

The revised 8307 base pair coding sequence of human thyroglobulin transiently expressed in eukaryotic cells

Simone A R van de Graaf, Erwin Pauws, Jan J M de Vijlder and Carrie Ris-Stalpers

Pediatric and Neonatal Endocrinology, Academic Medical Centre, University of Amsterdam, Amsterdam, The Netherlands

(Correspondence should be addressed to S A R van de Graaf, Academic Medical Centre G2-123, Laboratory of Pediatric and Neonatal Endocrinology, PO Box 22700, 1100 DE, Amsterdam, The Netherlands)

Abstract

We developed a transient transfection system for human thyroglobulin (TG) cDNA in both human thyroid cells and in COS-1 cells.

Four overlapping TG cDNA fragments were amplified by reverse transcription-PCR from RNA of normal thyroid tissue. The most 5' fragment includes the natural translation initiation site and the sequence encoding the signal peptide (SP). After subcloning, the nucleotide sequence was determined and compared with the published human sequence, resulting in the detection of 30 nucleotide variations. For validation purposes, all variations were screened in 6–12 normal human alleles. Twenty-one were present in all screened alleles and have to be revised in the published nucleotide sequence. Since one variation concerns a triplet insertion, the coding sequence of the mature human thyroglobulin is 8307 nucleotides encoding 2750 amino acids.

The TG cDNA constructs were transiently transfected in HTori 3 and COS-1 cells and protein expression was detected using a polyclonal anti-human-TG on fixed cells and after SDS-PAGE. In both cell-lines all four TG protein fragments were expressed. The mannose structures detected on the proteins by lectins and localization after expression in the cells suggest that only the N-terminal TG fragment (containing the SP) is directed to the endoplasmatic reticulum but is unable to reach the Golgi complex.

The described expression system in human thyrocytes will be a helpful tool in studying the structure–function relationship of human TG in thyroid hormonogenesis.

European Journal of Endocrinology 136 508–515

Introduction

Thyroglobulin (TG) is the major component of the thyroid gland and has the unique ability to synthesize and to store thyroid hormone (1, 2). Catalysed by thyroid peroxidase, tyrosine residues in the TG molecule are iodinated and subsequently coupled to form thyroxine (T_4) and triiodothyronine (T_3). Mature TG (i.e. iodinated, hormone-containing TG) is stored in the lumen of the thyroid follicles. Hormonal secretion to the venous flow requires endocytosis of mature TG within the thyrocyte and fusion with lysosomes for cleavage of TG, releasing T_4 from the protein backbone (3).

Thyroglobulin isolated from thyroid tissue is a homodimeric glycoprotein consisting of two identical 330 kDa subunits. Each monomer contains 2749 amino acids (4). An additional stretch of 19 amino acids at the N-terminus of the molecule encodes a signal peptide (SP). It has been shown that the signal peptide is removed after expression of a functional recombinant N-terminal TG fragment in a baculovirus system (5). Within the amino acid sequence of TG twenty potential N-glycosylation sites are available of which sixteen sites

are found to be glycosylated in TG isolated from thyroid tissue. Three types of carbohydrate chains have been identified in this case: high mannose, hybrid/complex and complex (6). The human TG homodimer molecule contains 134 tyrosine residues of which 6–34 are iodinated *in vivo* (7). Two diiodotyrosine residues can be coupled to form T_4 : the donor residue is donating its iodophenol group to the acceptor tyrosine residue (hormonogenic site). The preferential acceptor site is Tyr5, which is the first one to be iodinated *in vivo*, but several other tyrosine residues are also involved in thyroid hormonogenesis (8, 9). We have shown that in the N-terminal fragment of TG (198 amino acids), expressed in insect cells, Tyr5 and 130 are preferentially involved in T_4 formation, although in the mutated fragments, other tyrosine residues participate in T_4 synthesis (10).

The 8.7 kb TG mRNA contains a coding sequence of 8304 nucleotides (4). The human TG gene is located on chromosome locus 8q24.2–8q24.3 (11, 12), covers over 300 kb and contains at least 37 exons separated by introns as large as 64 kb (13). Mutations in the TG gene can result in aberrant TG protein expression giving rise to congenital hypothyroidism (14).

We have developed an *in vitro* eukaryotic expression system for human thyroglobulin cDNA. Four overlapping TG cDNA constructs were prepared using human thyroid RNA as a template and were transiently transfected in a human thyroid cell line (HTori 3) and in COS-1 cells. The expression of antigenic TG protein fragments on Western blot, type of carbohydrate chains on the expressed fragments and localization in the cell were investigated. This system provides the opportunity to elucidate further the function of naturally occurring mutations in the TG gene on thyroid hormone synthesis.

Materials and methods

RNA isolation and complementary DNA preparation

Total RNA was isolated from normal thyroid gland tissue, obtained with informed consent, using TRIzol Reagent (Life Technologies BV, Breda, The Netherlands). cDNA was synthesized using random hexamers and reverse transcriptase according to standard procedures.

DNA amplification

In order to amplify approximately 2.3 kb cDNA fragments and to minimize amplification errors, PCR amplification (15) was done with 100 ng cDNA using DNA polymerase from the Expand High Fidelity kit (Boehringer Mannheim, Mannheim, Germany). The relative positions of the oligonucleotides (synthesized on Expedite Nucleic Acid Synthesis System; Millipore, Bedford, MA, USA) used to amplify the four different TG fragments are specified in Table 1. Sense primers TGII, TGIII and TGIV were designed to create a Kozak's consensus sequence (16) surrounding an internal in-frame ATG. A schematic representation of the amplified human TG cDNA fragments is shown in Fig. 1. The

Table 1 Sequence of oligonucleotide primers. The cDNA positions of the most 5' nucleotide are in parentheses. Translation initiation codons for different TG fragments are underlined. Mismatches introduced to create Kozak's consensus sequence (16) are indicated by lower case letters.

Primers	Nucleotide sequence (5' → 3')
TGI _s	AGGAAAATGGCCCTGGTC (–6)
TGI _{as}	AGCACTCTGAGTTGAAGCACT (2110)
TGII _s	GGGCTacCATGCAAAGCCTCATG (1997)
TGII _{as}	CAGATGAAGCTGGAATGAGCC (4236)
TGIII _s	GGCAccATGTGTCAGTGCTGACTAC (3907)
TGIII _{as}	CATCGCACCGTCGTTTACA (6004)
TGIV _s	TGTGATGACAcCATGGAGTC (5782)
TGIV _{as}	GCTGGTCACTTGCTGTAGGT (+5)

s, sense; as, antisense.

overlapping regions between the fragments contain a unique restriction site (SpeI, HaeII and EcoRV respectively). This creates the possibility of reconstructing the full-length TG cDNA or of preparing other combinations of fragments.

Expression vectors

The EcoRV polylinker site of the pcDNA3 plasmid (Invitrogen, Leek, The Netherlands) was used to clone the amplified blunt TG fragments. The constructs are referred to as pTGI, pTGII, pTGIII and pTGIV. A termination signal located in the polylinker was used in constructs pTGI, pTGII and pTGIII leading to the addition of 16–24 non-TG amino acids at the C-terminal site.

DNA sequencing

Nucleotide sequence analysis was performed to verify the amplified TG cDNA fragments using T7, SP6 and

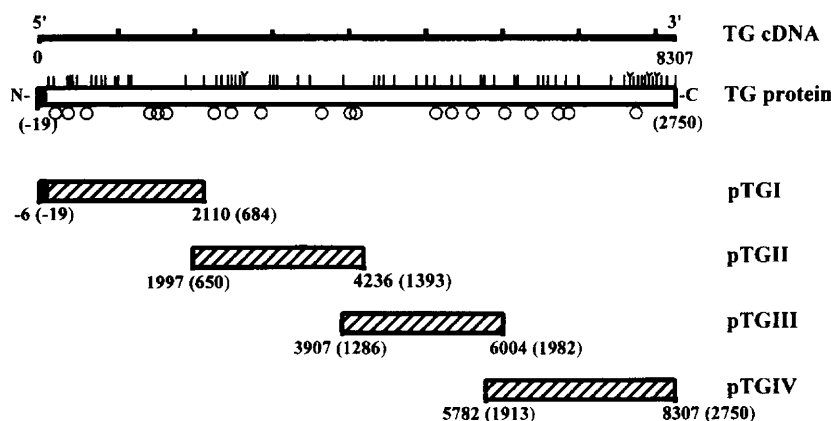


Figure 1 Schematic illustration of TG cDNA and protein. Top: relative positions of the nucleotide numbers of the TG open reading frame and positions of the first and last amino acid residue (between parentheses) are shown. Bottom: TG fragments (pTGx), amplified from cDNA with primers described in Table 1, with nucleotide and amino acid positions. The black box indicates the signal peptide. Relative (revised) positions according to (4) and Table 2. Tyrosine residues (Y = 1, Y = 2 residues) and putative N-glycosylation sites (O) are indicated.

internal TG specific oligonucleotides (synthesized on Expedite Nucleic Acid Synthesis System; Millipore). Reactions were performed using the cycle sequencing method procedure and the Dyedeoxyterminator Cycle Sequencing Kit (Applied Biosystems/Perkin Elmer, Foster City, CA, USA). After electrophoresis on a sequencing gel, the samples were analysed on the ABI Prism 377 DNA sequencer and aligned to the TG cDNA sequence reported by Malhiéry and Lissitzky (4) using AutoAssembler software (Applied Biosystems/Perkin Elmer).

Cell culture and transfection

COS-1 cells and HTori 3 cells (17) were cultured in Dulbecco's Modified Eagle's Medium (Life Technologies BV) containing 4.5 g glucose/l supplemented with NaHCO_3 , L-glutamine, antibiotics and 10% fetal calf serum (FCS, Life Technologies BV). Cells were grown to 60–70% confluence at 37°C, 5% CO_2 and 95% humidity, and transfected using the calcium phosphate precipitation method (18). For immunofluorescence microscopy cells were grown on coverslips in 3.5 cm dishes. One millilitre calcium phosphate precipitate containing 18 μg pTGx construct and 2 μg pcDNA3 was used per 5×10^6 cells. pcDNA3CAT (Invitrogen) was used for mock transfection. After 24-h incubation, cells were lysed with Reporter Lysis buffer (Promega, Madison, WI, USA). After centrifugation, the cell-lysate containing cytoplasmic proteins was stored at -70°C . The cells grown on coverslips were fixed with Kryofix (Merck, Darmstadt, Germany), 2.5 ml per 3.5 cm dish, and kept at room temperature until further analysis.

Thyroidal TG isolation

Normal thyroid gland tissue, stored at -70°C , was also used to purify TG. About 1 g tissue was pulverized under liquid N_2 and homogenized in phosphate-buffered saline (PBS: 9.2 mM Na_2HPO_4 , 1.3 mM NaH_2PO_4 , 140 mM NaCl, pH 7.4) containing protease inhibitors. After centrifugation at 100 000 g the supernatant was brought to 50% ammonium sulphate saturation, incubated at 4°C for 16 h and centrifuged at 100 000 g . The precipitated TG was solved in and dialysed against PBS followed by chromatography on Sepharose CL-6B (Pharmacia Biotech, Wisconsin, USA). Fractions containing TG were pooled and concentrated with an Amicon YM100 filter (Amicon, Danvers, MA, USA). For the glycosylation differentiation 1.5 μg was subjected to SDS-PAGE and electroblotted (procedures described below).

Protein determination

Protein concentration was determined in isolated total TG and in the cell-lysates of the transfected cells

using the BioRad Protein Assay (BioRad, München, Germany).

Immunoblot analysis

Cell-lysate (7 μg) and TG protein (1.5 μg) were subjected to SDS-PAGE according to Laemmli's method (7.5%) and electroblotted to nitrocellulose filters (0.1 μm ; Schleicher & Schuell, Dassel, Germany). Immunodetection of TG protein was performed at room temperature using PBS containing 0.1% Tween 20, completed with 5% low fat milk to prevent nonspecific binding, in steps containing primary antibody (RaHuTG; 19) or secondary antibody (GARPO; DAKO A/S, Glostrup, Denmark). Antigen-antibody complexes were visualized using the ECL kit (Amersham International plc, Amersham, Bucks, UK).

Immunoprecipitation

Expressed TG fragment proteins were immunoprecipitated from approximately 200 μg cell-lysate, using polyclonal rabbit anti-human TG (RaHuTG; 19) coupled to Protein A-Sepharose CL-4B (Pharmacia Biotech). The TG-Sepharose complexes were subjected to SDS-PAGE and electroblotted as described above. These blots were used for glycosylation differentiation.

Glycosylation differentiation

Nitrocellulose filters, containing immunoprecipitated TG protein fragments and isolated total TG protein, were incubated with different lectins labelled with digoxigenin. *Galanthus nivalis* agglutinin (GNA) recognizes terminal mannose linked to mannose. Both *Sambucus nigra* agglutinin (SNA) and *Maackia amurensis* agglutinin (MAA) recognize sialic acids linked to galactose, $\alpha(2-6)$ and $\alpha(2-3)$ respectively. *Datura stramonium* agglutinin (DSA) recognizes $\text{Gal}\beta(1-4)\text{GlcNAc}$. Lectins were detected with anti-digoxigenin conjugated to alkaline phosphatase and visualized with NBT/X-phosphate (DIG Glycan Differentiation kit; Boehringer Mannheim).

Immunofluorescence microscopy

Transfected cells fixed on coverslips were incubated for 30 min with 10% FCS in PBS to prevent nonspecific binding. The 16-h primary (RaHuTG; 19) incubation and 2-h secondary (FITC-conjugated goat anti-rabbit IgG; ICN Biomedicals, Costa Mesa, CA, USA) (in darkness) incubation were performed in PBS containing 2% bovine serum albumin at 4°C. Cells were embedded in Vectashield (Vector Laboratories, Burlingame, CA, USA) and then covered with a glass slide. Microscopy was performed with model AH 3-RFCA (Olympus, Tokyo, Japan) in combination with B excitation (BH2-DMB cube) for FITC visualization.

Table 2 Variations detected in TG constructs after aligning nucleotide sequence with published TG sequence (4).

Construct	nt position	nt mutation ref 4 → this paper	aa position	aa change	nt: allele-frequency in normal cDNA
TGI	78	GTT → GTG	7	—	G: 10/10
	405	CAT → CAG	116	His → Gln	G: 10/10
	1543 ^a	CAA → GAA	496	Gln → Glu	C: 10/10
	1810/11	GAC → AGC	585	Asp → Ser	AG: 10/10
TGII	1958	GAT → GGT	634	Asp → Gly	G: 10/10
	2185 ^a	CCC → TCC	710	Pro → Ser	C: 10/10
	2200	TCT → GCT	715	Ser → Ala	T: 4/10; G: 6/10
	2334	CCT → CCC	758	—	T: 4/10; C: 6/10
	2760	ACG → ACA	901	—	A: 10/10
	2952 → ^b	→ CAG	966 ^a	Gln	CAG: 10/10
	3082 ^b	ATG → GTG	1009 ^a	Met → Val	A: 6/8; G: 2/8
	3127 ^b	TAC → CAC	1024 ^a	Tyr → His	C: 8/8
	3141 ^b	GGT → GGG	1028 ^a	—	G: 8/8
	3176/77 ^b	ACT → ATC	1040 ^a	Thr → Ile	TC: 8/8
	3339 ^b	GAA → GAG	1094 ^a	—	G: 6/6
	4302 ^b	CCT → CCC	1415 ^a	—	C: 12/12
TGIII	4309 ^b	CGG → TGG	1418 ^a	Arg → Trp	T: 12/12
	4386 ^b	TGC → TGT	1443 ^a	—	T: 10/10
	4388 ^b	CAT → CCT	1444 ^a	His → Pro	C: 10/10
	4506 ^b	GCC → GCT	1483 ^a	—	C: 5/10; T: 5/10
	5512 ^b	AAC → GAC	1819 ^a	Asn → Asp	A: 4/10; G: 6/10
	5806 ^{b,c}	ACC → GCC	1917 ^a	Thr → Ala	G: 8/8
	5806 ^{b,c}	ACC → GCC	1917 ^a	Thr → Ala	G: 8/8
	6273 ^b	GAA → GAC	2072 ^a	Glu → Asp	C: 10/10
TGIV	6446/47 ^b	CTC → CCT	2130 ^a	Leu → Pro	CT: 10/10
	6509 ^b	CGC → CAG	2151 ^a	Arg → Gln	A: 10/10
	6725 ^b	CAC → CGC	2223 ^a	His → Arg	G: 6/6
	7038 ^b	GGA → GGG	2327 ^a	—	G: 6/6
	7408 ^b	CTG → TTG	2451 ^a	—	C: 3/10; T: 7/10
	7501 ^b	TGG → CGG	2482 ^a	Trp → Arg	T: 3/6; C: 3/6
	7589 ^b	CGA → CAA	2511 ^a	Arg → Gln	A: 6/6

^aVariation at this position due to PCR amplification artefact. ^bCAG insertion causes a shift in published positions (4); all positions downstream of this insertion are adjusted (nt + 3, aa + 1). ^cVariation at this position was detected in both pTGIII and pTGIV constructs. nt, nucleotide; aa, amino acid.

Results

Four overlapping cDNA fragments encoding for the human thyroglobulin were amplified by reverse transcription (RT)-PCR and cloned into an expression vector (pcDNA3) designed for high level expression in eukaryotic cells. Relative positions of the constructs are shown in Fig. 1, base pair and amino acid positions are according to (4) and Table 2. The sizes of the TG constructs range from 2097 to 2525 nucleotides encoding for 696–837 amino acids. In the N-terminal pTGI construct translation was initiated from the natural TG translation codon thus encoding the signal peptide; pTGII, pTGIII and pTGIV initiate from an internal methionine.

For validation purposes all PCR-generated constructs were sequenced. Thirty variations detected in the TG constructs compared with the published sequence (4) are shown in Table 2. To determine whether these variations are based on proofreading errors during amplification or on polymorphisms in the TG sequence,

RT-PCR followed by direct sequence analysis was performed on RNAs isolated from 3 to 6 different normal thyroid tissues. The allele frequency of the screened nucleotide variations is shown in the last column (Table 2). Nucleotide variations at positions 78, 405, 1810/11 1958, 2760, 2952, 3127, 3141, 3176/77, 3339, 4302, 4309, 4386, 4388, 5806, 6273, 6446/47, 6509, 6725, 7038 and 7589 were present in all screened alleles. The CAG insertion at position 2952 results in the insertion of a glutamine residue. The nucleotide and amino acid positions are adjusted after this insertion. Fourteen revisions result in an amino acid change. One tyrosine residue substitution was found (amino acid position 3127). No variations affecting the putative N-glycosylation sites or nucleotide deletions were detected. Nucleotide variations at positions 2200, 2334, 3082, 4506, 5512, 7408 and 7501 were only present in some alleles and are considered to be polymorphic. Four of these polymorphisms result in an amino acid change. Only nucleotide variations 1543 and 2185, resulting in an amino acid substitution,

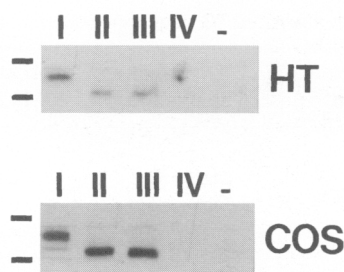


Figure 2 Immunoblot of TG fragments (I, II, III and IV) expressed in HTori 3 (HT) and COS-1 (COS) cells. Twenty-four hours after transfection, cells were lysed and cytoplasmic proteins were separated by SDS-PAGE (7 µg protein/lane), blotted and screened with the polyclonal antibody RaHuTG. Lane (-) contains cell-lysate of mock transfected cells. Positions of 117 000 (upper mark) and 89 000 (lower mark) molecular weight markers are shown.

could not be verified in other alleles and are considered PCR artefacts.

HTori 3 and COS-1 cells were transfected with plasmid DNA, and expressed TG protein was analysed by immunoblotting using polyclonal antiserum (RaHuTG) raised against total human TG. As shown in Fig. 2, TG protein fragments I, II, III migrated under denaturing conditions with apparent molecular weights of 109 000, 98 000 and 95 000 respectively in both HTori 3 and COS-1 transfected cells. Fragment TGIV was poorly recognized and could only be detected after immunoprecipitation with the polyclonal TG antibodies giving a signal at a molecular weight of 105 000 (data not shown). The molecular weights after SDS-PAGE differ from the pTGx calculated ones: 80 700, 82 800, 78 800 and 91 900 for TGI, TGII, TGIII and TGIV respectively.

The specific binding of lectins to different carbohydrate moieties was used to identify these structures in the TG protein fragments expressed in HTori 3 and COS-1 cells. Before SDS-PAGE, TG proteins were immunoprecipitated with RaHuTG. Total TG isolated from normal thyroid tissue was also analysed by recognition of four different lectins. Total TG reacted positive with all four lectins. Of the expressed TG fragments, only TGI was recognized by lectin GNA, reacting with terminal mannose linked to mannose, when expressed in either HTori 3 or COS-1 cells.

The native TG protein fragments were detected in transfected and fixed HTori 3 and COS-1 cells with RaHuTG and visualized with a fluorescent label. In Fig. 3 images are shown of the localization of the TG fragments in HTori 3 and COS-1 cells. Also, under these non-denaturing conditions the TGIV fragment showed weak staining which was, however, clearly above background, with the polyclonal antiserum in both HTori 3 and COS-1 cells. The signal did not increase when mixtures of monoclonals described by Den Hartog *et al.* (19) were used. The TG fragments showed a diffuse pattern of fluorescence in both cell-lines. However, the TGI pattern was different from

that of TGII, TGIII and TGIV, showing a more granulous pattern. The patterns did not change on longer incubation of the cells after transfection (up to 96 h; data not shown).

The revised coding sequence for human thyroglobulin has been submitted to NCBI/GenBank.

Discussion and conclusions

Although the structure and function of TG have been the subject of investigations for many years, several questions remain unanswered. In order to investigate the structure–function relationship of wild type and mutated TG in relation to thyroid (dys)hormonogenesis we have developed a eukaryotic expression system. We have used an immortalized human thyroid cell line, HTori 3, that does not express endogenous TG (data not shown), and COS-1 cells to transiently express transfected TG cDNA constructs.

TG cDNA was amplified from RNA isolated from thyroid tissue in four overlapping fragments of approximately 2.3 kb (Fig. 1). These fragments were selected based on the presence of a unique restriction site in the overlapping regions making it possible to reconstruct either the full-length TG cDNA or combinations of fragments in the future. After subcloning, the nucleotide sequences were determined and aligned with the already published human cDNA sequence (4). Thirty variations were observed (Table 2), 10 of them would not alter the wild type amino acid sequence and 19 could result in an amino acid substitution. Surprisingly, one variation comprised an in-frame CAG insertion and subsequently an amino acid insertion. Variations at nucleotide positions 78, 405 and 2760 were already mentioned in the literature (20, 5, 21 respectively). In order to investigate whether the variations were due to polymorphisms, errors in the published sequence or simply to PCR amplification artefacts, we checked 6–12 normal alleles for the presence or absence of these variations. Two variations appeared to be due to a PCR artefact (nt 1543 and 2185) both resulting in an amino acid substitution, 7 polymorphisms were scored (nt 2200, 2334, 3082, 4506, 5512, 7408 and 7501) of which 4 result in variation of amino acid residues. Twenty-one revisions have to be made to the published sequence (Table 2). The most striking revisions are the triplet insertion after nucleotide position 2952 since this leads to the insertion of a glutamine residue and subsequently to repositioning of nucleotides (+3) and amino acids (+1) downstream of this insertion. Furthermore, one tyrosine residue (1024) was substituted by a histidine residue, thus mature TG protein contains 66 instead of 67 tyrosine residues as reported (4). None of the revisions involved putative N-glycosylation sites. Moreover, the inclusion of the extra glutamine residue removes a gap in the comparison of the amino acid sequence between human (4) and bovine (22) TG and due to the amino acid revisions the homology

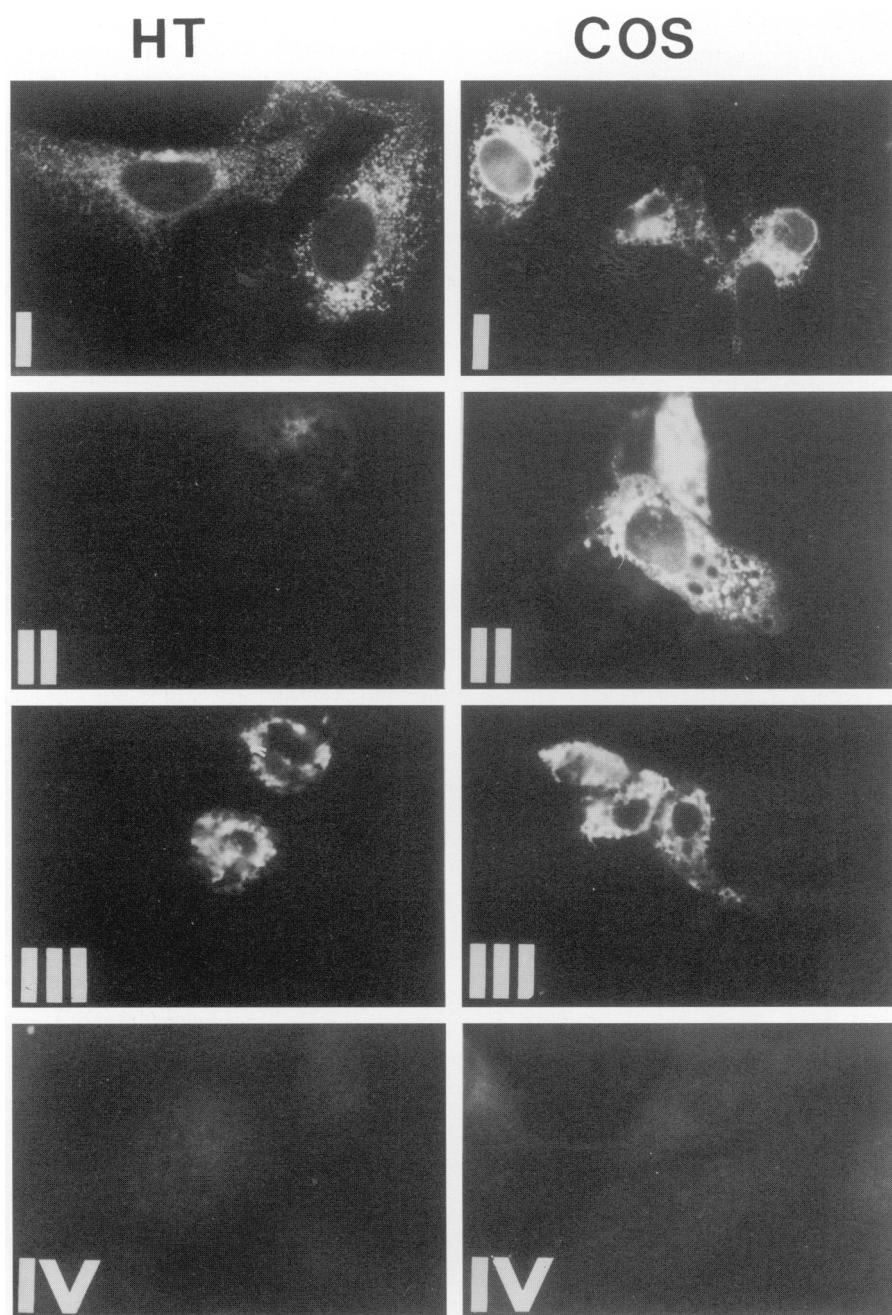


Figure 3 Immunocytochemical localization of the TG fragments expressed in HTori 3 (HT) and COS-1 (COS) cells. Cells were plated on coverslips before transfection and fixed after 24-h incubation. Screening was performed with the polyclonal antibody RaHuTG and FITC-conjugated secondary antibodies. Magnification is $\times 400$. Transfected constructs are indicated in the left corner of each image.

between human and bovine increases (from 77.22% to 77.52%).

The pTGx constructs were transiently transfected into HTori 3 and COS-1 cells. The expressed TGI, TGII and TGIII proteins were immunodetected on Western blot by a polyclonal rabbit anti-human TG, whereas TGIIV protein was below the detection limit (Fig. 2). Only after

immunoprecipitation was TGIIV protein detectable on SDS-PAGE (data not shown).

The molecular weights of the expressed protein fragments as determined under denaturing electrophoresis conditions (Fig. 2) are identical in HTori 3 and COS-1 cells. They differ from the calculated molecular weights, probably due to post-translational processes.

The signal peptide on the N-terminus of the TGI peptide is used in general for directing pre-proteins from the cytosol to the endoplasmic reticulum (ER) where the SP is often removed and polypeptide synthesis is completed, thereby raising the opportunity for post-translational processes like glycosylation. The N-terminal part of TG (198 amino acids) was found to be glycosylated after expression in a baculovirus expression system (5) emphasizing the proposed function of the signal peptide. The glycosylation status of isolated total TG and of the four TG protein fragments expressed in COS-1 and HTori 3 cells was evaluated using various lectins recognizing either terminal mannoses or sialic acids on galactoses or non-sialysed N-acetylglucosamines. Thyroidal total TG was recognized by all lectins indicating that all different glycosylation types (i.e. high mannose, complex and hybrid) were present, as already reported (6). As might be expected, only TGI reacted with the lectin recognizing terminal mannoses, while TGII, TGIII and TGIV, which contain no SP sequence, reacted with none of the lectins. No differences with respect to type of glycosylation detected were observed in HTori 3 and COS-1 cells. The staining pattern (Fig. 3) of TGI resembles staining of ER specific proteins and differs from the TGII, TGIII and TGIV staining pattern, suggesting that TGI is in the ER while the other TG protein fragments are located in the cytoplasm. The mannose complexes and ER localization suggest that in our expression system the TGI protein was not able to reach the Golgi complex, due to either aberrant folding of the synthesized protein fragment or to defects in post-translational processes as a result of the high level of expression.

In conclusion, the work presented here leads to several adjustments in the human TG mRNA sequence published by Malthiery and Lissitzky (4) (Table 2) and the mature TG protein has now extended to 2750 amino acids encoded by 8307 nucleotides. We have developed a eukaryotic expression system for human TG in a cell-line derived from human thyroid tissue. This will be a valuable tool with which to study the structure-function relationship of mutant thyroglobulin involved in dyshormonogenesis in, for example, patients with a TG synthesis defect.

Acknowledgements

We thank Sir AB Parkes, University of Cardiff, UK for providing us with HTori 3 cells and the dr Ludgardine Bouwman Foundation for financial support.

References

- 1 Van Herle AJ, Vassart G & Dumont JE. Control of thyroglobulin synthesis and secretion. *New England Journal of Medicine* 1979 301 239–249 & 307–314.
- 2 Dunn JT. Thyroglobulin: chemistry and biosynthesis. In *Werner's & Ingbar's The Thyroid*, pp 85–95. Eds LE Braverman & RD Utiger. Philadelphia: J.B. Lippincott Company, 1996.
- 3 Dunn AD, Crutchfield HE & Dunn JT. Proteolytic processing of thyroglobulin by extracts of thyroid lysosomes. *Endocrinology* 1991 128 3073–3080.
- 4 Malthiery Y & Lissitzky S. Primary structure of human thyroglobulin deduced from the sequence of its 8448 base complementary DNA. *European Journal of Biochemistry* 1987 165 491–498.
- 5 Den Hartog MT, Sijmons CC, Kristel PMP, Bakker O & de Vijlder JJM. Expression of a functional thyroglobulin fragment in a baculovirus system. *Journal of Molecular Endocrinology* 1993 11 161–166.
- 6 Yang SX, Pollock HG & Rawitch AB. Glycosylation in human thyroglobulin: location of the N-linked oligosaccharide units and comparison with bovine thyroglobulin. *Archives of Biochemistry and Biophysics* 1996 327 61–70.
- 7 Izumi M & Larsen PR. Triiodothyronine, thyroxine, and iodine in purified thyroglobulin from patients with Graves' disease. *Journal of Clinical Investigation* 1977 59 1105–1112.
- 8 Palumbo G, Gentile F, Condorelli GL & Salvatore G. The earliest site of iodination in thyroglobulin is residue number 5. *Journal of Biological Chemistry* 1990 265 8887–8892.
- 9 Rawitch AB, Gregg J & Turner CD. Nature and location of hormonogenic sites within the structure of thyroglobulin. In *Thyroglobulin: The Prohormone. Vol 2. Progress in Endocrine Research and Therapy*, pp 43–54. Eds MC Eggo & GN Burrow. New York: Raven Press, 1985.
- 10 Den Hartog MT, Sijmons CC, Bakker O, Ris-Stalpers C & de Vijlder JJM. The importance of the content and localization of tyrosine residues for thyroxine formation within the N-terminal part of human thyroglobulin. *European Journal of Endocrinology* 1995 132 611–617.
- 11 Baas F, Bikker H, Geurts van Kessel A, Melsert R, Pearson PL, de Vijlder JJM *et al.* The human thyroglobulin gene: a polymorphic marker localized distal to c-myc on chromosome 8 band q24. *Human Genetics* 1985 69 138–145.
- 12 Bergé-Lefranc JL, Cartouzou G, Mattei MG, Passage E, Malezet-Desmoulin C & Lissitzky S. Localization of the thyroglobulin gene by *in situ* hybridization to human chromosomes. *Human Genetics* 1985 69 28–31.
- 13 Baas F, van Ommen GJ, Bikker H, Arnberg AC & de Vijlder JJM. The human thyroglobulin gene is over 300 kb long and contains introns of up to 64 kb. *Nucleic Acids Research* 1986 14 5171–5186.
- 14 Ledent C, Parma J, Dumont J, Vassart & Targovnik H. Molecular genetics of thyroid diseases. *European Journal of Endocrinology* 1994 130 8–14.
- 15 Saiki RK, Gelfand DH, Stoffel S, Scharf SJ, Higuchi R, Horn GT *et al.* Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. *Science* 1988 239 487–491.
- 16 Kozak M. Comparison of initiation of protein synthesis in procaryotes, eucaryotes, and organelles. *Microbiology Review* 1983 47 1–45.
- 17 Lemoine NR, Mayall ES, Jones T, Sheer D, McDermid S, Kendall-Taylor P *et al.* Characterisation of human thyroid epithelial cells immortalised *in vitro* by simian virus 40 DNA transfection. *British Journal of Cancer* 1989 60 897–903.
- 18 Chen C & Okayama H. High efficiency transformation of mammalian cells by plasmid DNA. *Molecular Cell Biology* 1987 7 2745–2752.
- 19 Den Hartog MT, de Boer M, Veenboer GJM & de Vijlder JJM. Generation and characterization of monoclonal antibodies directed against noniodinated and iodinated thyroglobulin, among which are antibodies against hormonogenic sites. *Endocrinology* 1990 127 3160–3165.
- 20 Mason ME, Dunn AD, Wortsman J, Day RN, Day KH, Hoback SJ *et al.* Thyroids from siblings with Pendred's syndrome

- contain thyroglobulin messenger ribonucleic acid variants. *Journal of Clinical Endocrinology and Metabolism* 1995 **80** 497–503.
- 21 Parma J, Christophe D, Pohl V & Vassart G. Structural organization of the 5' region of the thyroglobulin gene. *Journal of Molecular Biology* 1987 **196** 769–779.
- 22 Mercken L, Simons MJ, Swillens S, Massaer M & Vassart G.

Primary structure of bovine thyroglobulin deduced from the sequence of its 8431-base complementary DNA. *Nature* 1985 **316** 647–651.

Received 9 January 1997

Accepted 24 February 1997