG-F reveals a single additional tyrosine residue at position 29 in the sequence. This tyrosine residue, however, does not appear to be that donor residue, since only a single iodinated tryptic peptide is obtained from TG-F (the amino terminal tryptic peptide) and no dehydroalanine was found (27). It would appear, therefore, that the second tyrosine residue which participates in hormone formation at this site is located outside of TG-F itself, perhaps substantially removed along the linear sequence of thyroglobulin but near residue 5 in the tertiary structure of the protein.

Further inspection of the amino acid sequence of TG-F reveals an interesting pattern of periodic β-bends predicted, particularly in the carboxyl terminal region of the fragment. This suggests a structure which may fold back on itself to form a compact domain containing the hormone bearing sequence. Moreover, several internal homologies (repeated homologous sequences) are found within TG-F. These repeats include an unusual sequence, Gln-Pro-X-Arg-Pro, which contains a trypsin resistant bond and a Glx-rich sequence, Cys-Glu-Leu-Gln-Arg-Glu, which occurs both near the amino terminus of TG-F and as a homologous sequence at the carboxyl terminus of the peptide, Cys-Gln-Leu-Gln-Lys-Gln. While these homologous sequence repeats are not large enough to imply gene duplication, per se, they suggest a possible repeating pattern of localized structural features within thyroglobulin. It has been suggested by Marriq et al. (28), based on mapping of CNBr digests and partial proteolytic digests, that significant internal homology exists in porcine thyroglobulin. More recently, Turner and Rawitch (29) have proposed significant internal homology in human thyroglobulin and preliminary data from cDNA sequences have demonstrated extensive internal homology in the amino terminal half of thyroglobulin (30).

ACKNOWLEDGEMENTS

This work was supported in part by NIH Grant AM18896 to ABR and AM21732 to GV. Helpful discussions with Dr. Carol Dziadik Turner are gratefully acknowledged, as is the technical assistance of Linda Fox and Marie Jeanne Simons.

REFERENCES

1. DeGroot, L.J. (1965) New Eng. J. Med. 272, 243-250.
2. Spiro, R. G. and Spiro, M. J. (1965) J. Biol. Chem. 240, 997-1001.
3. Arima, T., Spiro, M. J. and Spiro, R. G. (1972) J. Biol. Chem. 247, 1825-1835.
4. Pierce, J. G., Rawitch, A. B., Brown, D. M., and Stanley, P. G. (1965) Biochem. Biophys. Acta 111, 247-257.
5. Lissitzky, S., Rolland, M. and Reynaud, J. (1968) Eur. J. Biochem. 4, 464-471.
6. Spiro, M. J. (1973) J. Biol. Chem. 248, 4446-4460.
7. Chernoff, S. B. and Rawitch, A. B. (1981) J. Biol. Chem. 256, 9424-9431.
8. Dunn, J. T., Dunn, A. D., Heppner, D. G. and Kim, P. S. (1981) J. Biol. Chem. 256, 942-947.
9. Lejeune, P. J., Marriq, C., Rolland, M. and Lissitzky, S. (1983) FEBS Letters 156, 77-82.
10. Dunn, J. T., Kim, P. S. and Dunn, A. D. (1982) J. Biol. Chem. 257, 88-94.
11. Rawitch, A. B., Chernoff, S. B., Litwer, M. R., Rouse, J. B., and Hamilton, J. W. (1983) J. Biol. Chem. 258, 2079-2082.
12. Marriq, C., Rolland, M. and Lissitzky, S. (1982) E.M.B.O. Journal 1, 397-401.
13. Vassart, G., Refetoff, S., Brocas, H., Dinsart, C. and Dumont, J. E. (1975) Proc. Nat. Acad. Sci. USA 72, 3839-3843.
14. Vassart, G., Brocas, H., Lecocq, R. and Dumont, J. E. (1975) Eur. J. Biochem. 55, 15-22.

15. Brocas, H., Christophe, D., Pohl, V. and Vassart, G. (1982) FEBS Letters 137, 189-192.
16. 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) Nucl. Acid Res. 11, 2273-2285.
17. Christophe, D., Mercken, L., Brocas, H., Pohl, V. and Vassart, G. (1982) Eur. J. Biochem. 122, 461-469.
18. Mercken, L., Simons, M. and Vassart, G. (1982) FEBS Letters 149, 285-287.
19. Sankoff, D. (1972) Proc. Nat. Acad. Sci. USA 69, 4-6.
20. Edman, P. and Begg, G. (1967) Eur. J. Biochem. 1, 80-91.
21. Downing, M. R. and Mann, K. G. (1976) Anal. Biochem. 74, 298-319.
22. Mendez, E. and Lai, C. Y. (1975) Anal. Biochem. 68, 47-53.
23. Sandell, E. G. and Koltoff, I. H. (1937) Mikrochim. Acta 1, 9-25.
24. Chou, P. Y. and Fasman, G. D. (1977) J. Mol. Biol. 115, 135-175.
25. Dopheide, T. A. A. and Trikojus, V. M. (1964) Nature 201, 1128-1129.
26. Spiro, M. J. (1970) J. Biol. Chem. 245, 5820-5826.
27. Gavaret, J. M., Cahnmann, H. J. and Nunez, J. (1979) J. Biol. Chem. 254, 11218-11222.
28. Marriq, C., Stein, A., Rolland, M. and Lissitzky, S. (1978) Eur. J. Biochem. 87, 275-283.
29. Turner, C. D. and Rawitch, A. B. (1983) Fed. Proc. 42, 677 (abstract).
30. Mercken, L., Simons, M. J., Swillens, S. and Christophe, D. (1983) Ann. Endocr. 44, 26A (abstract).