

Preferential sites of proteolytic cleavage of bovine, human and rat thyroglobulin

The use of limited proteolysis to detect solvent-exposed regions of the primary structure

Fabrizio GENTILE and Gaetano SALVATORE

Centro di Endocrinologia e Oncologia Sperimentale del Consiglio Nazionale delle Ricerche
and Dipartimento di Biologia e Patologia Cellulare e Molecolare, University of Naples 'Federico II' Medical School, Italy

(Received June 23, 1993) – EJB 93 0926/2

The sites and the sequence of the proteolytic cleavages of bovine, human and rat thyroglobulin, during the limited proteolysis with thermolysin and trypsin, were determined by sequencing the NH₂ termini of the peptides produced and comparing them to the cDNA-derived sequences of bovine, human and rat thyroglobulin. Major cleavage sites of bovine thyroglobulin included residues 240, 502, 993, 1218, 1784 with thermolysin, and 240, 520, 1142, 1783, 2515 with trypsin. Cleavage sites of human thyroglobulin included residues 503, 982, 990, 1405, 1831 with thermolysin, and 522, 1627, 2513 with trypsin. Those of rat thyroglobulin included residues 501, 1776, 1784 with thermolysin, and 522, 1771, 1825, 2515 with trypsin (numbered as in bovine thyroglobulin). Thus, thyroglobulin from various species presents well localized and conserved regions particularly sensitive to proteolysis. The most sensitive region extended for 30 residues after residue 500. Another major cluster of cleavages was centered around residue 1800; this region was only partially sensitive in human thyroglobulin. A conserved tryptic site lay at the COOH terminus of the molecule. Most cleavage sites occurred within the inserted sequences that disrupt the Cys-rich, tandem repeats of thyroglobulin and either contain or are located near exon–intron junctions. Several cleavage sites lay in proximity of early iodinated or hormonogenic tyrosyl residues or of putative N-linked glycosylation sites. While a predominantly β -type secondary structure and a rigid three-dimensional structure were predicted for the Cys-rich repeats, stretches of predicted α -helices, β -strands and irregular structure were interspersed in the regions surrounding the cleavage sites. These data demonstrate the existence of conserved regions of thyroglobulin inherently sensitive to proteolysis, which most likely represent solvent-exposed regions of the primary structure, possibly forming loops at the surface of thyroglobulin.

Thyroglobulin, a dimeric glycoprotein with a molecular mass of 660 kDa, is the molecular site of synthesis of thyroid hormones, 3,3',5-triiodothyronine and 3,5,3',5'-tetraiodothyronine, formed within the peptide chain by the selective iodination and coupling of a small subset of tyrosyl residues. Both reactions are catalyzed by thyroid peroxidase. The hormones are freed, before being secreted into the blood, by the proteolytic degradation of thyroglobulin in the lysosomes of thyroid follicular cells.

Full-length coding sequences have been obtained from the cDNA of bovine [1] and human thyroglobulin [2], and partial sequences from the cDNA [3] or the gene [4] of rat thyroglobulin. However, little is known about the three-dimensional structure and the conformation of thyroglobulin.

Most studies of thyroglobulin function have concentrated on thyroid hormone synthesis. It has been long recognized that different tyrosyl residues have different reactivities toward iodine, so that iodination proceeds in a sequential order, which is controlled by the native structure of thyroglobulin [5, 6]. Five hormonogenic tyrosines have been identified and localized at positions 5, 1290, 2553, 2567 and 2746 of the human cDNA-derived sequence [7–15]. The association of consensus sequences with iodination and hormonogenesis in human thyroglobulin was reported [14]. However, because this association was not invariable, it is probable that conformational constraints play a role in hormonogenesis.

On the other hand, it is conceivable that the efficient and prompt release of the hormones from such a large protein precursor requires that different parts of the molecule undergo proteolytic cleavage in a predetermined order. It has been shown that the lysosomal degradation of thyroglobulin proceeds via selective endoproteolytic cleavages that produce the early release of the hormones from both extremities of thyroglobulin [16, 17]. Moreover, human thyroid cathepsins cleave thyroglobulin at a limited number of sites [18]. Fur-

Correspondence to F. Gentile, Centro di Endocrinologia e Oncologia Sperimentale del C.N.R., c/o Dipartimento di Biologia e Patologia Cellulare e Molecolare, Via Pansini, 5, I-80131 Naples, Italy
Fax: +39 81 770 1016.

Enzymes. Thermolysin from *Bacillus thermo-proteolyticus* rokko (EC 3.4.24.4); trypsin from bovine pancreas (EC 3.4.21.4).

ther study seemed to us necessary to understand whether these data reflect the adaptation of a specific lysosomal degradation pathway, the effect of conformation on the accessibility of thyroglobulin to proteolysis, or both.

In this paper, following an initial report [19], we have used limited proteolysis to probe the structure of thyroglobulin and to identify the regions of the primary structure that most probably are superficially located and accessible. We have used two enzymes, thermolysin and trypsin (treated with *N*^ε-tosyl-L-phenylalanylchloromethane), with different specificities and many potential target sites in the sequence of thyroglobulin, so that the selection of the cleavage sites could reflect the conformational constraints of thyroglobulin more than its primary structure. In order to obtain a more meaningful picture, we designed a comparative approach between the three animal species, i.e. calf, humans and rat, in

which the primary structure of thyroglobulin is entirely or partially known.

MATERIALS AND METHODS

Materials

Thermolysin from *Bacillus thermo-proteolyticus rokko* and trypsin from bovine pancreas (treated with *N*^ε-tosyl-L-phenylalanylchloromethane) were both obtained from Sigma. Products for electrophoresis were obtained from Bio-Rad, AcrylAide cross-linker and GelBond PAG film from FMC BioProducts, Immobilon P from Millipore, reagent-grade solvents from Baker; other analytical-grade chemicals from Serva Feinbiochemica.

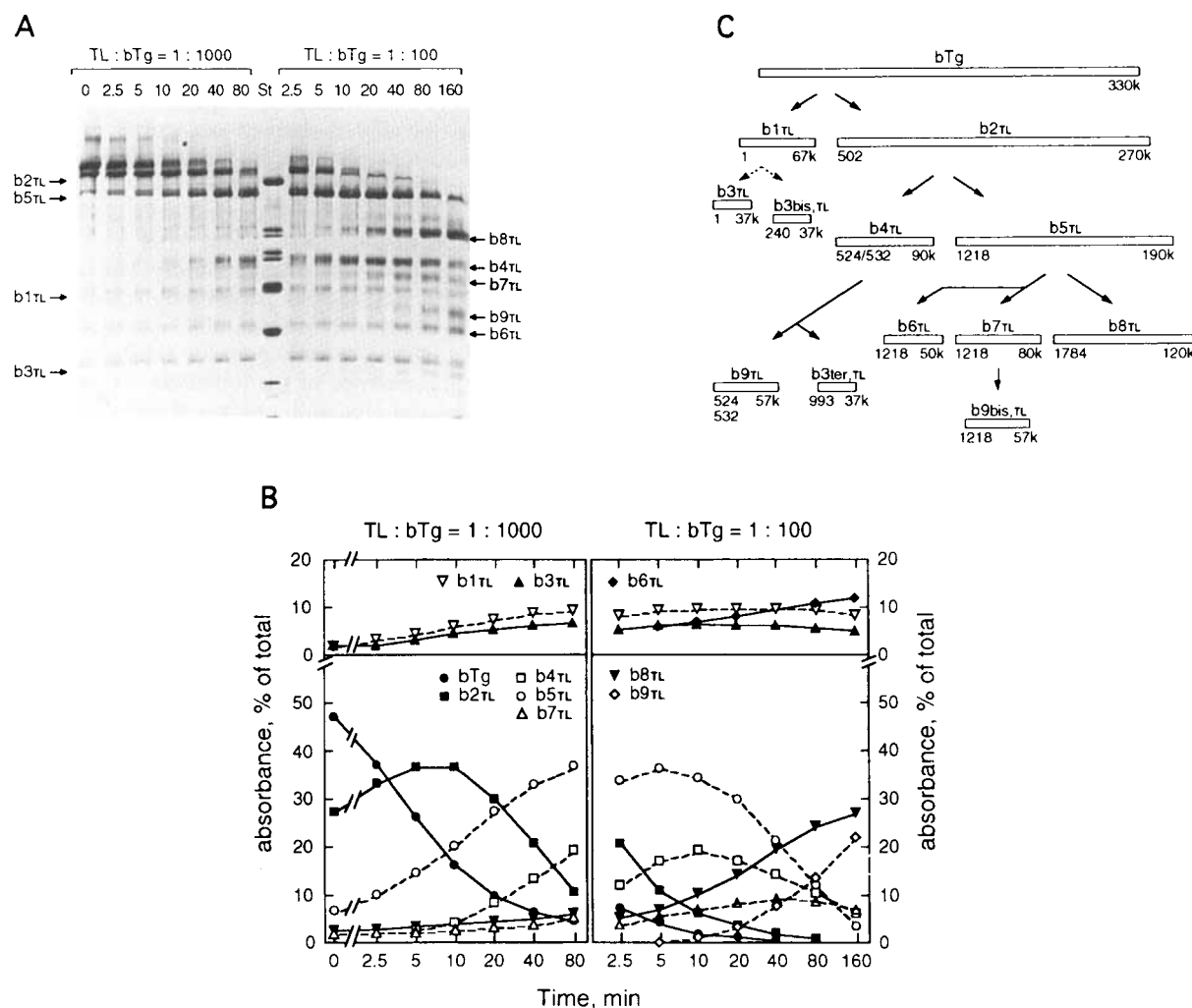


Fig. 1. Time course of limited proteolysis of bovine thyroglobulin with thermolysin. Bovine thyroglobulin (bTg) at the concentration of 1 mg/ml in 130 mM NaCl, 50 mM Tris/HCl pH 8.2 was incubated with thermolysin (TL) at 30°C. Aliquots were taken at timed intervals and the digestion was stopped by boiling in reducing SDS/PAGE sample buffer with 10 mM EDTA for 1.5 min. Two enzyme/substrate mass ratios were used (indicated at the top of A and B). (A) Reducing SDS/PAGE of the digestion products on a 4–13% total acrylamide gradient gel. 0 indicates control thyroglobulin (25 µg) with no enzyme added; the other numbers indicate time of incubation (min). St, molecular mass standards: myosin (205 kDa), β -galactosidase (116 kDa), phosphorylase *b* (94 kDa), bovine serum albumin (68 kDa), ovalbumin (45 kDa), carbonic anhydrase (30 kDa), soybean trypsin inhibitor (20 kDa). The peptides are marked on both sides of the gel. (B) Densitometric scan of the gel of A; absorbance was measured at 280 nm. For clarity, some curves are shown in separate insets at the top and the curve of peptide b6_{TL} was omitted in the left half of B. (C) Proposed flow diagram of the cleavages: each bar represents a peptide, with the peptide number above the bar (e.g. b2_{TL}), the apparent mass (in kDa) below the bar at right (e.g. 330 k) and the NH₂-terminal residue number as in [1] below the bar at left (e.g. 1).

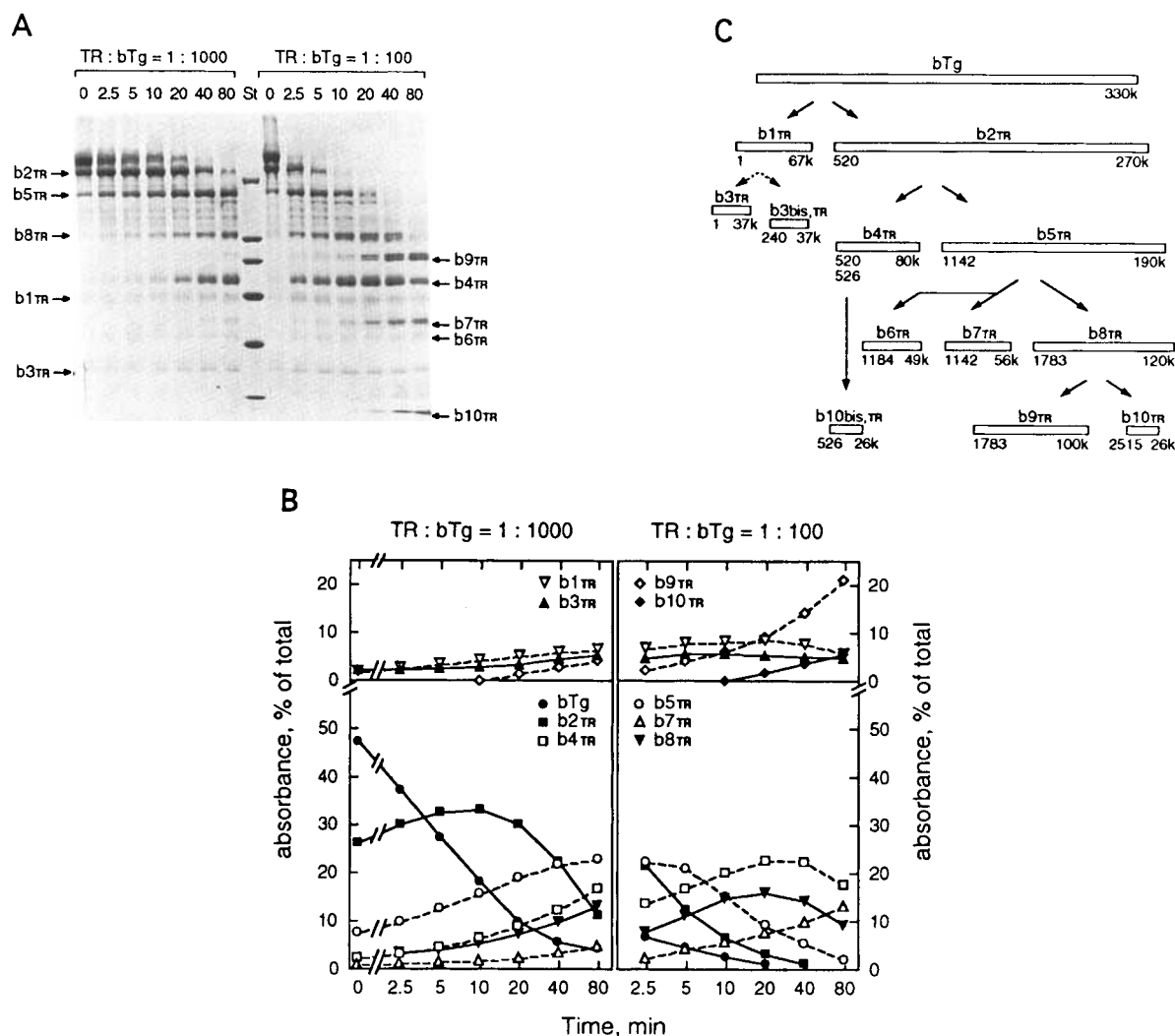


Fig. 2. Time course of limited proteolysis of bovine thyroglobulin with trypsin. Bovine thyroglobulin (bTg) at the concentration of 1 mg/ml in 130 mM NaCl, 50 mM Tris/HCl pH 8.2 was incubated with trypsin (TR) at 30°C. Aliquots were taken at timed intervals and the digestion was stopped by boiling in reducing SDS/PAGE sample buffer for 1.5 min. Two enzyme/substrate mass ratios were used (indicated at the top of A and B). (A) Reducing SDS/PAGE of the digestion products on a 4–13% total acrylamide gradient gel. 0 indicates control thyroglobulin (25 µg) with no enzyme added; the other numbers indicate time of incubation (min). St, molecular mass standards as in Fig. 1. The peptides are marked on both sides of the gel. (B) Densitometric scan of the gel of A; absorbance was measured at 280 nm. For clarity, some curves are shown in separate insets at the top, and the curve of peptide b6_{TR} was omitted. (C) Proposed flow diagram of the cleavages: each bar represents a peptide, with the peptide number above the bar (e.g. b1_{TR}), the apparent mass (in kDa) below the bar at right (e.g. 330 k) and the NH₂-terminal residue number as in [1] below the bar at left (e.g. 1).

Thyroglobulin preparation

Thyroglobulin was prepared under conditions minimizing proteolysis, from fresh bovine thyroids from the local abattoir, human thyroids from informed euthyroid patients hemilaryngectomized for non-thyroidal disease, and fresh Wistar rat thyroids. The tissue was minced with scissors and thyroglobulin extracted briefly on ice in 0.1 M sodium phosphate pH 7.2 and purified by fractional precipitation with 1.4–1.8 M ammonium sulphate, 50 mM Tris/HCl pH 7.2, and gel filtration on Sephacryl S-300 (Pharmacia LKB) in 130 mM NaCl, 50 mM Tris/HCl pH 7.2. The protein concentration was estimated by the absorbance at 280 nm, with an absorption coefficient of 10 cm⁻¹ for a 1% solution. The iodine content was measured as in [20].

Limited proteolysis of thyroglobulin

Thyroglobulin at the concentration of 1 mg/ml in 130 mM NaCl, 50 mM Tris/HCl pH 8.2 was incubated with proteases at 30°C. We selected thermolysin, which cleaves peptide bonds at the amino group of the hydrophobic amino acids isoleucine, leucine, valine, phenylalanine, methionine and alanine, and trypsin, which cleaves peptide bonds at the carbonyl group of the basic amino acids arginine and lysine. The trypsin used throughout had been treated with tosyl-phenylalanylchloromethane. In reducing SDS/PAGE, thyroglobulin typically shows two major bands [21, 22]. Time-course experiments of proteolysis were initially performed at the enzyme/substrate ratio of 1:1000 (by mass). Under these conditions we observed in all cases the almost complete dis-

Table 1. Peptides obtained by the limited proteolysis of bovine thyroglobulin. Lower-case letters indicate residues identified with less than full certainty. Yields for successive cycles of Edman degradation are reported in Tables S1 and S2 (Supplement). The NH₂-terminal residue numbers were assigned by comparison to the cDNA-derived sequence of bovine thyroglobulin [1]. b3_{TL}^a was obtained by a 30-min digestion at enzyme/substrate = 1:1000 (by mass); b3_{TL}^b was obtained by a 80-min digestion at enzyme/substrate = 1:100 (by mass).

Enzyme	Peptide or band no.	Apparent mass	NH ₂ -terminal sequence	NH ₂ -terminal residue no.
		kDa		
Thermolysin	b1 _{TL}	67	NIFEYQ	1
	b2 _{TL}	270	LSVGLNXXN	502
	b3 _{TL} ^a	37	b3 _{TL} : NIFEXXV	1
			b3bis _{TL} : XXIFAGL	240
	b3 _{TL} ^b	37	b3bis _{TL} : DXXFAGLDXA	240
			b3ter _{TL} : LAESPrAPSpV	993
	b4 _{TL}	90	b4 _{TL} : LRKPvGSFG	524
			b4bis _{TL} : FGFEVNLQEN	532
	b5 _{TL}	190	LGAAAGQRXq	1218
	b6 _{TL}	50	LGAAAGQR	1218
	b7 _{TL}	80	LGAAA	1218
	b7bis _{TL}	78	LGAAA	1218
	b8 _{TL}	120	LTPLEGTQDIL	1784
	b9 _{TL}	57	b9 _{TL} : LRKPvGSF	524
			b9bis _{TL} : LgAAAGQr	1218
			b9ter _{TL} : FgFEVNLQe	532
Trypsin	b1 _{TR}	67	NIFEYQVD	1
	b2 _{TR}	270	IDVALrKP	520
	b3 _{TR}	37	b3 _{TR} : NIFEXQv	1
			b3bis _{TR} : DIFAGL	240
	b4 _{TR}	80	b4 _{TR} : KPvVGSFGFE	526
			b4bis _{TR} : IDVALXKPvV	520
	b5 _{TR}	190	TSPGYSPAXXAe	1142
	b6 _{TR}	49	VAGSQPA	1184
	b7 _{TR}	56	TSPGYSPAXXAE	1142
	b8 _{TR}	120	GLTPLEGTQD	1783
	b9 _{TR}	100	GLTPLEGTQD	1783
	b10 _{TR}	26	b10 _{TR} : TSsKTAF	2515
			b10bis _{TR} : KPvVGSF	526

appearance of the slower band of the thyroglobulin doublet within 80 min. The further degradation of thyroglobulin was studied at higher enzyme/substrate ratios, i.e. 1:100 and 1:50. Under these conditions, the disappearance of the faster band of the doublet was observed within 40–320 min, depending on the species and the enzyme used. Conditions were chosen that enabled continuity between experiments at different ratios, so that the peptide pattern produced at the latest time points of the digestion at the lower ratio could be superimposed onto that of the earliest time points of the digestion at the higher ratio. Aliquots were taken after 2.5, 5, 10, 20, 40 and 80 min and, sometimes, 160 and 320 min; the digestion was stopped by boiling for 1.5 min in SDS/PAGE sample buffer (10 mM Tris/HCl pH 6.8, 1% SDS, 0.7 M 2-mercaptoethanol, 1.36 M glycerol, bromophenol blue as tracking dye). In the digestions with thermolysin, 10 mM EDTA was added just before boiling.

Separation and identification of the peptides produced

The digestion products were analyzed by SDS/PAGE in reducing conditions on 4–13% total acrylamide gradient

Table 2. Peptides obtained by the limited proteolysis of human thyroglobulin. Lower-case letters indicate residues identified with less than full certainty. Yields for successive cycles of Edman degradation are reported in Table S3 (Supplement). The NH₂-terminal residue numbers were assigned by comparison to the cDNA-derived sequence of human thyroglobulin [2].

Enzyme	Peptide or band no.	Apparent mass	NH ₂ -terminal sequence	NH ₂ -terminal residue no.
		kDa		
Thermolysin	h1 _{TL}	70	NIFEYQ	1
	h2 _{TL}	270	LSVGLD	503
	h3 _{TL}	47	FVPSXT	597
	h4 _{TL}	205	h4 _{TL} : LSFYQ	982
			h4bis _{TL} : FSPXD	990
	h5 _{TL}	160	LXGDHF	1405
Trypsin	h6 _{TL}	57	h6 _{TL} : LQGDHFG	1405
			h6bis _{TL} : NIFEXQV	1
	h7 _{TL}	110	LXXELFS	1831
	h1 _{TR}	80	NIFEYQ	1
	h2 _{TR}	265	DGTMNKP	522
	h3 _{TR}	56	NIFEXQVD	1
	h4 _{TR}	29	h4 _{TR} : NIFEXQ	1
			h4bis _{TR} : TSXKTA	2513
	h5 _{TR}	230	DGTMNKP	522
	h6 _{TR}	145	DGTMNK	522
	h7 _{TR}	105	DALGNSK	1627

gels polymerized on Gelbond-PAG plastic backing (FMC BioProducts). Molecular mass standards were: myosin (205 kDa), β -galactosidase (116 kDa), phosphorylase *b* (94 kDa), bovine serum albumin (68 kDa), ovalbumin (45 kDa), carbonic anhydrase (30 kDa) and soybean trypsin inhibitor (20 kDa). The gels were stained with 0.1% Coomassie brilliant blue R-250 in 25% (by vol.) 2-propanol, 10% (by vol.) acetic acid, destained in 25% (by vol.) methanol, 10% acetic acid, equilibrated in 0.7 M glycerol and air dried. The time curves of the intensities of the bands were obtained by densitometry, using a 2202 Ultrosan laser densitometer (LKB). For identification, all the major peptides were isolated by SDS/PAGE and Western blotting in conditions of maximal accumulation. Samples with 0.5 mg proteolytic digest were subjected to electrophoresis in the presence of 100 mM sodium thioglycollate on 4–13% gradient gels or 6% gels prepared at least 24 h before; they were transferred to polyvinylidene difluoride membranes (Immobilon P, Millipore) by semi-dry blotting in 25 mM Tris base, 10 mM glycine transfer buffer, at the constant current of 0.8 mA/cm² for 1 h. The membranes were washed thoroughly in double-distilled water, stained briefly with 1% Coomassie brilliant blue R-250 in 50% methanol and destained in 50% methanol, 10% acetic acid. The bands were cut and subjected to NH₂-terminal peptide microsequencing at the Protein Structure Laboratory (University of California at Davis, CA). The sequences of the peptides obtained by limited proteolysis of bovine and human thyroglobulin were compared to the respective cDNA-derived sequences [1, 2]; cleavage sites are identified by the number of the residue at whose amino group the cleavage occurred; with rat thyroglobulin, all available cDNA-derived sequences were used [1–3].

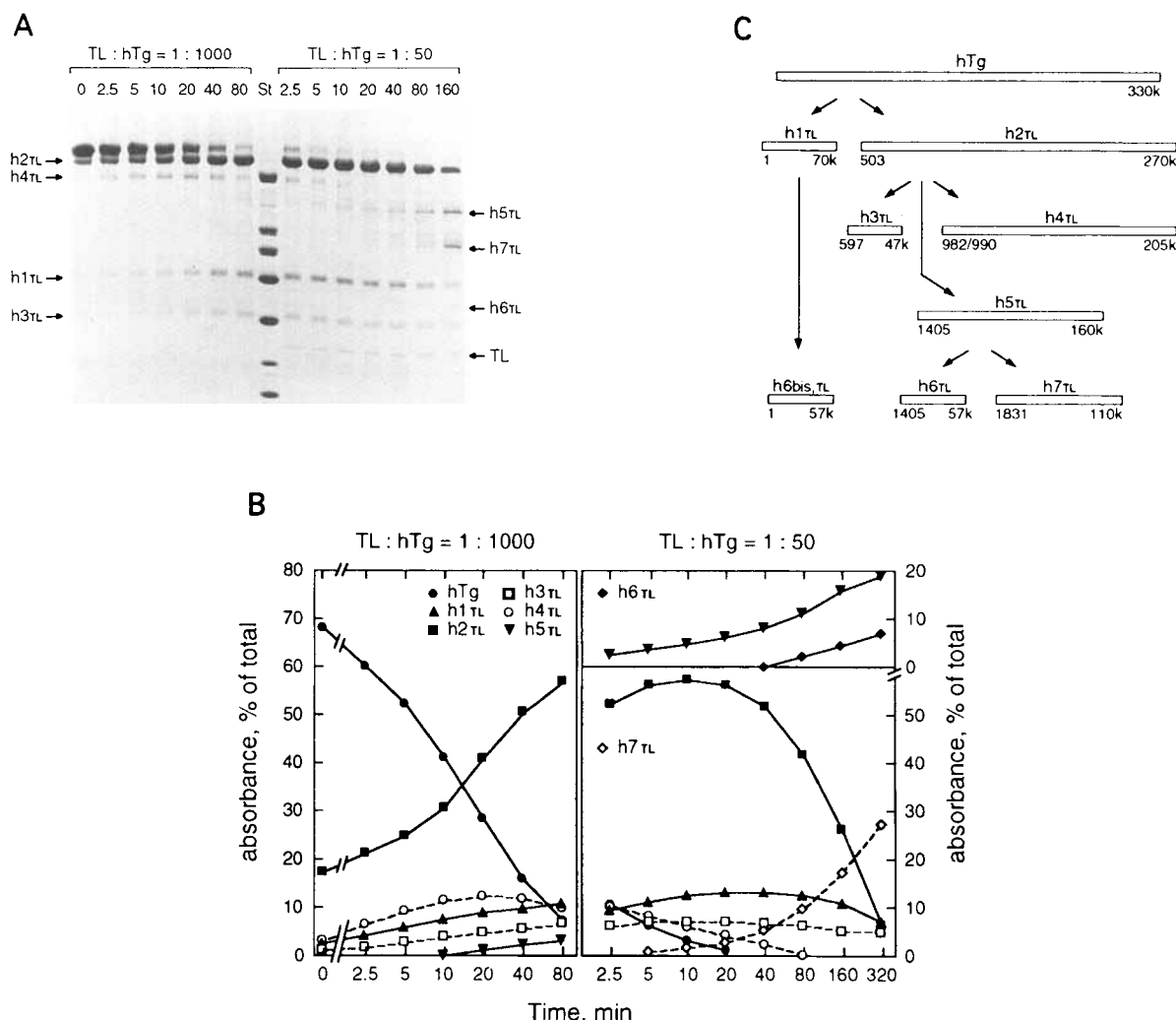


Fig. 3. Time course of limited proteolysis of human thyroglobulin with thermolysin. Human thyroglobulin (hTg) was incubated with thermolysin (TL) under the same experimental conditions as in Fig. 1. Two enzyme/substrate mass ratios were used (indicated at the top of A and B). (A) Reducing SDS/PAGE of the digestion products on a 4–13% total acrylamide gradient gel. 0 indicates control thyroglobulin (25 μ g) with no enzyme added; the other numbers indicate time of incubation (min) (the incubation time of 320 min was on a separate gel, not shown). St, molecular mass standards as in Fig. 1. The peptides are marked on both sides of the gel. The position of thermolysin is also marked. (B) Densitometric scan of the gel of A; absorbance was measured at 280 nm. For clarity, some curves in the right half of B are shown in a separate inset at the top. (C) Proposed flow diagram of the cleavages: each bar represents a peptide, with the peptide number above the bar (e.g. h1_{TL}), the apparent mass (in kDa) below the bar at right (e.g. 330 k) and the NH₂-terminal residue number as in [2] below the bar at left (e.g. 503).

Sequence analysis and structure predictions

Similarity scores between amino acids were calculated and sequence alignments were performed with the MacVector software (IBI), using the homology matrix pam 250, based on Schwartz-Dayhoff's scoring matrix for observed amino acid substitutions [23] (deletion penalty –12, gap penalty –6); one-residue deletions were placed between residues 420–421, 1780–1781, 2068–2069 and 2174–2175 of bovine thyroglobulin, 774–775, 965–966, 1401–1402, 1428–1429, 2066–2067 and 2172–2173 of human thyroglobulin, and 926–927 of rat thyroglobulin. Secondary structure predictions were obtained with the same software, by a combination of the methods proposed by Chou and Fasman [24], and Garnier et al. [25]; assignments were made only when both methods agreed and the graphical representation was freely adapted. Hydrophilicity plots were calculated using the hydropathy scale of Kyte-Doolittle [26], with reversed signs, using a nine-residue moving window.

RESULTS

Limited proteolysis of thyroglobulin; isolation of the peptides produced and identification of the cleavage sites

Bovine thyroglobulin

The limited digestions of bovine thyroglobulin with thermolysin and trypsin are shown in Figs 1 and 2, respectively. Each figure contains an analysis of the digestion in reducing SDS/PAGE (A), a densitometric scan of the gel (B), and a flow diagram of the cleavages (C), based on both the abundance of the peptides at successive time points and their NH₂ termini. The apparent molecular masses and the NH₂-terminal sequences of the peptides produced are shown in Table 1. The thyroglobulin used contained 1.00% (by mass) iodine and appeared in reducing SDS/PAGE as a typical doublet, whose faster band (F band) derives from the cleavage of the slower band at residue 520 [22]. Another band (formerly described as the A band [21]), was visible. A di-

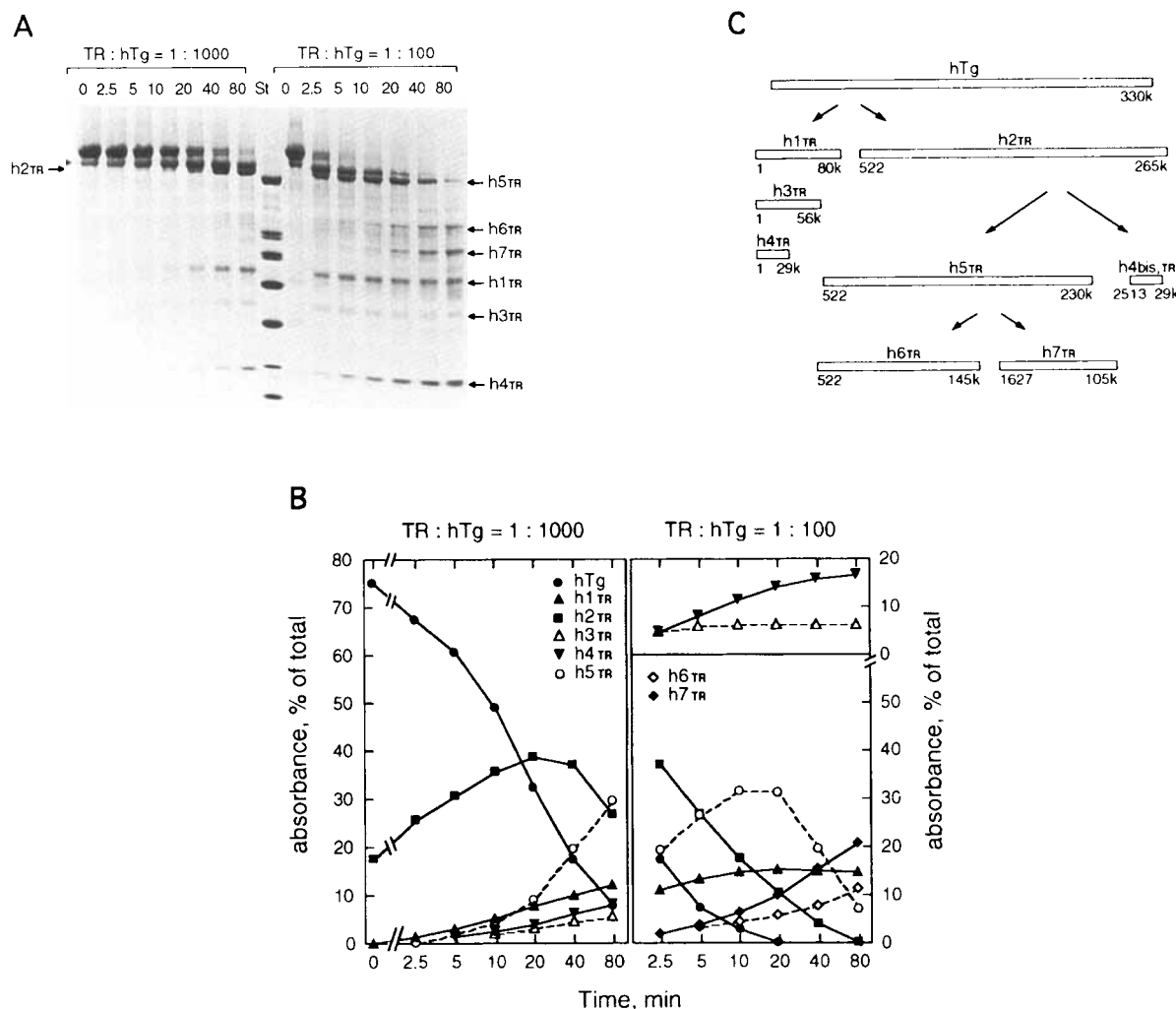


Fig. 4. Time course of limited proteolysis of human thyroglobulin with trypsin. Human thyroglobulin (hTg) was incubated with trypsin (TR) under the same experimental conditions as in Fig. 2. Two enzyme/substrate mass ratios were used (indicated at the top of A and B). (A) Reducing SDS/PAGE of the digestion products on a 4–13% total acrylamide gradient gel. 0 indicates control thyroglobulin (25 µg) with no enzyme added; the other numbers indicate time of incubation (min). St, molecular mass standards as in Fig. 1. The peptides are marked on both sides of the gel. (B) Densitometric scan of the gel of A; absorbance was measured at 280 nm. For clarity, some curves in the right half of B are shown in a separate inset at the top. (C) Proposed flow diagram of the cleavages: each bar represents a peptide, with the peptide number above the bar (e.g. h1_{TR}), the apparent mass (in kDa) below the bar at right (e.g. 80 k) and the NH₂-terminal residue number as in [2] below the bar at left (e.g. 522).

gestion pattern qualitatively similar was obtained with a thyroglobulin having an iodine content of 0.65% (by mass).

In the digestion with thermolysin at low enzyme/substrate ratio, the earliest cleavage occurred at residue 502 and produced a peptide (b1_{TL}) corresponding to the NH₂ terminus of thyroglobulin, and a peptide (b2_{TL}) superimposed onto the F band of control thyroglobulin. Another sequence corresponding to the NH₂ terminus of thyroglobulin (b3_{TL}) was contained in band b3_{TL}, together with a sequence (b3bis_{TL}) starting at residue 240. It is uncertain whether peptides b3_{TL} and b3bis_{TL} were derived from the cleavage of peptide b1_{TL}, although it seems possible. Oligosaccharide chains may have contributed to make their migration slower than expected. In other cases the sum of the apparent masses of adjacent peptides exceeded that of the parent peptide (Figs 1–6), which may mean that conformational properties of the isolated peptides, besides their masses, were affecting their electrophoretic

mobility. Peptide b2_{TL} was apparently cleaved at residue 1218, because its decline coincided with the increase of bands b4_{TL} and b5_{TL}; the former contained two closely related sequences (b4_{TL} and b4bis_{TL}). In the digestion at the higher enzyme/substrate ratio, the decline of peptide b5_{TL} coincided with the increase of bands b6_{TL}, b7_{TL}, and b8_{TL}. On some gels, band b7_{TL} was resolved into two closely migrating peptides (b7_{TL} and b7bis_{TL}) that had the same NH₂ terminus. This was followed by the increase of a band (b9_{TL}) that contained three sequences: two of them (b9_{TL} and b9ter_{TL}) apparently derived from peptides b4_{TL} and/or b4bis_{TL}; sequence b9bis_{TL} probably derived from peptides b7_{TL} and/or b7bis_{TL}. A peptide starting at residue 993 (b3ter_{TL}) was found in a late sample of band b3_{TL} and probably represented the COOH-terminal part of peptides b4_{TL} and/or b4bis_{TL}.

The pattern of peptides produced by trypsin was very similar. Peptides b1_{TR} and b2_{TR} apparently derived from

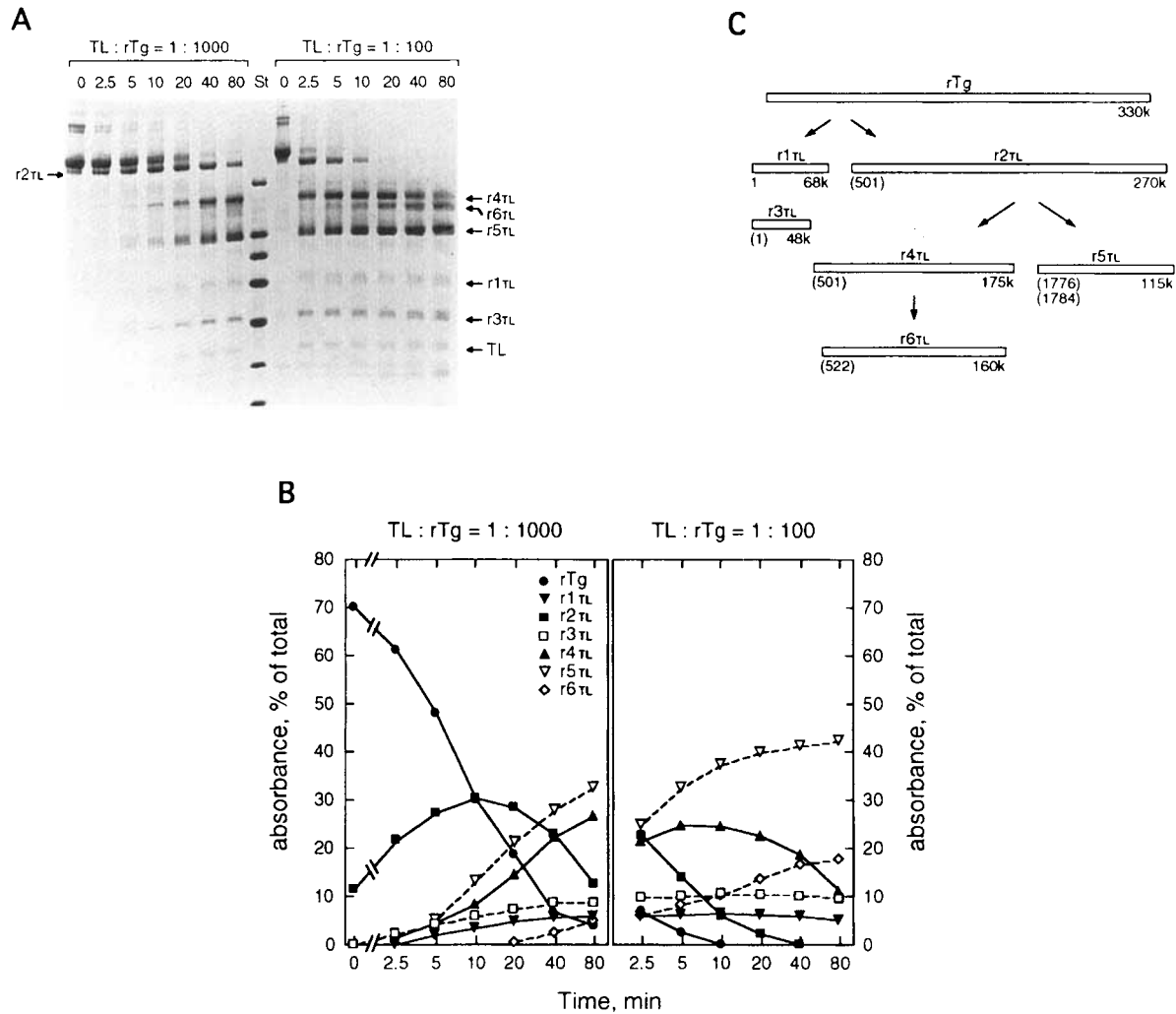


Fig. 5. Time course of limited proteolysis of rat thyroglobulin with thermolysin. Rat thyroglobulin (rTg) was incubated with thermolysin (TL) under the same experimental conditions as in Fig. 1. Two enzyme/substrate mass ratios were used (indicated at the top of A and B). (A) Reducing SDS/PAGE of the digestion products on a 4–13% total acrylamide gradient gel. 0 indicates control thyroglobulin (25 μ g) with no enzyme added; the other numbers indicate time of incubation (min). St, molecular mass standards as in Fig. 1. The peptides are marked on both sides of the gel. The position of thermolysin is also marked. (B) Densitometric scan of the gel of A; absorbance was measured at 280 nm. (C) Proposed flow diagram of the cleavages: each bar represents a peptide, with the peptide number above the bar (e.g. r1_{TL}), the apparent mass (in kDa) below the bar at right (e.g. 270 k) and the NH₂-terminal residue number, deduced from comparison to the cDNA-derived sequence of bovine thyroglobulin [1], below the bar at left (e.g. 501). The NH₂-terminal residue number of peptide r3_{TL} (in parentheses) was only tentatively established.

the early cleavage at residue 520. Two peptides (b3_{TR} and b3bis_{TR}) were contained in band b3_{TR}. Band b4_{TR}, with its two related sequences (b4_{TR} and b4bis_{TR}), and peptide b5_{TR} resulted from the cleavage of peptide b3_{TR} at residue 1142. The rapid decline of peptide b5_{TR} coincided with the increase of peptides b6_{TR}, b7_{TR} and b8_{TR}. The degradation of peptide b8_{TR} was also rapid and yielded peptide b9_{TR} and one of the sequences (b10_{TR}) found in band b10_{TR}. Sequence b10bis_{TR}, instead, probably derived from peptides b4_{TR} and/or b4bis_{TR}.

Human thyroglobulin

Figs 3 and 4 show the limited digestions of human thyroglobulin with thermolysin and trypsin, respectively. Table 2 contains the apparent molecular masses and the NH₂-terminal sequences of the peptides produced. The thyroglobulin used

contained 0.35% iodine (by mass) and appeared in reducing SDS/PAGE as a typical doublet, whose faster band derives from the cleavage of the thyroglobulin protomer at residue 504 [22]. Digestions of thyroglobulins whose iodine content varied over 0.10–0.35% (by mass) produced peptide patterns that were qualitatively similar.

In the digestion with thermolysin at low enzyme/substrate ratio, the early cleavage at residue 503 produced a peptide (h1_{TL}) corresponding to the NH₂ terminus of thyroglobulin and a peptide (h2_{TL}) superimposed onto the F band. Bands h3_{TL} and h4_{TL} were present at low levels and possibly derived from the degradation of a minor fraction of peptide h2_{TL}. Peptide h2_{TL} persisted at higher levels and for a longer time than in the other digestions, even at high enzyme/substrate ratio (1:50). Its decline coincided with the increase of peptide h5_{TL}. The fate of the region between residues 503

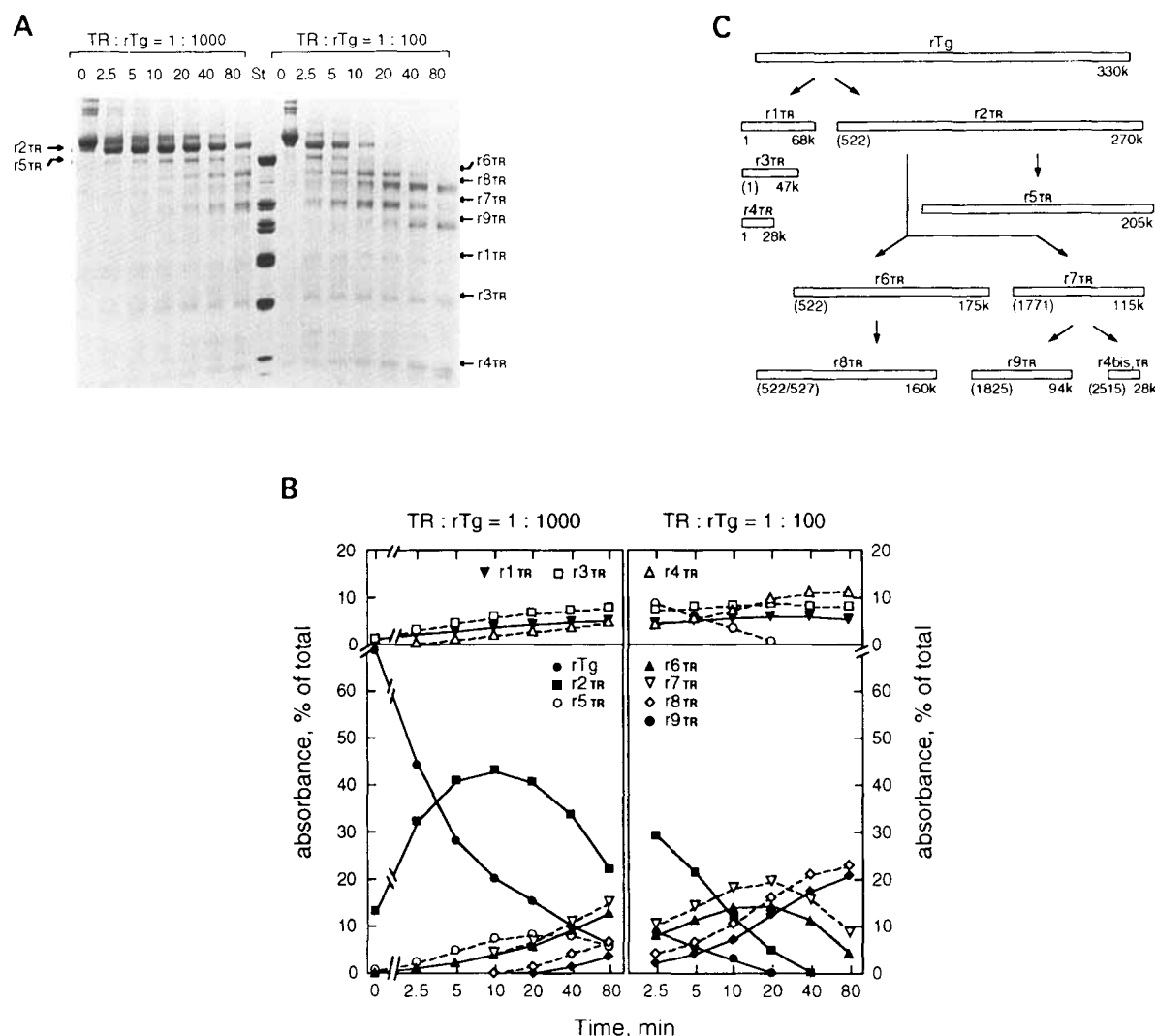


Fig. 6. Time course of limited proteolysis of rat thyroglobulin with trypsin. Rat thyroglobulin (rTg) was incubated with trypsin (TR) under the same experimental conditions as in Fig. 2. Two enzyme/substrate mass ratios were used (indicated at the top of A and B). (A) Reducing SDS/PAGE of the digestion products on a 4–13% total acrylamide gradient gel. 0 indicates control thyroglobulin (25 μ g) with no enzyme added; the other numbers indicate time of incubation (min). St, molecular mass standards as in Fig. 1. The peptides are marked on both sides of the gel. (B) Densitometric scan of the gel of A; absorbance was measured at 280 nm. For clarity, some curves are shown in separate insets at the top. (C) Proposed flow diagram of the cleavages: each bar represents a peptide, with the peptide number above the bar (e.g. r1_{TR}), the apparent mass (in kDa) below the bar at right (e.g. 270 k) and the NH₂-terminal residue number, deduced from comparison to the cDNA-derived sequence of bovine thyroglobulin [1], below the bar at left (e.g. 522). The amino-terminal residue number of peptides r1_{TR} and r4_{TR} was established also by comparison to the gene-derived rat thyroglobulin sequence [4]. The NH₂ terminus of peptide r3_{TR} (in parentheses) was only tentatively established.

and 1405 is unclear. The cleavage of peptide h5_{TL} at residue 1831 produced one of the sequences (h6_{TL}) contained in band h6_{TL} and peptide h7_{TL}. Band h6_{TL} also contained a sequence (h6bis_{TL}) that corresponded to the NH₂ terminus of thyroglobulin and probably derived from peptide h1_{TL}.

In the tryptic digestion, the NH₂-terminal peptide h1_{TR} and peptide h2_{TR} were the earliest intermediates, produced by the cleavage of thyroglobulin at residue 522. Other NH₂-terminal peptides included peptide h3_{TR} and one of the sequences (h4_{TR}) contained in band h4_{TR}. Peptide h5_{TR} and the other sequence (h4bis_{TR}) found in band h4_{TR} were seemingly derived from the cleavage of peptide h2_{TR} at residue 2513. At higher enzyme/substrate ratio, peptides h6_{TR} and h7_{TR} resulted from the cleavage of peptide h5_{TR} at residue 1627.

The presence of additional oligosaccharide units at residues 57 [27] and 2562 [28], and of an additional putative glycosylation site at residue 510 of human thyroglobulin [2], with respect to bovine thyroglobulin [1] (Table S5, Supplement) may explain why peptides h1_{TL}, h5_{TR} and h1_{TR} had greater apparent masses than the corresponding peptides b1_{TL}, b10_{TR} and b1_{TR}, respectively (compare Tables 1 and 2).

Rat thyroglobulin

Figs 5 and 6 show the limited digestions of rat thyroglobulin with thermolysin and trypsin, respectively. Table 3 shows the apparent molecular masses and the NH₂-terminal

Table 3. Peptides obtained by the limited proteolysis of rat thyroglobulin. Lower-case letters indicate residues identified with less than full certainty. Yields for successive cycles of Edman degradation are reported in Table S4 (Supplement). The NH₂-terminal residue numbers were assigned by comparison to the partial cDNA-derived sequence (r) [3] or gene-derived sequence (r*) of rat thyroglobulin [4], and to the full-length cDNA-derived sequences of bovine (b) [1] and human (h) thyroglobulin [2]. Assignments of residue numbers in parentheses are tentative.

Enzyme	Peptide or band no.	Apparent mass kDa	NH ₂ -terminal sequence	NH ₂ -terminal residue no.
Thermolysin	r1 _{TL}	68	NIFEy	1 (r*, b, h)
	r2 _{TL}	270	LIPVsLD	501(b), 502(h)
	r3 _{TL}	48	NiFeXqv	(1) (r*, b, h)
	r4 _{TL}	175	LIPVsLD	501(b), 502(h)
	r5 _{TL}	115	r5 _{TL} : LGgQEFL	1776(b), 1173(h)
Trypsin			r5bis _{TL} : LTLLEGT	1784(b), 1782(h)
	r6 _{TL}	160	VAMXKXVV	522(b), 523(h)
	r1 _{TR}	68	XIFEXQ	1 (r*, b, h)
	r2 _{TR}	270	VAMXKSV	522(b), 523(h)
	r3 _{TR}	47	NXXEXQV	(1) (r*, b, h)
	r4 _{TR}	28	r4 _{TR} : NIFEYQV	1 (r*, b, h)
			r4bis _{TR} : X(N/V)XKTA	733(r) 2515(b), 2513(h)
	r5 _{TR}	205	DIQGTS	–
	r6 _{TR}	175	VAMGKSV	522(b), 523(h)
	r7 _{TR}	115	QMTMSL	1771(b), 1768(h)
	r8 _{TR}	160	r8 _{TR} : VAMGK	522(b), 523(h)
			r8bis _{TR} : (S/P)VVGTF	527(b), 528(h)
	r9 _{TR}	94	SEAPEG	41(r) 1825(b), 1823(h)

sequences of the peptides produced. The thyroglobulin used contained 0.70% (by mass) iodine.

The earliest products of the digestion with thermolysin were the NH₂-terminal peptide r1_{TL} and a peptide (r2_{TL}) comigrating with the F band, whose sequence had no matches in the partial cDNA-derived sequence of rat thyroglobulin. Using a stringency of 70% of the maximum possible score, unique matches were found at residues 501 of bovine thyroglobulin and 502 of human thyroglobulin. A sequence found in band r3_{TL} was also tentatively placed at position 1. Peptide r2_{TL} appeared to be cleaved into peptide r4_{TL} and band r5_{TL}; the latter contained two sequences (r5_{TL} and r5bis_{TL}) that had unique matches at bovine residues 1776 and 1784 (human residues 1773 and 1782). Peptide r4_{TL} was then degraded into peptide r6_{TL}.

Similar peptides were obtained with trypsin. The first cleavage produced the NH₂-terminal peptide r1_{TR} and peptide r2_{TR}, whose sequence had unique matches at bovine residue 522 and human residue 523. Band r3_{TR} contained a sequence that was also tentatively placed at position 1. Another sequence corresponding to the NH₂ terminus of thyroglobulin (r4_{TR}) was detected in band r4_{TR}. Low levels of a peptide (r5_{TR}) were present whose NH₂-terminal sequence had no matches in bovine or human thyroglobulin. The cleavage of peptide r2_{TR} at a site corresponding to bovine residue 1771 (human residue 1768) produced peptides r6_{TR} and r7_{TR}. The digestion of peptide r6_{TR} yielded a band (r8_{TR}) that contained two sequences (r8_{TR} and r8bis_{TR}), matching bovine residues 522 and 527 (human residues 523 and 528), respectively. Further degradation and cleavage of peptide r7_{TR} at a site corresponding to bovine residue 2515 (human residue 2513) produced peptides r9_{TR} and r4bis_{TR}. The latter was contained in band r4_{TR} together with peptide r4_{TR}.

Analysis of the cleavage sites

Primary structure

Table 4 shows a synopsis of the cleavage sites in the three species. Clusters of cleavage sites occurred in definite

Table 4. Sites of proteolytic cleavage of bovine, human and rat thyroglobulin. Sites located at closely related positions were grouped and, when present in at least two species, were named with letters (A–D). Unique sites include the cleavage sites not conserved between species. The cleavage sites are identified by the numbers of the residues at whose amino groups the cleavages occurred. Residue numbers were assigned by comparing the sequences of the peptides from each species to the respective cDNA-derived sequence. Rat peptides were compared with the cDNA-derived sequences of bovine (b) [1] and human (h) [2] origin.

Site	Bovine thyroglobulin		Human thyroglobulin		Rat thyroglobulin	
	thermolysin	trypsin	thermolysin	trypsin	thermolysin	trypsin
A	502		503		501(b), 502(h)	
	524	520		522	522(b), 523(h)	522(b), 523(h)
	532	526				527(b), 528(h)
B		993	982, 990			
C					1776(b), 1173(h)	1771(b), 1768(h)
	1784	1783			1784(b), 1782(h)	
			1831			1825(b), 1823(h)
D		2515		2513		2515(b), 2513(h)
Unique	240	240				
	1218	1142, 1184	597			
			1405	1627		

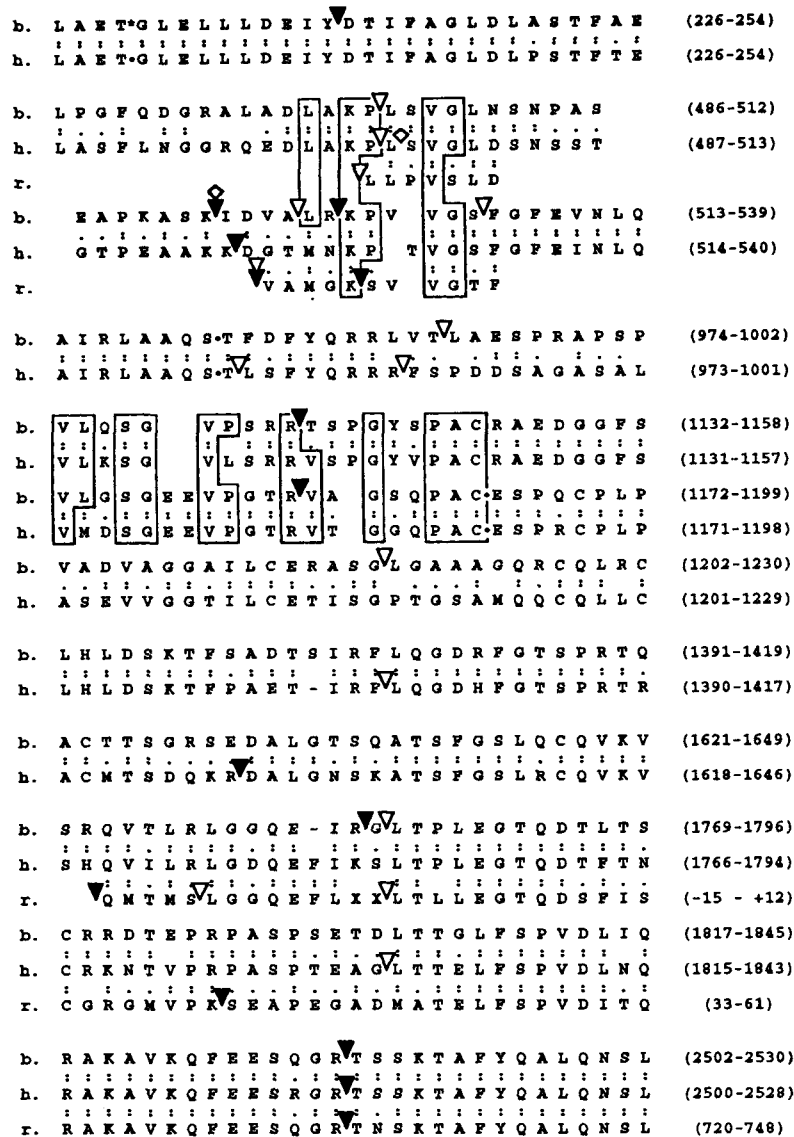
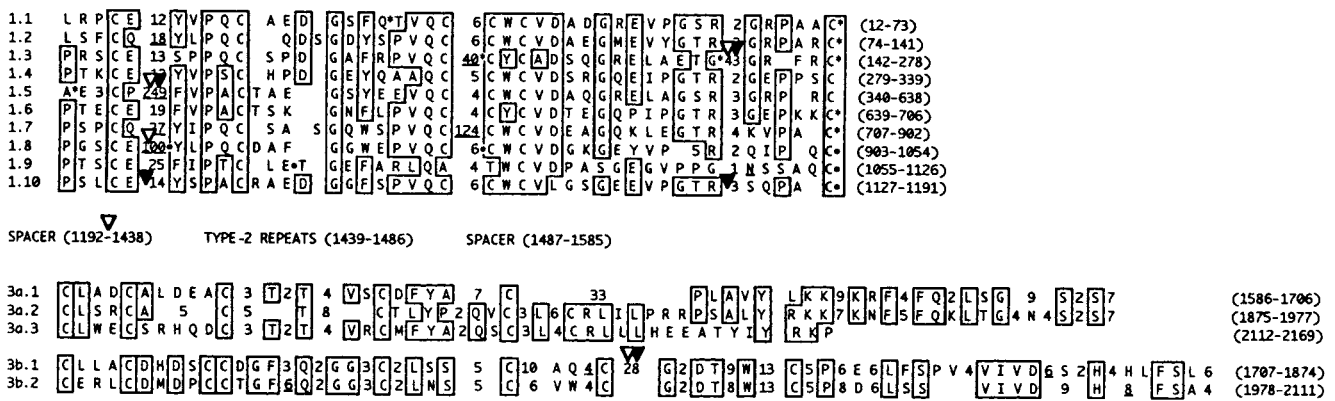


Fig. 7. Sites of proteolytic cleavage of human, bovine and rat thyroglobulin. Aligned cDNA-derived sequences of (b) bovine thyroglobulin [1], (h) human thyroglobulin [2], (r) rat thyroglobulin [3]. Amino acids are represented in single-letter code. Numbers identify the extremes of each segment of sequence. Negative numbers indicate the part of the rat thyroglobulin sequence determined by peptide sequencing (Table 3). Deletions between residues 1400–1401 of human thyroglobulin and 1780–1781 of bovine thyroglobulin, introduced for best alignment, are marked by dashes. The similarity scores between amino acids are indicated as follows: ":" = $\geq +1$; "." = $+1$ to -1 ; empty space = < -1 . The sites of cleavage are indicated: (∇) thermolysin; ($\blacktriangledown$) trypsin; ($\diamond$) native thyroglobulin [22]. Two repetitive sequences in the regions 490–540 and 1130–1200 are outlined: the repeated segments of the aligned sequences are stacked and similar residues are boxed. Gaps (empty spaces) have been arbitrarily inserted to optimize the alignment. Exon–intron junctions [30] are marked by dots if observed, and by asterisks if deduced from comparison to another species.

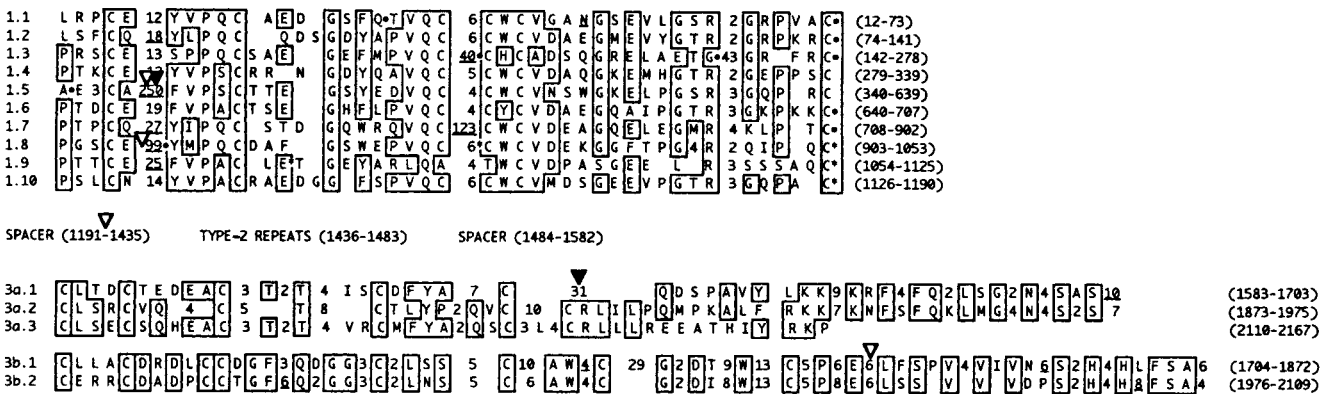
regions of the sequence. The sequence around the cleavage sites is shown in Fig. 7. All the cleavage sites reflected the protease specificities, except at residue 240 of bovine thyroglobulin. However, there were many potential target sites for thermolysin and trypsin all along the sequence that were not attacked. In all species, the cleavage sites of group A lay within or at the border of two short repetitive sequences. The tryptic cleavages of bovine thyroglobulin at residues 1142 and 1184 also occurred at similar sites within two previously unrecognized repeats within repeat 1.10. This observation agrees with the former proposal that the building block of the NH_2 -terminal part of thyroglobulin originated from the

duplication of a smaller unit [30]. Possibly these repetitive structures are especially sensitive to proteolysis. The tryptic cleavages at residues 520 of bovine thyroglobulin, 522 and 1627 of human thyroglobulin, and 1825(b) of rat thyroglobulin, occurred within the Lys-Xaa-Asp/Glu sequence. The absence of this sequence at the site of bovine thyroglobulin corresponding to human residue 1627 coincided with the lack of cleavage. Thermolysin cleaved bovine thyroglobulin at residue 1784, within the Gly-Leu-Thr sequence. Local divergence of the sequence may explain why human thyroglobulin was not cleaved at this site, but at residue 1831, where the Gly-Leu-Thr sequence also occurred.

A



B



C



Fig. 8. Repetitive structure of thyroglobulin. cDNA-derived sequences of (A) bovine thyroglobulin [1], (B) human thyroglobulin [2], (C) rat thyroglobulin [3]. Amino acids are represented in single-letter code. The two major types (type 1 and 3) of tandem repeats are aligned and the residues at their extremes are indicated by numbers in parentheses (negative numbers are used as in Fig. 7). Insertions and deletions were introduced to optimize the similarity. Conserved residues are boxed when present at least five times in type-1 and twice in type-3 repeats. Numbers within the sequence indicate the length of segments with no similarity between repeats. Underlined numbers or letters mark putative glycosylation sites [1–3] (see Table S5, Supplement). Regions separating different kinds of tandem repeats are named 'spacer'. Exon–intron junctions [30] are marked by dots if observed, and asterisks if deduced from comparison to other species. The cleavage sites are indicated: (▽) thermolysin; (▼) trypsin. The first 11 residues of bovine and human thyroglobulin and the acetylcholinesterase-similar domain [29], corresponding to residues 2181–2750 in bovine thyroglobulin, 2179–2748 in human thyroglobulin and 399–967 in rat thyroglobulin, are not shown.

The NH₂-terminal sequences of peptides r5_{TL} and r5_{TR} (Table 3) permitted us to extend the cDNA-derived sequence of rat thyroglobulin of 15 residues, two of which unidentified, toward the NH₂ terminus.

Relationship with the repetitive structure of thyroglobulin

Most cleavage sites had an interesting relationship with the repetitive structure of thyroglobulin, which is shown in Fig. 8. More than 75% of thyroglobulin at its NH₂-terminal side consists of tandem repeats rich in cysteine, including ten type-1 repeats, three short type-2 repeats, three type-3a repeats and two type-3b repeats. Unrelated sequences interrupt some repeats at fixed positions. A domain with 28% se-

quence similarity to the acetylcholinesterase of *Torpedo californica* constitutes the COOH-terminal one-fourth of thyroglobulin [29]. Fig. 8 also shows the exon–intron junctions in the amino-terminal half of thyroglobulin, most of which are either at the borders of the type-1 repeats, or within/at the borders of the inserted sequences that interrupt them [30].

The early cleavages of group A occurred with both enzymes in the long insert of repeat 1.5 of bovine and human thyroglobulin, and at similar positions of rat thyroglobulin. With bovine thyroglobulin, cleavages were also detected in the 40-residue insert of repeat 1.3 and in the 14-residue insert of repeat 1.10. Cleavage sites of group B lay in the major insert of repeat 1.8, while unique cleavages of bovine and human thyroglobulin with thermolysin occurred in the region

separating type-1 from type-2 repeats. Several sites of group C lay in the main insert of repeat 3b.1 of bovine thyroglobulin and in the corresponding region of rat thyroglobulin, while trypsin cleaved human thyroglobulin in the main insert of repeat 3a.1. Site D was located in the acetylcholinesterase-similar domain. Thus, most cleavage sites lay within sequences that disrupt the arrangement of the Cys-rich repeats. The sensitivity of these regions to proteolysis may reflect structural peculiarities that will be examined in the Discussion, also in view of the secondary structure predictions reported in the Supplement.

DISCUSSION

The present study on the limited proteolysis of bovine, human and rat thyroglobulin with thermolysin and trypsin demonstrates that thyroglobulin presents well localized and conserved regions highly sensitive to proteolysis.

Cleavage most consistently occurred with both enzymes at sites of group A (see Table 4), located in the vicinity of residue 500. Recent data show that the faster band of the electrophoretic doublet of thyroglobulin is produced by cleavages in the same region, at residue 520 in bovine thyroglobulin and 504 in human thyroglobulin [22]. Also the first cleavages of thyroglobulin by chymotrypsin occur at bovine residue 503 and human residue 504 [22]. The present data confirm that this is the region most sensitive to proteases in bovine, human, and also rat thyroglobulin. Cleavage sites of group B were common to bovine and human thyroglobulin. Another major cluster of sites of cleavage by both enzymes (group C) was located in the vicinity of residue 1800. The corresponding region of human thyroglobulin was only partially sensitive. Tryptic cleavage also occurred at site D in all species; the resulting COOH-terminal peptide may contain up to three hormonogenic tyrosines (2553, 2567 and 2746 in human thyroglobulin) [7, 8, 12–15]. Cleavage at site D was earlier in human thyroglobulin than in the others; this may reflect structural differences that may affect the priority of usage of the hormonogenic sites.

In the digestion of human thyroglobulin with thermolysin, there seemed to be two subsets of peptides $h2_{TL}$, one yielding peptides $h3_{TL}$ and $h4_{TL}$, and a larger one yielding peptide $h5_{TL}$ (Fig. 3). In the tryptic digestion of rat thyroglobulin, low amounts of peptide $r5_{TR}$ were produced as an alternative to peptides $r6_{TR}$ and $r7_{TR}$ (Fig. 6). Both cases may reflect the presence of a minor percentage of covalent dimers, besides non-covalent ones, which were seen in all species in non-reducing SDS/PAGE (not shown; see [22]).

Unique prominent cleavages occurred in bovine thyroglobulin in the region around residue 1180, whose predicted secondary structure was more irregular than in the rest of the molecule and in the same region of human thyroglobulin. Cleavages unique to human thyroglobulin occurred at residues 1405 with thermolysin and 1627 with trypsin. The latter site had no match in bovine thyroglobulin, but shared analogy with other tryptic sites. Similar patterns of peptides were obtained with bovine thyroglobulin containing 0.65% or 1.00% of iodine, on the one hand, and with human thyroglobulin whose iodine content ranged over 0.10–0.35%, on the other hand. These facts limit the possibility that the difference in the cleavage sites of bovine and human thyroglobulin only reflected differences in their iodine content.

The preferential cleavage at the NH_2 and $COOH$ termini may reflect the adaptation of thyroglobulin to efficient hor-

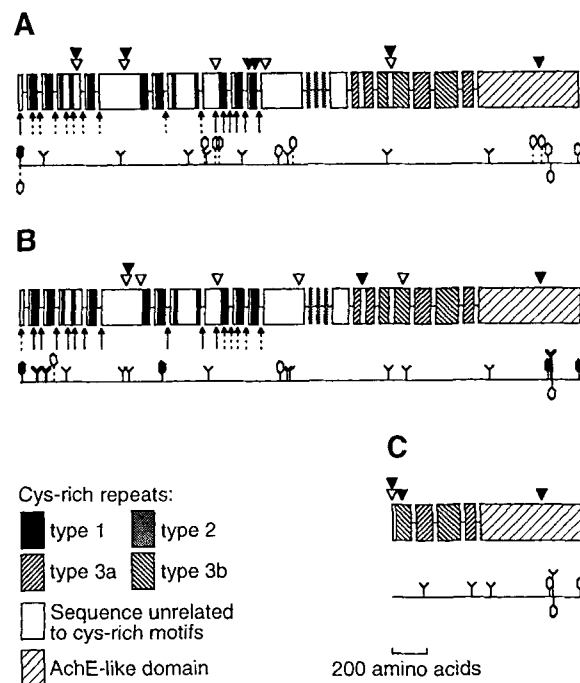


Fig. 9. Relationship between the cleavage sites and the repetitive structure and post-translational modifications of thyroglobulin. Diagrammatic representations of (A) bovine thyroglobulin, (B) human thyroglobulin, (C) rat thyroglobulin. The various kinds of repeated motifs and the other regions of thyroglobulin are represented as boxes. Empty boxes represent the unrelated sequences interrupting some repeats or separating different kinds of repeats. Spaces have been introduced for clarity between contiguous motifs and bear no relationship to the gene introns. AchE, acetylcholinesterase. The cleavage sites are shown: (∇) thermolysin; ($\blacktriangledown$) trypsin. Exon–intron junctions [30] are marked by tiny arrows with continuous stalks, if observed, and dashed stalks, if deduced from other species. Y, actual glycosylation site [27, 28]; Y, putative N-linked glycosylation site [1–3]. Only putative sites containing a consensus asparagine with positive hydrophilic score (see Table S5, Supplement) are shown. Filled hexagons with continuous stalks indicate observed hormonogenic Tyr residues, empty hexagons indicate those deduced from comparison to other species [7–15]. Hexagons with dashed stalks indicate hormonogenic donor Tyr residues [33–35].

mone release. The production of a limited number of hormone-bearing peptides by selective endoproteolytic cleavages may facilitate the release of tri- and tetra-iodothyronine, as suggested by the studies on the lysosomal degradation of thyroglobulin [16–18]. In hog thyroid lysosomes, hormone-depleted thyroglobulin was found that had a normal mobility in non-reducing SDS/PAGE, but resolved upon reduction into peptides of 50–260 kDa. This was interpreted to mean that selective endoproteolytic attack produced discrete fragments that remained bound by disulfide bonds, while the hormones at the thyroglobulin ends were removed by further exo- or endo-proteolytic action [16]. In the limited proteolysis of rabbit thyroglobulin with thyroid lysosomal cathepsins, tri- and tetra-iodothyronine were first released as peptide intermediates of 10–50 kDa [17]. Rabbit thyroglobulin was cleaved at residues 532, 795, 2487 (human numbering) by cathepsin B, at residues 551, 1835, 2468 by cathepsin D, and residues 2389, 2452, 2490 by cathepsin L [18]. Many of these cleavage sites are near those observed by us with different enzymes. Thus, the lysosomal degradation of thyroglobulin seems governed by the intrinsic susceptibility of some of its regions.

A striking characteristic of most cleavage sites was their location within the inserts that interrupt the Cys-rich tandem repeats of thyroglobulin (Figs 8 and 9). It was suggested that these inserts played a role in shaping the surface of thyroglobulin, due to the occurrence of exon–intron junctions within their context or at their borders, and to the presence of predicted glycosylation sites and amphipathic α -helices [30]. In fact, exon–intron junctions often map at protein surfaces and coincide with regions that contain sites of proteolytic cleavage [31], display greater sequence variation in a family of related sequences and form loops at the surface of proteins [32]. Our data provide experimental backing for this hypothesis.

The superficial location of the cleavage sites is also suggested by their relationship with tyrosyl residues accessible to iodination and putative glycosylation sites (Fig. 9). Tyrosines that undergo early iodination in human thyroglobulin [14] include Tyr972, near the cleavage sites of group B, and Tyr1447, not far from the cleavage site at residue 1405. Tyr239 is iodinated early in human [14] and turtle thyroglobulin [15]. Tyr2553, not far from cleavage site D, is hormonogenic [7, 8, 12–15]. Tyr2567 is iodinated early in human thyroglobulin [14] and is hormonogenic in hog [7]. ‘Donor’ tyrosines, which take part in hormonogenesis by transferring an iodophenoxyl group to an ‘acceptor’ iodotyrosine, were found at residues 986 or 1008 [33], and 2469 and/or 2522 [34] of bovine thyroglobulin, near cleavage sites B and D, respectively. A donor tyrosine was found at residue 1375 of bovine thyroglobulin that, if present in human thyroglobulin, would lie close to the cleavage site at residue 1405 [33]. Putative glycosylation sites containing an asparagine with positive hydrophilic score (Fig. 9; Table S5, Supplement) lay within 7–27 residues from the cleavage sites at bovine residues 502, 1142, 1783, 1784 and human residues 503, 522, 1831; other ones lay within 49–65 residues from the cleavage sites at bovine residues 993, human residues 982, 1405, 2513, and rat residue 2513.

Not all the cleavage sites were in predicted hydrophilic regions. Indeed, hydrophobicity plots can be used to predict the accessibility of strongly hydrophilic regions, but are less successful with those partially buried [26]. Two epitopes of mouse thyroglobulin recognized by autoreactive T cells were centered on the hormonogenic site corresponding to human residue 2553 [36]. Several epitopes were found in two screenings of recombinant peptides of human thyroglobulin with heteroantibodies [37, 38]. Only one of them (residues 1791–1862) lay near one of the cleavage sites reported here [37]. Although the antibodies were directed against native thyroglobulin and would have recognized exposed epitopes, recognition may have been incomplete. Fusion proteins from prokaryotic systems may lack post-translational and conformational epitopes, while some antibodies recognize the three-dimensional structure of thyroglobulin [39, 40]. Moreover, other structural requirements besides surface exposure may contribute to the proteolytic susceptibility.

Due to their content of Gly, Pro and Cys residues at fixed positions, the conserved regions of the Cys-rich repeats of thyroglobulin have a predicted structure mostly composed of β -strands and β -turns (Figs S1 and S2), in which many bends may form and be stabilized by disulfide bridges, yielding folding units with a rigid three-dimensional structure [41]. The Cys-rich repeats of thyroglobulin share structural analogy or partial sequence similarity with those found in various proteins, including the epidermal growth factor precursor [42], the low-density lipoprotein receptor [43], entactin [44],

the cell surface antigen epithelial glycoprotein [45], the laminin B1 chain [46], and the β subunits of the leukocyte adhesion proteins MAC-1, LFA-1 and p150,95 [47]. In some of these, the Cys-rich repeats are unusually stable and protease-resistant. This seems to be the case also with the Cys-rich motifs of thyroglobulin.

In contrast, the unrelated sequences that disrupt the repetitive organization of thyroglobulin are Cys-poor and may adopt a less constrained conformation. They contain mixtures of α -helices and β -strands, that may be part of globular folding units, and stretches of variable length with irregular structure, that may form connecting loops at the surface of thyroglobulin.

Thus, several circumstances, including the relationship with the splice junctions and with the pattern of sequence repetition, the presence of iodinated and hormonogenic Tyr residues and/or putative glycosylation sites, and the predicted secondary structure, suggest that the regions of the sequence surrounding the sites of proteolytic cleavage represent solvent-exposed regions, possibly forming loops at the surface of thyroglobulin. This is most evident for the major insert of repeat 1.5. Similar considerations apply, to variable extents, to the inserts of repeat 3b.1 and 1.8, the region separating type-1 from type-2 repeats, and the insert of repeat 3a.1. The exposed location of the region containing a cluster of hormonogenic Tyr residues near cleavage site D is also more than probable. Some of these regions also contain amphipathic α -helices, that will be described elsewhere.

We thank Drs. Jan Wolff, Danny Sackett and Bruno Di Jeso for reading the manuscript and providing useful suggestions; Dr. James Schilling, John Gardner and Kim Brown, at the Protein Structure Laboratory (University of California at Davis) for services in peptide sequencing; Franco D’Agnello, Mario Berardone and Fortunato Moscato for the photographs, and Giovanni Sequino for helping with the laboratory animals. This work was supported in part by *Progetto Finalizzato Biotecnologie e Biostrumentazione, Consiglio Nazionale delle Ricerche, Italy*.

REFERENCES

1. Mercken, L., Simons, M.-J., Swillens, S., Massaer, M. & Vassart, G. (1985) Primary structure of bovine thyroglobulin deduced from the sequence of its 8431-base complementary DNA, *Nature* **316**, 647–651.
2. Malthiery, Y. & Lissitzky, S. (1987) Primary structure of human thyroglobulin deduced from the sequence of its 8448-base complementary DNA, *Eur. J. Biochem.* **165**, 491–498.
3. Di Lauro, R., Obici, S., Condliffe, D., Ursini, V. M., Musti, A., Moscatelli, C. & Avvedimento, V. E. (1985) The sequence of 967 amino acids at the carboxyl end of rat thyroglobulin. Location and surroundings of two thyroxine-forming sites, *Eur. J. Biochem.* **148**, 7–11.
4. Musti, A. M., Avvedimento, E. V., Polistina, C., Ursini, V. M., Obici, S., Nitsch, L., Cocozza, S. & Di Lauro, R. (1986) The complete structure of the rat thyroglobulin gene, *Proc. Natl Acad. Sci. USA* **83**, 323–327.
5. Gavaret, J. M., Dème, D., Nunez, J. & Salvatore, G. (1977) Sequential reactivity of tyrosyl residues of thyroglobulin upon iodination catalyzed by thyroid peroxidase, *J. Biol. Chem.* **252**, 3281–3285.
6. Lamas, L., Taurog, A., Salvatore, G. & Edelhoch, H. (1974) Preferential synthesis of thyroxine from early iodinated tyrosyl residues in thyroglobulin, *J. Biol. Chem.* **249**, 2732–2737.
7. Marriq, C., Rolland, M. & Lissitzky, S. (1982) Structure-function relationship in thyroglobulin: amino acid sequence of two different thyroxine-containing peptides from porcine thyroglobulin, *EMBO J.* **1**, 397–401.

8. Marriq, C., Rolland, M. & Lissitzky, S. (1983) Amino acid sequence of the unique 3,5,3'-triiodothyronine-containing sequence from porcine thyroglobulin, *Biochem. Biophys. Res. Commun.* **112**, 206–213.
9. Rawitch, A. B., Chernoff, S. B., Litwer, M. R., Rouse, J. B. & Hamilton, J. W. (1983) Thyroglobulin structure-function. The amino acid sequence surrounding thyroxine, *J. Biol. Chem.* **258**, 2079–2082.
10. Lejeune, P.-J., Marriq, C., Rolland, M. & Lissitzky, S. (1983) Amino acid sequence around a hormonogenic tyrosine residue in the N-terminal region of human thyroglobulin after *in vivo* and *in vitro* iodination, *Biochem. Biophys. Res. Commun.* **114**, 73–80.
11. Rawitch, A. B., Litwer, M. R., Gregg, J., Dziadik Turner, C., Rouse, J. B. & Hamilton, J. W. (1984) The isolation of identical thyroxine containing amino acid sequences from bovine, ovine and porcine thyroglobulins, *Biochem. Biophys. Res. Commun.* **118**, 423–429.
12. Dunn, J. T., Anderson, P. C., Fox, J. W., Fassler, C. A., Dunn, A. D., Hite, L. A. & Moore, R. C. (1987) The sites of thyroid hormone formation in rabbit thyroglobulin, *J. Biol. Chem.* **262**, 16948–16952.
13. Fassler, C. A., Dunn, J. T., Anderson, P. C., Fox, J. W., Dunn, A. D., Hite, L. A., Moore, R. C. & Kim, P. S. (1988) Thyrotropin alters the utilization of thyroglobulin's hormonogenic sites, *J. Biol. Chem.* **263**, 17366–17371.
14. Lamas, L., Anderson, P. C., Fox, J. W. & Dunn, J. T. (1989) Consensus sequences for early iodination and hormonogenesis in human thyroglobulin, *J. Biol. Chem.* **264**, 13541–13545.
15. Roe, M. T., Anderson, P. C., Dunn, A. D. & Dunn, J. T. (1989) The hormonogenic sites of turtle thyroglobulin and their homology with those of mammals, *Endocrinology* **124**, 1327–1332.
16. Rousset, B., Selmi, S., Bornet, H., Bourgeat, P., Rabilloud, R. & Munari-Silem, Y. (1989) Thyroid hormone residues are released from thyroglobulin with only limited alteration of the thyroglobulin structure, *J. Biol. Chem.* **264**, 12620–12626.
17. Dunn, A. D., Crutchfield, H. E. & Dunn, J. T. (1991) Proteolytic processing of thyroglobulin by extracts of thyroid lysosomes, *Endocrinology* **128**, 3073–3080.
18. Dunn, A. D., Crutchfield, H. E. & Dunn, J. T. (1991) Thyroglobulin processing by thyroidal proteases. Major sites of cleavage by cathepsin B, D and L, *J. Biol. Chem.* **266**, 20198–20204.
19. Gentile, F. (1992) Regions of the primary structure at the outer surface of human, bovine and rat thyroglobulin, *J. Endocrinol. Invest.* **15**, Suppl. 2, p. 114 (abstr. 133).
20. Palumbo, G., Tecce, M. F. & Ambrosio, G. (1982) A non-incinerative rate-sensing method for the determination of iodine in iodoproteins, *Anal. Biochem.* **123**, 183–189.
21. van der Walt, B., Kotzé, B., Edelhoch, H. & van Jaarsveld, P. P. (1983) Characterization of the major polypeptide chains of reduced bovine thyroglobulin, *Biochim. Biophys. Acta* **744**, 90–98.
22. Gentile, F., Palumbo, G. & Salvatore, G. (1992) The origin of the electrophoretic doublet of thyroglobulin, *Biochem. Biophys. Res. Commun.* **186**, 1185–1191.
23. Schwartz, R. M. & Dayhoff, M. O. (1979). Matrices for detecting distant relationships, in *Atlas of protein sequence and structure* (M. O. Dayhoff, ed.) vol. 5, suppl. 3, p. 353, National Biomedical Research Foundation, Washington DC.
24. Chou, P. Y. & Fasman, G. D. (1978) Prediction of the secondary structure of proteins from their amino acid sequence, *Adv. Enzymol. Relat. Areas Mol. Biol.* **47**, 45–148.
25. Garnier, J., Osguthorpe, J. D. & Robson, B. (1978) Analysis of the accuracy and the implications of simple methods for predicting the secondary structure of globular proteins, *J. Mol. Biol.* **120**, 97–120.
26. Kyte, J. & Doolittle, R. F. (1982) A simple method for displaying the hydropathic character of a protein, *J. Mol. Biol.* **157**, 105–132.
27. Franc, J.-L., Venot, N. & Marriq, C. (1990) Characterization of the two oligosaccharides present in the preferential hormonogenic domain of human thyroglobulin, *Biochem. Biophys. Res. Commun.* **166**, 937–944.
28. Rawitch, A. B., Liao, T. H. & Pierce, J. G. (1968) The amino acid sequence of a tryptic glycopeptide from human thyroglobulin, *Biochim. Biophys. Acta* **160**, 360–367.
29. Schumacher, M., Camp, S., Maulet, Y., Newton, M., MacPhee-Quigley, K., Taylor, S. S., Friedmann, T. & Taylor, P. (1986) Primary structure of *Torpedo californica* acetylcholinesterase deduced from its cDNA sequence, *Nature* **319**, 407–409.
30. Parma, J., Cristophe, D., Pohl, V. & Vassart, G. (1987) Structural organization of the 5' region of the thyroglobulin gene. Evidence for intron loss and "exonization" during evolution, *J. Mol. Biol.* **196**, 769–779.
31. Craik, C. S., Sprang, S., Fletterick, R. & Rutter, W. J. (1982) Intron-exon splice junctions map at protein surfaces, *Nature* **299**, 180–182.
32. Craik, C. S., Rutter, W. J. & Fletterick, R. (1983) Splice junctions: association with variation in protein structure, *Science* **220**, 1125–1129.
33. Ohmiya, Y., Hayashi, H., Kondo, T. & Kondo, Y. (1990) Location of dehydroalanine residues in the amino acid sequence of bovine thyroglobulin. Identification of "donor" tyrosine sites for hormonogenesis in thyroglobulin, *J. Biol. Chem.* **265**, 9066–9071.
34. Palumbo, G. (1987) Thyroid hormonogenesis. Identification of a sequence containing iodophenyl donor site(s) in calf thyroglobulin, *J. Biol. Chem.* **262**, 17182–17188.
35. Marriq, C., Lejeune, P.-J., Venot, N. & Vinet, L. (1991) Hormone formation in the isolated fragment 1–171 of human thyroglobulin involves the couple tyrosine 5 and tyrosine 130, *Mol. Cell. Endocrinol.* **81**, 155–164.
36. Champion, B. R., Page, K. R., Parish, N., Rayner, D. C., Dawe, K., Biswas-Hughes, G., Cooke, A., Geysen, M. & Roitt, I. M. (1991) Identification of a thyroxine-containing self-epitope of thyroglobulin which triggers thyroid autoreactive T cells, *J. Exp. Med.* **174**, 363–370.
37. Dong, Q., Ludgate, M. & Vassart, G. (1989) Towards an antigenic map of human thyroglobulin: identification of ten epitope-bearing sequences within the primary structure of thyroglobulin, *J. Endocrinol.* **122**, 169–176.
38. Henry, M., Malthiery, Y. & Zanelli, E. (1990) Epitope mapping of human thyroglobulin. Heterogeneous recognition by thyroid pathologic sera, *J. Immunol.* **145**, 3692–3698.
39. Kondo, E. & Kondo, Y. (1984) Monoclonal antibodies to hog thyroglobulin recognizing disulfide-dependent conformational structures, *Mol. Immunol.* **21**, 581–588.
40. De Baets, M. H., Theunissen, R., Kok, K., de Vijlder, J. J. M. & van Breda Vriesman, P. J. C. (1987) Monoclonal antibodies to human thyroglobulin as probes for thyroglobulin structure, *Endocrinology* **120**, 1104–1111.
41. Malthiery, Y., Bergè-Lefranc, J.-L., Franc, J.-L., Henry, M., Lejeune, P.-J., Ruf, J. & Lissitzky, S. (1989) Thyroglobulin structure and function: recent advances, *Biochimie* **71**, 195–210.
42. Doolittle, R. F. (1985) The genealogy of some recently evolved vertebrate proteins, *Trends Biochem. Sci.* **10**, 233–237.
43. Yamamoto, T., Davis, C. G., Brown, M. S., Schneider, W. J., Casey, M. L., Goldstein, J. L. & Russell, D. W. (1984) The human LDL receptor: a Cys-rich protein with multiple Alu sequences in its mRNA, *Cell* **39**, 27–38.
44. Durkin, M. E., Chakravarti, S., Bartos, B. B., Liu, S.-H., Friedman, R. L. & Chung, A. E. (1988) Amino acid sequence and domain structure of entactin. Homology with epidermal growth factor precursor and low density lipoprotein receptor, *J. Cell. Biol.* **107**, 2749–2756.
45. Simon, B., Podolsky, D. K., Moldenhauer, G., Isselbacher, K. J., Gattoni-Celli, S. & Brand, S. J. (1990) Epithelial glycoprotein is a member of a family of epithelial cell surface antigens homologous to nidogen, a matrix adhesion protein, *Proc. Natl Acad. Sci. USA* **87**, 2755–2759.

46. Sasaki, M., Kato, S., Kohno, K., Martin, G. R. & Yamada, Y. (1987) Sequence of the cDNA encoding the laminin B1 chain reveals a multidomain protein containing Cys-rich repeats. *Proc. Natl Acad. Sci. USA* 84, 935–939.
47. Kishimoto, T. K., O'Connor, K., Lee, A., Roberts, T. M. & Springer, T. A. (1987) Cloning of the β subunit of the leukocyte adhesion proteins: homology to an extracellular matrix receptor defines a novel supergene family. *Cell* 48, 681–690.

Supplementary material to:

Preferential sites of proteolytic cleavage of bovine, human and rat thyroglobulin

The use of limited proteolysis to detect solvent-exposed regions of the primary structure.

Fabrizio GENTILE and Gaetano SALVATORE

RESULTS

NH₂-terminal microsequence analysis of the proteolytic peptides of thyroglobulin

The yields of the phenylthiohydantoin derivatives for sequential steps of Edman degradation of the peptides produced by limited proteolysis of bovine, human and rat thyroglobulin are reported in Tables S1–S4.

Secondary structure predictions

Full-length predictions of the secondary structure of bovine thyroglobulin and human thyroglobulin are displayed in Figs S1 and S2, respectively. Local predictions for the regions surrounding the cleavage sites in different species are compared in Fig. S3.

The structures predicted for the Cys-rich motifs differed noticeably from those of other regions of thyroglobulin. In the NH₂-terminal half of the molecule, the type-1 repeats consisted mainly of β -strands and β -turns; short α -helices occurred in the minor inserts between the first two Cys residues of the motif. Predicted structures for type-2 and type-3 repeats were also mostly of β -type, with the exception of repeat 3a.2 of human thyroglobulin. A more varied composition, consisting of stretches of α -helix interspersed with β -strands and regions of unpredicted, irregular structure, characterized the major inserts of type-1 repeats that harbored sites of proteolytic cleavage. These included: (a) the 43-residue insert of repeat 1.3, which was cleaved in bovine thyroglobulin; (b) the major insert of repeat 1.5, in which there were cleavage sites of group A associated with stretches of α -helix; (c) the major insert of repeat 1.8, in which there were cleavage sites of group B associated with β -strands.

The acetylcholinesterase-similar domain also consisted of α -helices, β -strands and regions of undefined, irregular structure. Cleavage site D lay between stretches of α -helix and β -strand. The region around residue 1218 of bovine thyroglobulin contained long stretches of unpredictable secondary structure, whereas several β -strands and β -turns were detected in the corresponding region of human thyroglobulin. The cleavage site at position 1405 of human thyroglobulin lay within a β -strand. The possible influence of differences of the secondary structure on the sensitivity of different regions of thyroglobulin to proteolysis is examined in the Discussion.

Table S1. NH₂-terminal microsequence analysis of the peptides produced by the limited proteolysis of bovine thyroglobulin with thermolysin. The phenylthiohydantoin derivatives detected by HPLC analysis for subsequent cycles of Edman degradation are shown. Italics indicate less than certain identification. The numbers that follow the residues are the yields in picomoles. "tr" indicates traces. b3_{TL}^a was obtained by a 30-min digestion at enzyme/substrate = 1:1000 (by mass); b3_{TL}^b was obtained by a 80-min digestion at enzyme/substrate = 1:100 (by mass).

Cycle	b1 _{TL}	b2 _{TL}	b3 _{TL} ^a		b3 _{TL} ^b		b4 _{TL}		b5 _{TL}	b6 _{TL}	b7 _{TL}	b7bis _{TL}	b8 _{TL}	b9 _{TL}	b9bis _{TL}	b9ter _{TL}
			b3 _{TL}	b3bis _{TL}	b3ter _{TL}	b3bis _{TL}	b4 _{TL}	b4bis _{TL}								
1	Asn 10	Leu 24	Asn 6	—	—	Leu 5	Asp 3	Leu 45	Phe 12	Leu 26	Leu 25	Leu 30	Leu 30	Leu 50	—	Phe 13
2	Ile 12	Ser 15	Ile 2	—	—	Ala 4	—	Arg 8	Gly 7	Gly 16	Gly 20	Gly 15	Gly 22	Thr 15	Arg 10	Gly 12
3	Phe 10	Val 10	Phe 4	Ile 4	—	Glu 3	—	Lys 18	Phe 6	Ala 14	Ala 15	Ala 14	Ala 14	Pro 12	Lys 15	Ala 12
4	Glu 10	Gly 14	Glu 1	Phe 2	—	Ser 2	Phe 3	Pro 23	Glu 6	Ala 18	Ala 17	Ala 15	Ala 15	Leu 10	Pro 23	Ala 13
5	Tyr 1	Leu 17	—	Ala 5	Pro 4	Ala 2	Val 25	—	—	Ala 18	Ala 19	Ala 16	Ala 16	Glu 8	Val 22	Ala 14
6	Gln 8	Asn 4	—	Gly 3	Arg 4	Gly 3	Val 22	Asn 4	Gly 14	Gly 15	—	—	—	Gly 8	Val 25	Gly 15
7	—	—	Val 2	Leu 4	Ala 3	Leu 1	Gly 20	Leu 14	Gln 13	Gln 12	—	—	—	Thr 4	Gly 20	Gln 10
8	Asn 4	—	—	—	Pro 3	Asp 1	Ser 7	Gln 7	Arg 3	Arg 4	—	—	—	Gln 5	Ser 8	Arg 4
9	—	—	—	—	Ser tr	—	Phe 15	Glu 8	—	—	—	—	—	Asp 5	Phe 15	Glu 7
10	—	—	—	—	Pro tr	Ala tr	Gly 15	Asn 4	Gln 6	—	—	—	—	Thr 5	—	—
11	—	—	—	—	Val 1	—	—	—	—	—	—	—	—	Leu 15	—	—

Table S2. NH₂-terminal microsequence analysis of the peptides produced by the limited proteolysis of bovine thyroglobulin with trypsin. Lettering and numbers as in Table S1.

Cycle	b1 _{TR}	b2 _{TR}	b3 _{TR}		b4 _{TR}		b5 _{TR}	b6 _{TR}	b7 _{TR}	b8 _{TR}	b9 _{TR}	b10 _{TR}	b10bis _{TR}
			b3 _{TR}	b3bis _{TR}	b4 _{TR}	b4bis _{TR}							
1	Asn 5	Ile 8	Asn 4	Asp 3	Lys 10	Ile 4	Thr 5	Val 4	Thr 12	Gly 30	Gly 22	Thr 50	Lys 25
2	Ile 5	Asp 10	Ile 2	Thr 3	Pro 12	Asp 4	Ser 7	Ala 4	Ser 10	Leu 7	Leu 10	Ser 18	Pro 25
3	Phe 5	Val 10	Phe 3	Ile 3	Val 14	—	Pro 7	Gly 4	Pro 14	Thr 8	Thr 5	Ser tr	Val 20
4	Glu 3	Ala 9	Glu 5	Phe 3	Val 13	Ala 5	Gly 7	Ser 3	Gly 11	Pro 10	Pro 10	Lys 17	Val 20
5	Tyr tr	Leu 12	—	Ala 4	Gly 10	Leu 6	Tyr 5	Gln 3	Tyr 6	Leu 6	Leu 7	Thr 12	Gly 20
6	Gln 3	Arg 2	Gln 5	Gly 3	Ser 8	—	Ser 6	Pro 3	Ser 10	Glu 6	Glu 6	Ala 18	Ser 5
7	Val 1	Lys 4	Val 4	Leu 3	Phe 8	Lys 3	Pro 10	Ala 2	Pro 12	Gly 8	Gly 6	Phe 18	—
8	Asp 2	Pro 25	—	—	Gly 10	Pro 6	Ala 9	—	Ala 7	Thr 4	Thr 7	—	—
9	—	—	—	—	Phe 8	Val 4	—	—	—	Gln 4	Gln 5	—	—
10	—	—	—	—	Glu 6	Val 5	—	—	—	Asp 3	Asp 4	—	—
11	—	—	—	—	—	—	Ala 12	—	Ala 6	—	—	—	—
12	—	—	—	—	—	—	Glu 8	—	Glu 5	—	—	—	—

Table S3. NH₂-terminal microsequence analysis of the peptides produced by the limited proteolysis of human thyroglobulin. Lettering and numbers as in Table S1.**(A) Digestion with thermolysin**

Cycle	h1 _{TL}		h2 _{TL}		h3 _{TL}		h4 _{TL}				h5 _{TL}		h6 _{TL}				h7 _{TL}	
							h4 _{TL}		h4bis _{TL}				h6 _{TL}		h6bis _{TL}			
1	Asn	37	Leu	45	Phe	8	Leu	11	Phe	10	Leu	16	Leu	34	Asn	9	Leu	14
2	Ile	37	Ser	18	Val	4	Ser	tr	—	—	—	—	Gln	18	Ile	7	—	—
3	Phe	36	Val	28	Pro	6	Phe	9	Pro	11	Gly	16	Gly	20	Phe	7	—	—
4	Glu	34	Gly	36	Ser	2	Tyr	7	—	—	Asp	20	Asp	12	Glu	5	Glu	8
5	Tyr	12	Leu	39	—	—	Gln	2	Asp	2	His	6	His	1	—	—	Leu	11
6	Gln	32	Asp	27	Thr	1					Phe	11	Phe	11	Gln	3	Phe	23
7													Gly	14	Val	4	Ser	8

(B) Digestion with trypsin

Cycle	h1 _{TR}		h2 _{TR}		h3 _{TR}		h4 _{TR}				h5 _{TR}		h6 _{TR}		h7 _{TR}	
1	Asn	23	Asp	14	Asn	14	Asn	23	Thr	31	Asp	13	Asp	22	Asp	23
2	Ile	29	Gly	10	Ile	13	Ile	27	Ser	30	Gly	9	Gly	24	Ala	24
3	Phe	27	Thr	4	Phe	16	Phe	10	—	—	Thr	2	Thr	11	Leu	23
4	Glu	19	Met	4	Glu	10	Glu	12	Lys	6	Met	5	Met	12	Gly	21
5	Tyr	12	Asn	8	—	—	—	—	Thr	11	Asn	4	Asn	13	Asn	14
6	Gln	19	Lys	7	Gln	12	Gln	14	Ala	16	Lys	3	Lys	10	Ser	14
7			Pro	24	Val	12					Pro	7			Lys	11
8					Asp	7										

Table S4. NH₂-terminal microsequence analysis of the peptides produced by the limited proteolysis of rat thyroglobulin. Lettering and numbers as in Table S1.**(A) Digestion with thermolysin**

Cycle	r1 _{TL}		r2 _{TL}		r3 _{TL}				r4 _{TL}		r5 _{TL}				r6 _{TL}	
					r3 _{TL}		r3bis _{TL}				r5 _{TL}		r5bis _{TL}			
1	Asn	10	Leu	7	Asn	1	Ala	3	Leu	8	Leu	23	—	—	Val	11
2	Ile	5	<i>Leu</i>	5	<i>Ile</i>	1	—	—	<i>Leu</i>	6	Gly	10	Thr	8	Ala	8
3	Phe	5	Pro	6	Phe	2	—	—	Pro	6	<i>Gly</i>	9	Leu	3	Met	7
4	Glu	4	Val	5	<i>Glu</i>	1	—	—	Val	5	Gln	6	Leu	7	—	—
5	<i>Tyr</i>	tr	<i>Ser</i>	3	—	—	<i>Gln</i>	1	<i>Ser</i>	5	Glu	10	—	—	Lys	5
6			Leu	8	<i>Gln</i>	2	—	—	Leu	7	Phe	6	Gly	8	—	—
7			Asp	5	<i>Val</i>	1	<i>Arg</i>	1	Asp	4	Leu	9	Thr	6	Val	6
8															Val	8

(B) Digestion with trypsin

Cycle	r1 _{TR}		r2 _{TR}		r3 _{TR}				r4 _{TR}				r5 _{TR}		r6 _{TR}		r7 _{TR}		r8 _{TR}				r9 _{TR}					
					r3 _{TR}		r3bis _{TR}		r4 _{TR}		r4bis _{TR}								r4ter _{TR}		r8 _{TR}				r8bis _{TR}		r8ter _{TR}	
1	—	—	Val	16	Asn	4	Ala	2	Asn	45	—	—	—	—	Asp	30	Val	16	Gln	13	Val	20	Ser	14	Pro	14	Ser	20
2	Ile	3	Ala	16	—	—	Thr	2	Ile	35	Asn	17	Val	14	Ile	6	Ala	16	Met	14	Ala	11	Val	11	—	—	Glu	21
3	Phe	5	Met	10	—	—	—	—	Phe	26	—	—	—	—	Gln	6	Met	13	Thr	13	Met	8	Val	12	—	—	Ala	14
4	Glu	2	—	—	Glu	3	Pro	2	Glu	27	Lys	29	—	—	Gly	4	Gly	14	Met	10	Gly	16	—	—	—	—	Pro	11
5	—	—	Lys	5	—	—	—	—	Tyr	9	Thr	15	—	—	Thr	6	Lys	7	Ser	11	Lys	6	Thr	9	—	—	Glu	13
6	Gln	2	Ser	11	Gln	2	—	—	Gln	22	Ala	12	—	—	Ser	3	Ser	9	Leu	14	—	—	Phe	10	—	—	Gly	14
7	—	—	Val	13	Val	2	—	—	Val	13	—	—	—	—	Val	12	—	—	—	—	—	—	—	—	—	—	—	—

Hydrophilicity predictions

Regions with predicted hydrophilic and hydrophobic character alternate continuously along the thyroglobulin sequence [2]. Hydrophilicity scores in the regions around the cleavage sites varied from slightly negative to positive (Fig. S3). The site of cleavage by thermolysin at human residue 990, which was attacked later than that at bovine residue 993, was associated with a more hydrophilic profile. Also the unique site of tryptic cleavage at bovine residue

1142 was more hydrophilic than the corresponding site of human thyroglobulin. In Table S5, the hydrophilicity scores of the consensus asparagine residues at the putative glycosylation sites of bovine, human and rat thyroglobulin are reported.

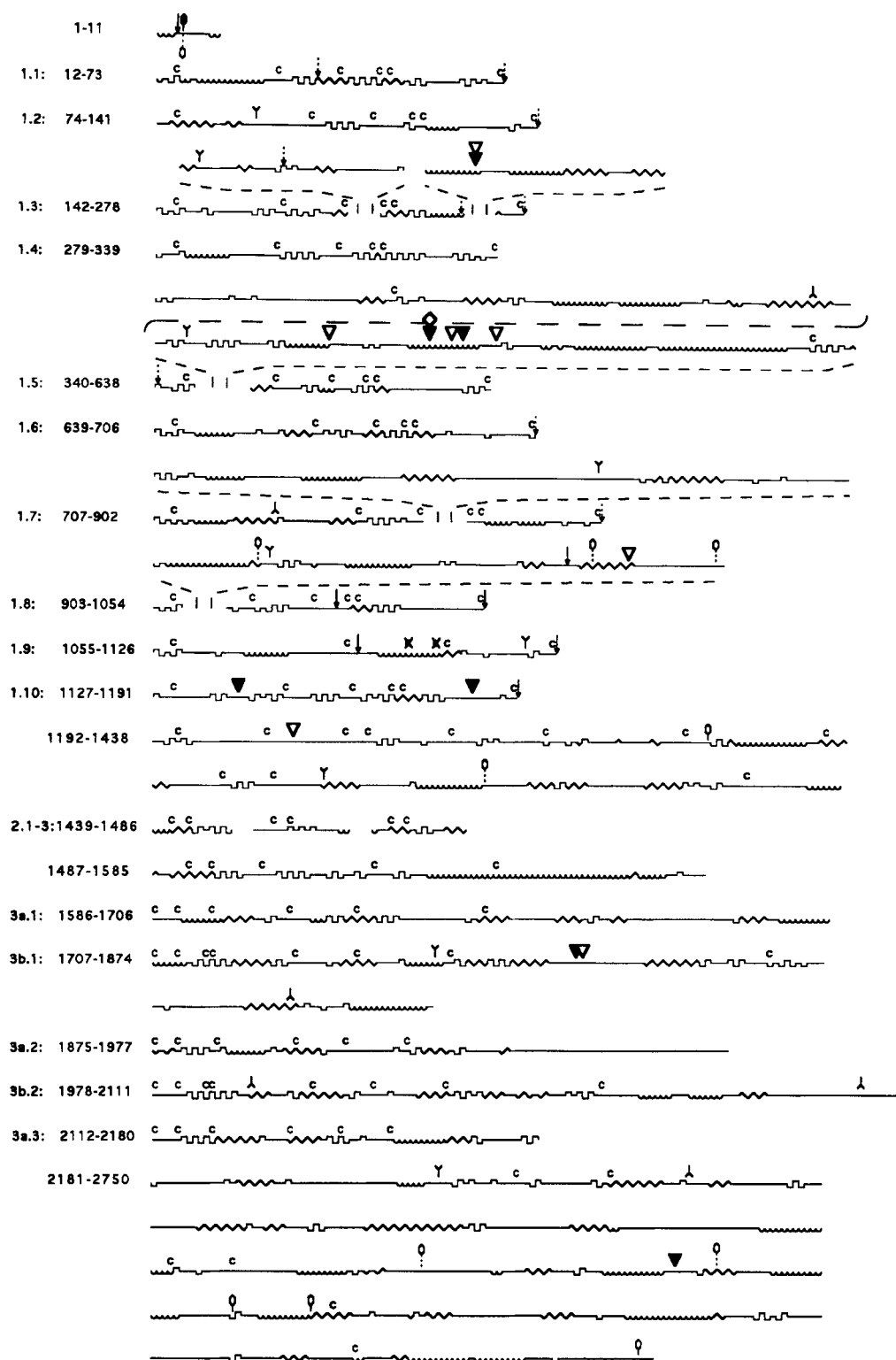


Fig.S1. Secondary structure prediction of bovine thyroglobulin. The prediction was obtained by a combination of the methods by Chou and Fasman [24] and Garnier et al. [25]; assignments were made only when both methods agreed. The Cys-rich repeats of thyroglobulin are shown on separate lines. The numbers at left indicate the type of repeat and the amino acid residue numbers at its extremities. The five major inserts in type-1 repeats are also shown on separate lines, above the respective repeats. The various types of secondary structure are represented as follows: α -helix; β -strand; β -turns; — regions where predictions could not be made. Cleavage sites are marked: ∇ thermolysin; $\blacktriangledown$ trypsin; $\blacklozenge$ native thyroglobulin [22]. Other symbols: c cysteine residue; X amino acid substitution at a site occupied by cysteine in similar repeats; Y putative N-linked glycosylation site [1] containing a consensus asparagine with positive or λ negative hydrophilic score (Table S5); ϕ observed or ϕ supposed hormonogenic tyrosine [7-15]; ϕ donor tyrosine [33, 34]; $\downarrow$ exon-intron junction detected in bovine or in $\downarrow$ human thyroglobulin [30].

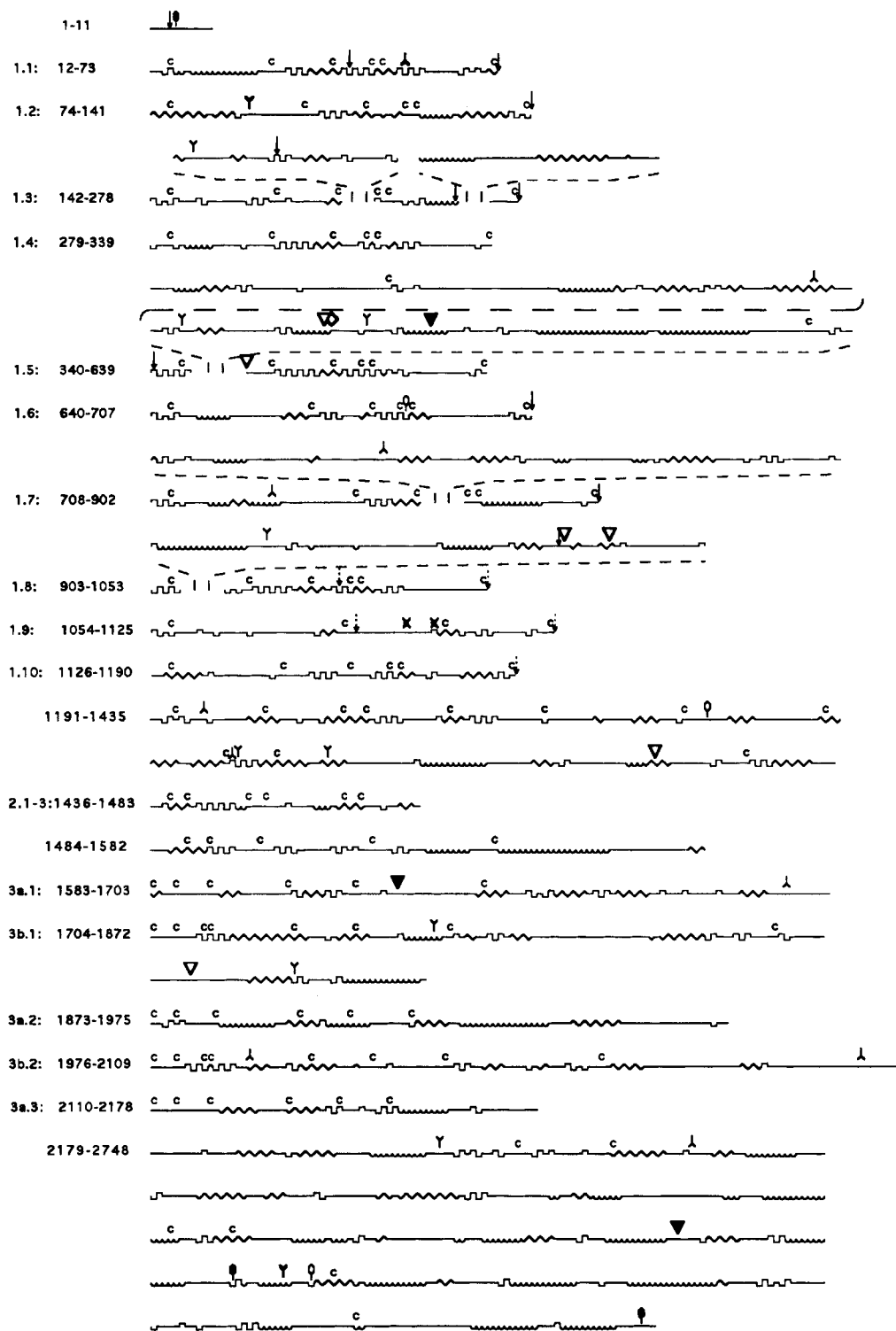


Fig.S2. Secondary structure prediction of human thyroglobulin. The prediction was obtained by a combination of the methods by Chou and Fasman [24] and Garnier et al. [25]; assignments were made only when both methods agreed. The Cys-rich repeats of thyroglobulin are shown on separate lines. The numbers at left indicate the type of repeat and the amino acid residue numbers at its extremities. The five major inserts in type-1 repeats are also shown on separate lines, above the respective repeats. The various types of secondary structure are represented as follows: α -helix; β -strand; β -turns; — regions where predictions could not be made. Cleavage sites are marked: ∇ thermolysin; $\triangledown$ trypsin; $\diamond$ native thyroglobulin [22]. Other symbols: c cysteine residue; X amino acid substitution at a site occupied by cysteine in similar repeats; Y putative or Y actual N-linked glycosylation site [1] containing an asparagine with positive or λ , Λ negative hydrophilic score (Table S5); ϕ observed or ϕ supposed hormonogenic tyrosine [7-15]; ϕ donor tyrosine [35]; $\downarrow$ exon-intron junction detected in human or in $\downarrow$ bovine thyroglobulin [30].

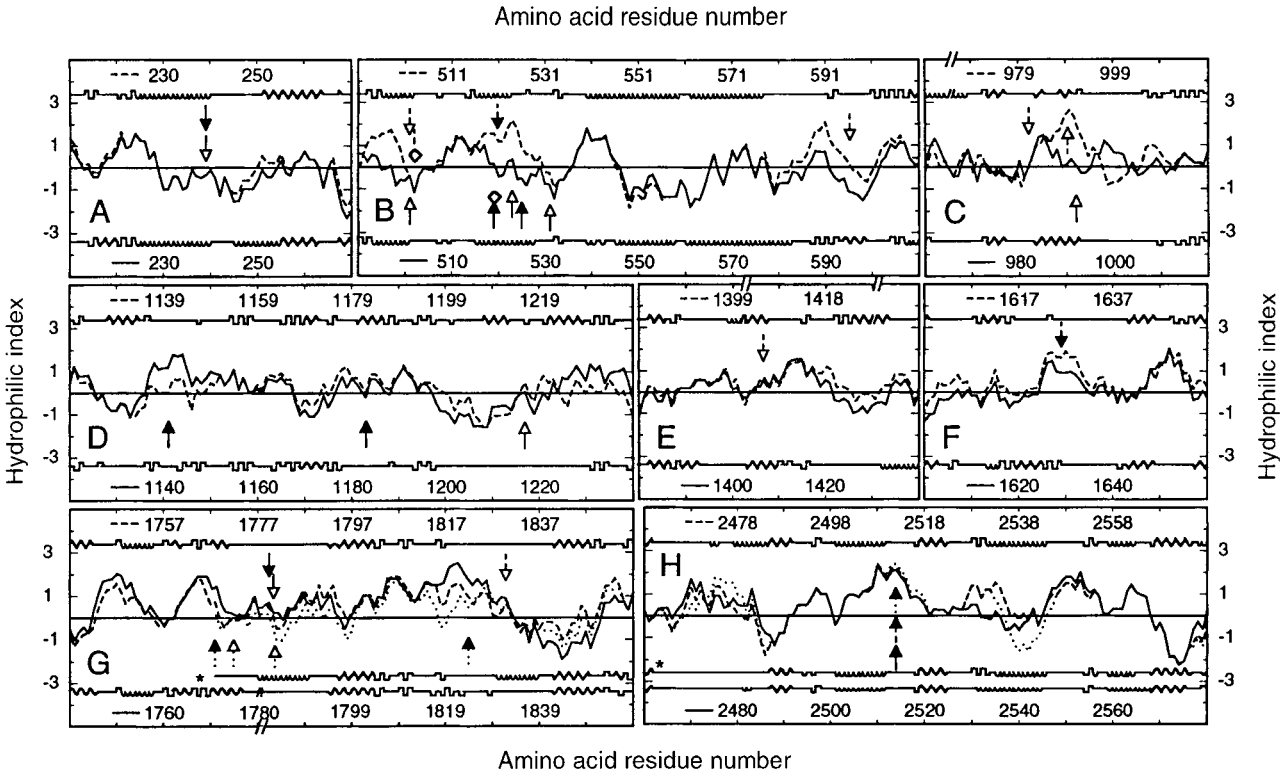


Fig.S3. Hydrophilicity plots and secondary structure predictions in the regions of thyroglobulin surrounding the cleavage sites. Hydrophilicity plots, obtained by use of the hydropathy scale of Kyte and Doolittle [26], with reversed signs, are shown in the middle of each panel. Continuous lines refer to bovine thyroglobulin, dashed lines to human thyroglobulin, and dotted lines to rat thyroglobulin. Residue numbers are reported on the lower x axis for bovine thyroglobulin, on the upper x axis for human thyroglobulin, and are not shown for rat thyroglobulin. For best alignment, gaps were introduced between residues 1780–1781 of bovine thyroglobulin, and 965–966, 1401–1402, and 1428–1429 of human thyroglobulin. Secondary structure predictions, obtained by a combination of the methods by Chou and Fasman [24] and Garnier et al. [25], are shown near the lower x axis for bovine thyroglobulin, right above the latter for rat thyroglobulin (marked by asterisks) and near the upper x axis for human thyroglobulin. The types of secondary structure are represented as in Figs S1 and S2. The cleavage sites are shown: (▽) thermolysin; (▼) trypsin; (◇) native thyroglobulin [22]; stalks extending from the symbols indicate the species in which cleavages occurred, according to the same convention as for hydrophilicity plots.

Table S5. Putative N-linked glycosylation sites of thyroglobulin. Residue numbering by alignment to the cDNA-derived sequences of bovine [1], human [2], or rat thyroglobulin [3]. K-D index: hydrophilic score calculated by the hydropathy scale of Kyte and Doolittle [26], with reversed signs. Numbers in parentheses simply indicate the alignment of unique putative sites of rat thyroglobulin with the corresponding positions in bovine and human thyroglobulin. Actual glycosylation was reported at sites 57, 91 [27] and 2562 of human thyroglobulin [28].

Bovine thyroglobulin		Human thyroglobulin		Rat thyroglobulin		Bovine thyroglobulin		Human thyroglobulin		Rat thyroglobulin	
residue	K-D index	residue	K-D index	residue	K-D index	residue	K-D index	residue	K-D index	residue	K-D index
		57	−0.46			1346	+0.16	1345	+0.19		
91	+0.26	91	+0.80					1696	−0.03		
179	−0.01	179	+0.10			1757	+0.56	1754	+0.26		
464	−0.44	465	−0.44			(1849)	−	(1847)	−	65	−0.04
476	+0.52	477	+0.56			1851	−0.77	1849	0		
		510	+0.79			(1920)	−	(1918)	−	136	+0.80
728	−0.41	729	−1.02			1995	−1.02	1993	−1.02	211	−0.81
		797	−0.83			2104	−1.06	2102	−0.73	321	−0.26
834	+0.97					(2151)	−	(2149)	−	368	+1.24
928	+0.60	928	+0.27			2232	+1.20	2230	+1.02	450	+0.38
1121	+1.23					2277	−0.46	2275	−0.71	495	−1.02
		1200	−0.18			(2426)	−	(2424)	−	644	−0.03
		1328/29	−0.6/+0.3			(2564)	−	2562	+1.42	782	+1.42