

CLINICAL STUDY

Genomic organization of the 5' region of the human thyroglobulin gene

Christian M Moya¹, Fernando M Mendive¹, Carina M Rivolta¹, Gilbert Vassart² and Héctor M Targovnik^{1,2}

¹División Genética, Hospital de Clínicas 'José de San Martín' and Cátedra de Genética y Biología Molecular, Facultad de Farmacia y Bioquímica, Universidad de Buenos Aires, 1120-Buenos Aires, Argentina, and ²Service de Génétique Médicale, Hôpital Erasme and IRIBHN, Université Libre de Bruxelles, 1070 Bruxelles, Belgium

(Correspondence should be addressed to H M Targovnik, División Genética, Hospital de Clínicas 'José de San Martín' and Cátedra de Genética y Biología Molecular, Facultad de Farmacia y Bioquímica, Universidad de Buenos Aires, Av. Córdoba 2351 4^{to} piso – sala 5, 1120-Buenos Aires, Argentina; Email: htargovn@huemul.ffyb.uba.ar)

(C M Moya and F M Mendive contributed equally to the study)

Abstract

Objective: The purpose of the present work is to establish the intron–exon organization from exon 12 to exon 23 of the human thyroglobulin gene and to construct a physical map of the 5' terminal half of the gene.

Design: Screening of a genomic library and subsequent restriction map, hybridization and sequencing methods have been employed to characterize the recombinant positive phages.

Methods: A human genomic DNA library was screened by *in situ* hybridization. Southern blotting experiments were performed to characterize the phage inserts. Intron/exon junction sequences were determined by the Taq polymerase-based chain terminator method. Finally, the thyroglobulin gene was mapped using the Gene Bridge 4 radiation hybrid clone panel.

Results: We isolated and characterized four λ phage clones that include nucleotides 3002 to 4816 of the thyroglobulin mRNA, encompassing exons 12 to 23 of the gene. The exon sizes range between 78 and 219 nucleotides. We found that the GT-AG splicing sequences rule was perfectly respected in all the introns. A total of 7302 intronic bases was analyzed. Hormogenic tyrosine 5 and 1291 are encoded by exons 2 and 18. Also, seven alternative spliced variants are associated with the 5' region. Thyroglobulin gene maps to 5.5 centiRays from the AFMA053XF1 marker, in chromosome 8.

Conclusions: The present study shows that the first 4857 bases of thyroglobulin mRNA are divided into 23 exons and the four phages isolated include 32.6 kb genomic DNA, covering 1815 nucleotides of exonic sequence distributed in 12 exons, from exon 12 to 23.

European Journal of Endocrinology 143 789–798

Introduction

Synthesis of tri-iodothyronine (T₃) and thyroxine (T₄) follows a metabolic pathway that depends on the integrity of the thyroglobulin (TG) structure (1, 2). This large glycoprotein is a homodimer of 660 kDa synthesized and secreted by the thyroid cells into the lumen of the thyroid follicle. The human TG is coded by a large gene about 300 kb long (3), located on chromosome 8q24.2–8q24.3 (4–6). The number of exons has been estimated to be above 48 (7), but intron/exon junctions in the human gene are only known for the first 11 exons (8–10) and the last 13 exons (7, 11).

The mRNA encoding human TG is 8.5 kb long (12–14). The general organization of the mRNA sequence showed a 41-nucleotide (nt) 5'-untranslated segment preceding an open reading frame of 8307 bases. The preprotein monomer is composed of a 19-amino acid

signal peptide followed by a 2749-amino acid polypeptide, including 66 tyrosine residues of which 3–17 are iodinated *in vivo*. Analysis of the primary structure of the TG protein for internal homology led to its division into three regions (12): type-1 repetitive region which contains eleven type-1 repeat elements located between positions 12 and 1191 and between 1492 and 1546 (12, 15, 16); type-2 repetitive region composed of three type-2 elements located between amino acids 1437 and 1484 (12); and type-3 repetitive region with five elements between residues 1584 and 2168 (12). This internal repetitive protein organization makes TG an example of gene evolution by intragenic duplication events and gene fusions (9).

In contrast, the last 581 C-terminal residues show no internal homology and contain remarkable sequence similarity with *Torpedo californica* acetylcholinesterase (17).

The TG maturation process occurs within thyrocytes and includes intramonomer disulfide bonding formation, glycosylation, phosphorylation and sulfatation (18). The oxidation, iodination and coupling reactions are mediated by thyroperoxidase at the thyrocyte apical cell surface (19). The native three-dimensional TG structure is necessary for the coupling of iodotyrosyl residues to form T₄ and T₃. Five hormonogenic acceptor tyrosines have been identified and localized at positions 5, 1291, 2554, 2568 and 2747 in human TG (20) and several tyrosines have been proposed as donor sites by transferring an iodophenoxy group to an acceptor iodotyrosine. After coupling, the acceptor residues are converted to hormones and the donor residues are converted to dehydroalanine. The last stages of the biosynthesis of the thyroid hormones follow TG internalization from the lumen and its proteolytic cleavage in the lysosomal system.

Recently, two probable molecular control mechanisms of thyroid hormonogenesis were identified in the N-terminal half of the sequence (21, 22). The first suggests that an apical endogenous membrane receptor interacts with the secreted immature TGs and prevents their premature lysosomal transfer and degradation (21). Molina *et al.* (22) propose that the 11 type-1 repeats regulate the mature TG degradation by a selective and reversible inhibition of the lysosomal proteases.

Reverse transcriptase-polymerase chain reaction (RT-PCR) demonstrates the existence of 11 alternative splice transcripts for the TG gene (14, 23–27). Whether these must be considered as ‘noise’ originating from inaccuracy of the splicing machinery or are endowed with some function in thyroid hormonogenesis remains to be determined.

Knowledge of the structural organization of the human TG gene will help to elucidate the function of the different domains of the protein. On the other hand, the efficient detection of mutations in TG-linked families requires a more thorough description of the genomic structure of TG gene, including precise intron/exon boundaries.

In the present work, we report the establishment of the intron–exon borders of exons 12 to 23 of the human TG gene, including sequence data from splicing signals and the flanking intronic regions. Additionally, a physical map covering the majority of the 5′ region of the gene was also constructed.

Materials and methods

Probe amplification by RT-PCR

Four RT-PCR fragments: PCR 2.2, PCR 3.1, PCR 3.2 and PCR 3.3, previously described as 1.2F-2R, 3F-1.3R, 1.3F-2.3R and 2.3F-3R respectively (28), were used in the screening of a human genomic library and in Southern blot analysis. These four RT-PCR probes

map in the central region of the TG mRNA and together encompass nt 3012 to 6010 (according to new cDNA numbering (13)).

Total RNA was prepared from human thyroid tissue by the method of Chomczynski and Sacchi (29). Two micrograms total RNA were first reverse transcribed with 200 U Moloney murine leukemia virus reverse transcriptase (Gibco BRL, Life Technology, Gaithersburg, MD, USA) and 20 U Rnase inhibitor (Rnasin, Promega, Madison, WI, USA) in 20 µl solution containing a standard reverse transcription buffer (Gibco BRL), 1 mmol/l of each dNTP (dATP, dCTP, dTTP and dGTP) and 50 pmol of the reverse primer during 1 h at 42 °C. The reverse transcriptase was inactivated at 95 °C.

The PCR reactions were performed in 100 µl, using a standard PCR buffer (Gibco BRL) containing the 20 µl RT reaction, 2.5 mmol/l MgCl₂, 4% dimethylsulfoxide, 2 U Taq polymerase (Gibco BRL) and 50 pmol of each reverse and forward primers. No dNTP were added in the PCR reaction, so, as the RT reaction is diluted 1/5, the final concentration of nucleotides was 200 µmol/l. The samples were subjected to 40 cycles of amplification. Each cycle consisted of denaturation at 95 °C for 30 s, primer annealing at 55 °C for 30 s, and primer extension at 72 °C for 1 min. After the last cycle, the samples were incubated for an additional 5 min at 72 °C to ensure that the final extension step was complete.

The DNA sequences of each of the oligonucleotides used for RT-PCR and the positions of the first and last nucleotide of each probe are shown in Table 1 and Fig. 1 respectively. The amplified products (PCR 2.2: 1076 bp, PCR 3.1: 681 bp, PCR 3.2: 683 bp and PCR 3.3: 884 bp) were analyzed in a 1.5% agarose gel.

Screening of a human genomic library

An aliquot of human genomic library constructed with λ Dash II phage vector (Stratagene, La Jolla, CA, USA) was used to infect *Escherichia coli* XL1-Blue MRA. Infected bacteria were plated on petri dishes and screened by the filter replica method using PCR 2.2, PCR 3.1, PCR 3.2 and PCR 3.3 as hybridization probes.

Prehybridization and hybridization were carried out at 42 °C in 50% formamide, 5 × Denhardt’s solution, 5 × SSPE (1 × SSPE is 0.15 mol/l NaCl, 0.01 mol/l NaH₂PO₄ (pH 7.7), and 0.001 mol/l EDTA), 0.5% SDS and 100 µg/ml sonicated denatured salmon sperm DNA.

The probes were labeled with α[³²P]dATP by random priming. Filters were washed with 2 × SSC (1 × SSC is 0.15 mol/l NaCl and 0.015 mol/l sodium citrate, pH 7.0), 0.1% SDS, followed by 1 × SSC, 0.1% SDS, and finally 0.1 × SSC, 0.1% SDS, twice each at 65 °C. The filters were exposed to X-ray film at –70 °C with an intensifying screen.

Table 1 Oligonucleotides used as primers in amplification of hybridization probes.

RT-PCR fragments	Forward primer		Reverse primer	
	Nucleotide sequence (5' → 3')		Nucleotide sequence (5' → 3')	
PCR 2.2	CTATCAGAGACGCCGCTTTTCCCC		CTAAACGCTCCCTGGTCAGACAACC	
PCR 3.1	GCGGACGCTGGGAGTCACAGCTGC		CAGGACATGGGACACAGGCC	
PCR 3.2	GGCCTGTGTCCCATGTCCTG		TGGCTCCTGAGGCTGAGAAC	
PCR 3.3	GTTCTCAGCCTCAGGAGCCA		GGTCCGCATCGCACCCGTCGTTAC	

Screening of recombinant positive phages by PCR

The use of PCR provides a rapid and precise screening system to determine the relative localization of recombinant positive phages. This approach was used to identify the phage containing exon 23.

The DNA from each positive phage stock was purified by proteinase K treatment and phenol extraction. PCR reaction and cycles were performed as described for the probe amplification by RT-PCR. The forward (position of 5' end TG sequence: 4700) and reverse (position of 5' end TG sequence: 4762) primers are shown in Table 2. The amplified products, 63 bp, were analyzed in a 15% acrylamide gel.

Preparation of λ phage DNA

Bacteriophage DNA was prepared with the Wizard lambda preps DNA purification system (Promega). After elution from the mini column, DNA was extracted twice with phenol-chloroform, salt concentration was adjusted to 2 mol/l with ammonium acetate and the DNA was precipitated with ethanol.

Southern blot analysis of recombinant positive phages

Restriction and blotting experiments were carried out using standard procedures (30–32). One microgram λ

phage DNA was digested with 10 U EcoRI endonucleases (Gibco-BRL). Prehybridization, hybridization and washing of the filters were performed as described for the screening of the genomic library, except that washing was performed using 0.5% SDS.

In order to map the recombinant clones, the membranes were hybridized with PCR 2.2, PCR 3.1, PCR 3.2 and PCR 3.3 probes in separate experiments. Recombinant phages DNA were also used as probes to check the possibility of overlapping between contiguous phages.

DNA sequencing

The exon and intron/exon junction sequences were determined by Taq polymerase-based chain terminator method (fmol, Promega) from λ phage clones DNA. Primers were especially designed for each intron–exon junction. Oligonucleotide sequences and the position of their 5' ends are shown in Table 2. Intron–exon borders were characterized by alignment between the cDNA (13) and genomic sequences, using the PC Gene computer program (Intelligenetics, Inc., Geneva, Switzerland).

Chromosomal mapping

A panel of 93 radiation hybrid clones (Genebridge panel, Research Genetics, Huntsville, AL, USA) was used to map the TG gene by PCR screening. PCR were

Table 2 Oligonucleotides used as primers in sequence reactions.

Exon number	Forward primer		Reverse primer	
	Position of 5' end	Nucleotide sequence (5' → 3')	Position of 5' end	Nucleotide sequence (5' → 3')
12	3012	CTATCAGAGACGCCGCTTTTCCCC ¹	3114	CTCCCAACTTCCAAACGC
13	3149	GGTGTGTAGATGAGAAAGGAGGG	3211	GAATCTGCAGAGAGCG
14	3228	CTGCGAGAAATCTCGAACC	3308	AGGTCTTTTGGAGATGGG
15	3335	GAGAATATGCCAGGCTGC	3433	ACTGGGCACTGCTGCTCG
16	3446	GCAATGTGCTCAAGAGTGGAG	3592	CCTCTTCTCCGCTGTCCATGACACACC
17	3686	CAATCCTGTGTGAGACAATCTCGGG	3803	CTCTCCAGGCTACATATCAATGGC
18	3854	AGCTGTGGCAGACCATCCAGACCC	3980	GCTGTCAGCTCATCCAATATGAAAACC
19	4003	GTGAAGACTTTTGGCACCTGG	4150	GTAAGTCAGGAAGAGAGCCACTGGG
20	4184	TCCTTGGGCGCTTACAGATCTGATCC	4348	TTGTCAAGACTTGGTAGAATCCTTCCG
21	4393	GGAAGCTATTCCTCAAGATGAG	4483	CAGGACATGGGACACAGGCC ²
22	4529	GTGTCAGTACTGTCAGAGG	4699	TCAAACACTGTGAGTCCTCAA
23	4700	TGATGCAGAAGTTTGAGAAGGTTCC	4762	CCACAGGAGCATTGGC

¹Forward primer used in the PCR 2.2 probe amplification; ²reverse primer used in the PCR 3.1 probe amplification.

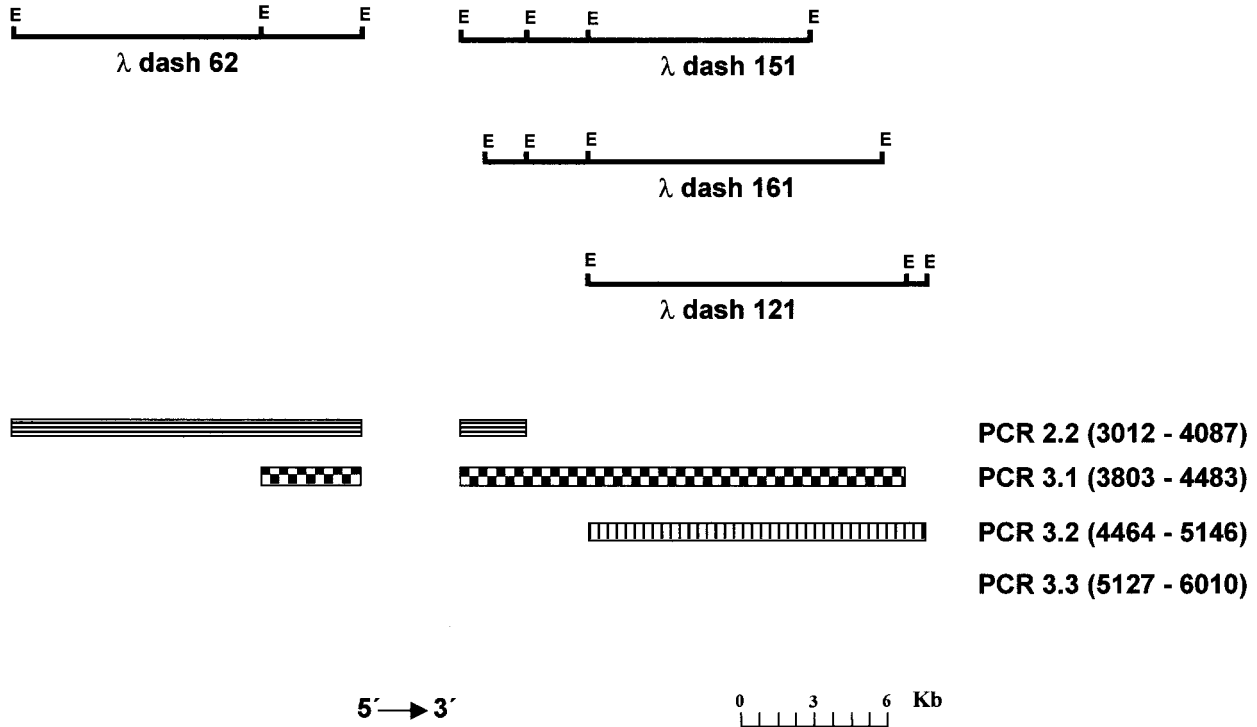


Figure 1 Relative positions and hybridization patterns of the λ phages used in this study. The EcoRI (E) fragments scoring positive in hybridization are shown, drawn to scale in the 5' to 3' direction. The lower bars represent the hybridization patterns with cDNA specific probes obtained by RT-PCR. None of the phages hybridize with PCR 3.3. The positions of the first and last nucleotide of each probe are shown in parentheses.

performed in 50 μ l solution, using Qiagen core kit (Qiagen, GmbH, Hilden, Germany), containing 2.5 mmol/l MgCl₂, 0.2 mmol/l dNTP, 50 ng target DNA, 50 pmol of each of the primers, and 1 U Taq DNA polymerase. The primers used were 5' cgttctgtccccacagt 3' as forward primer (5' end is located 208 nt upstream from exon 1, in the TG promoter region) and 5' gctccatggcctcagaact 3' as reverse primer (5' end is located 22 nt downstream from exon 1, in intron 1) (33). In order to allow easy sequencing, M13 sequences have been incorporated into PCR fragments by adding specific sequences at the 5' end of the primers. Cycles were performed as described for probe amplification by RT-PCR. The amplified products (374 bp, 338 of them TG sequences) were analyzed in a 1.5% agarose gel. No amplification product was obtained from hamster DNA used as negative control. The TG specificity of the PCR fragment was established by sequencing, using M13 primers sequences and the Thermo Sequenase Fluorescent Labeled Primer Cycle Sequencing Kit (Amersham Life Science, Amersham, Bucks, UK) followed by analysis on an ABI 370 A automated sequencer. Both strands were sequenced. The PCR results were analyzed using the Rhmapper computer program through the publicly available World Wide Web server of the

Whitehead Institute/MIT center (<http://carbon.wi.mit.edu:8000/cgi-bin/contig/rhmapper.pl>).

Results

About 1.9×10^6 phages of a human genomic library were screened by filter hybridization with four human TG cDNA probes (PCR 2.2, PCR 3.1, PCR 3.2 and PCR 3.3) corresponding to 3 kb TG mRNA. Fifty plaques scored positive and two of these were randomly selected and purified to homogeneity. The DNA was prepared from the corresponding phages and digested with EcoRI. Two different restriction patterns were obtained leading to the identification of two recombinant phages termed λ Dash 62 and λ Dash 151 respectively. Southern blot experiments with the same screening probes were used to localize the phage inserts with respect to the mRNA sequence and determine their relative orientation. Figure 1 shows the localization of λ Dash 62 and λ Dash 151 clones according to their positive hybridization fragments.

λ Dash 62 is located 5' to λ Dash 151 since the exonic fragments of λ Dash 62 hybridize with cDNA probes PCR 2.2 and PCR 3.1 and do not hybridize with PCR 3.2, whereas the λ Dash 151 exonic fragments are positive with all three probes and do not overlap with λ

Dash 62. These observations were verified by hybridizing the EcoRI blot with λ Dash 62 used as probe. The hybridization of λ Dash 62 and λ Dash 151 with PCR 3.3 was negative, since the region is outside the sequence covered by this probe.

In order to identify the intron/exon boundaries of the TG gene and to analyze the regions responsible for pre-mRNA processing, we performed cycle sequencing reactions from primers designed according to a sequencing strategy (i.e. the establishment of one intron-exon junction allowed us to design the following primer to sequence the neighbouring intron/exon border and so on). All investigated intron-exon junctions and splicing sites were sequenced from exonic primers (Table 2) heading into the flanking introns, that is a forward primer, directed to the 5' donor site and a reverse primer, directed to the 3' acceptor site. Our sequencing results show that λ Dash 62 contains exons 12 to 18 and that λ Dash 151 contains exons 19 to 21. The DNA sequences surrounding the intron/exon junctions are shown in Fig. 2. The negative hybridization of the λ Dash 62 with λ hTg₃ (8), containing a segment of intron 11, reveals a gap between both clones, located in intron 11.

We previously reported on the intron 21/exon 22/intron 22 junction sequences in the clone named λ Dash 161 (34), obtained in the screening reported in this study followed by screening of recombinant positive phages by PCR. λ Dash 151 and λ Dash 161 displayed similar hybridization patterns, as shown in Fig. 1. An additional new clone named λ Dash 121 was obtained using the same approach. The predicted 63 bp fragment is included in exon 23 and was observed in one of the twenty stocks investigated by PCR. The presence in λ Dash 121 of a fragment of 0.7 kb which is negative for PCR 3.1 and positive for PCR 3.2 allows localization of the λ Dash 121 3' relative to λ Dash 151 and λ Dash 161 (Fig. 1). As expected, PCR 3.3 does not hybridize to λ Dash 121 and λ Dash 161.

The EcoRI map of each of the recombinant phages was compatible with the existence of an overlap of about 13 kb between λ Dash 151 and λ Dash 161 and of about 12 kb between λ Dash 161 and λ Dash 121. This has been confirmed by cross-hybridization using each phage as probe.

Sequence analysis showed that λ Dash 161 contains exons 19 to 22 and λ Dash 121 contains exons 21 to 23.

The four phages include 32.6 kb genomic DNA, covering 1815 nt of exonic sequence (from nucleotides 3002 to 4816 of the mRNA, according to the new cDNA numbering system (13)) distributed in 12 exons, from exon 12 to 23 whose sizes range between 78 and 219 nt (Fig. 2). In total, 7302 nt of intronic sequence were obtained (Fig. 2). Donor and acceptor splicing sites sequences are shown in Table 3. When we compared our data with general splicing consensus

sequences (35), we found that the GT-AG rule is respected in all introns.

We also established a general picture of the complete 5' region of the gene (Fig. 3a) by integrating restriction, hybridization and sequence experiments from a previous study (8–10) and our present data. Here, we present the physical EcoRI restriction map and the distribution of the first 23 exons. There are still two intronic gaps of unknown size, in introns 11 and 18, because no phages could be found. As has been reported, two homonogenic acceptor sites map in the amino domain of the TG (20). The detailed analysis showed that they are located at positions 5 and 1291 of the TG monomer within exons 2 and 18 respectively, out of the repetitive region. Comparisons between the exon organization and type-1 and type-2 cysteine-rich repetitive elements are summarized in Fig. 3b. Ten of the eleven type-1 repeat motifs span between exon 2 and 16, while the eleventh repeat is entirely located in exon 22. The 3' end of exon 20 and a fraction of exon 21 encode the three type-2 repetitive elements.

The knowledge of exon limits allowed us to demonstrate that the alternative splice variants characterized by Bertaux *et al.* (25) consist of a series of transcripts where complete exons or a group of consecutive complete exons were skipped. Bertaux *et al.* (25) found four transcripts smaller than previously known TG mRNA. Sequencing reactions showed that these transcripts were 844, 689, 532 and 313 bp, smaller than normal size. Here, we show that the regions absent in each of these RNA molecules correspond to exons 17; 17 and 18; 17, 18 and 19; and 17, 18, 19 and 20 respectively.

Figure 3c shows the seven alternative spliced variants, in wild type TG mRNA, that have been associated with the 5' region (14, 23–25). Exons 3, 4 + 6 and 22 are also alternatively spliced.

Finally, the TG gene was mapped using the Gene Bridge 4 radiation hybrid clone panel. The results of the PCR reactions (data not shown) for the 93 clones of the panel map the Tg gene 5.5 centiRays (cR) from the AFMA053XF1 marker, in the D8S557-D8S529 interval (144.8–144.9 centiMorgan (cM)) of chromosome 8.

These results are in perfect agreement with the data obtained previously by *in situ* hybridization to human chromosomes (4–6) and with that of Deloukas (36), who used forward and reverse primers localized in exons 47 and 48, respectively, to map TG on the same radiation hybrid clone panel.

Discussion

In addition to the first 11 intron/exon boundaries available from previous studies, we now report the sequencing and characterization of the intron-exon organization of exons 12 to 23 of the human TG gene. The corresponding gene region was studied in detail.

```

.....agattgctgtgaggattaaatattaaggtatgtatttagctagtgccctcacacataacag
atgagtagatagagcctacctattattactattatctcatgtgctgaagacacttcaactgacctactgatttctgtatg
gaagttattgtagaggtctactataatattctttcaagcttcagcctaagtgttattcttaaggacaactctgttgag
tgctgcctactgtgcaaggagcctttgtaatactgtgggtgaaatgtttgcaataaataaagaaggcctcc
ttatgttggaaccaggacatacgtggttctcagctacagagccacacagagcaggtggtcatattctgcttttc
tccttcctgactccag CCTTAAGCTT---EXON 12 (3002-3139)---GCTGGGACTG gtaaggagggatag
gcacctcaggtggccaagtgcacccctttttatttttagaggacagagccccacactgggagggcaagtgcagctgga
ctcttaggctcctcttagcagctggtgaaaaattccttagcttctcgggctccagttatctcactgtgtaacagg
agcaggaacttctaccagcctgaaagacaagtgggtggttctgctggggcctcagagtgatgtggacaatgtccggc
tggggctctagactggggacagaaggcgaccaggctgcactgctatgagtgagaaggaaagtgtttatgggacac
tcagggtgaagtcctcagtgcaattcctgaatgttctcccctggctctttccag GCCACTGCTG---EXON 13
(3140-3217)---ATTCCACAGT gtaagcgaagactgcagagttctcctcctgaccccccttgggtgggcatcactg
gtcagctcagctgtggttgcctaacgctggagactccatggcccacccctcagccaggtcccgggtgagccagac
gcatttcccaagtcccaggcctgcaagcttgcgctgacctggcacttctagttggaaggctcccagctttgcagg
tgaaggaacacagctgagactgtgaaaggagccctggttctcccctctgtacatctgctgggcaatttcagccactc
ctgtagtttcagcaaccgctcgcctctgaacctgcagcccaacccctctctgaggtcagccatgtgtaatccac
agtcacttagttcaaaagctggcagtgctccacatgagctcttctcctcagagattctacctctgaaagcaca
cctgcatcaccaggctcatgacctgaaaggtgggctcagcctgggatcccctcttctctggctcccacagacca
gtcctttacaagcactctctcctccacag GCCCGACAAC---EXON 14 (3218-3330)---CTGCCATAGAA g
taagggtctggaagcacaggatggagccgggactgtccccgctggtcagagatgattttctttttgttcaatag
ttcagcttagttaaaagctcacatgatttctcagctgggtgtttgctcgggctgctagaacctgtgagcctgtgacc
ttgggctatttacttacctttctgtgtcccagttgttttggtttaatacataagatcaggccacattttgtcctagct
aatgttaggtttgttttgagactccaagaaaatattcttcaagaactctttgaaaaataaacaacatgctatca
gttttcttaaggcatcttggccaaggccctatgcttacaggtgtcccagg.....
.....gactctagcttctgcaaatcatttctggaatttagtgaggctggtcaggagcatggctaggcttggg
tctagatgggtttgggctggaactgacctgctcactcaccagctgtgagttgttgcctccttctctga
gagatctcagtttccccacccaaaagcgggagaactgtgtccatgcctcatgaggtgttgaagcctcaatcaa
atcatgcagtaaaaggctgggaacagagagcactctcccactccacatctctcgtctggagaagccatcacc
tgttttttggaaaggacacttttgacatctgccagcagcaggttggccactagagcatcttggccacagtgactgac
atgaccocggctttgtctcag ACAGGAGAGT---EXON 15 (3331-3433)---AGTGCCAGT gtgagtagca
gcccctcctggcatggcttctcactctctgtggggcactgagcgtggagtccaagagctggctctgtcccagc
tctactgcttccatagctgtggggcagcagctgctgacctactatgctcagtttctcatctatgaaatcggggca
atattccccctctctgcagtttcttgggagaagcaaacgtgggagggagttaggtgtgcttgtgtctgagaagtgc
aaaccagcattgatccccatcaaaggggcccctagggcttggagggctggccggatgtgtggcctgcccgtctgg
ttccgcagtgaaagcaggtgatgtccaaaggcaatgtgcagatgtcctctctgtctaaac.....
.....gacctctcttctctgtaggtcctanactgactatgttgccttatttcaagaactgacctgct
acgtacttgttaactcctttaaagactcaagcttgaaggcagcagtggaagggagagatcacagactgagc
caggtccagfgtttgcctgttgcattcaacctcctggtggggagggcagccccagaccctggaaggcctgcaggc
aggtgggctgaggggtggcctgggtggtgggaggggtgattgggcatctgagcagggagtgggcactgattccccagc
caactcctctgtctgtcag GCCCAAGCCT---EXON 16 (3434-3634)---GCCTGTGAGC gtaagtc
atgaccocggggggaacgacagggcctgcatactgttctttagtccagactgggtgaagtagtagagaagctggga
aggcaggggtgggaggggagcaggaagtgcaggagggatgtagtgtgggtgagctctctgtgtgaaaacttacatc
tacttctgaagacgctcctcgtgggaactgaagtcaagaagaagcctgggaagggaaacccaagaactacactaga
gtgggcttctc.....
.....tgcctttagagcgcagagagctcgagtgc
tgagctggacatcaagctacacgagacattcatgcttagtctgtttaaactgagagcagatgggctcctcaagta
gagggcagggcagagaggaaggttagcagggccagggcagccctgagagcttgacaggtccagggaggggagagga
ggagggcccaggtgtgagtgagaaggagacccccaaagcagggatgctcagctcctcagggctggcaggtggccac
cagcgtccacttctctcctccag GCCCGCGGTG---EXON 17 (3635-3847)---GCCTGCCAAG gtgagt
ggcatcagagcatctatcctgcaccccgctcctctcagggctagggctgggaccagataggaagcgtggctgctc
catttccccaccaaatgtagcagcgtggtggtatgccagatggcaccagagatgagagagagctccatggagcatt
ttctagaaggagcagatgctctcagagactcctggagttaggcttcttctgtccctctctggaaccatctgctcct
tgagagaattgccattcttccaacttcttgggatgcttggattcaggaataactggcagtcagtgcttcaacagc
tggataagtg.....aaatggcctgctgtatcgaagatggcatggaagattccagagatgcccagacctcaaat
ataaagcaccctggttggaggtgagggcagactgagcttctgtgaggggtgctcagtaatttcagagacaagctgt
tatgctggaaagtacttagactcagaatgacaatgcaaatggcagtgcttgtgtgtttgagctcaatgggaaggg
ttatttttgcagaggaatccccaaacaaagaaaataactctacagcccttgcctctgctgacctctggtgcttgc
tcgag GGCCCCAGCT---EXON 18 (3848-4002)---CCAGATCCAG gtacatgctggccttccccacag

```

Figure 2 (Part I)

All exon borders and intron/exon junctions were localized precisely and sequenced. The positions of the first 23 exons of the TG gene were correlated with the repetitive organization of the protein.

The nucleotide sequence of human TG cDNA and the deduced sequence of the protein showed that 70% of each monomeric chain is characterized by the presence of three families of repetitive units, all of which include

strongly conserved cysteine residues (12, 15, 16). Type-1 motif is present in the N-terminal portion of the molecule and is repeated eleven times. A more refined analysis consider two subtypes of the type-1 repeat, 1A and 1B (16). Type-1A contains six conserved cysteine residues. The ninth and eleventh units possess only four cysteines and both units were assigned as type-1B. Up to now, the location of the extremities of the first ten

aggccttgactgaactcagggttacgggtgcagaaaacctgagggtcacattaacctgagggtttgagttcaacc
 tcactgttttatcaatgtgcaaaaaaaacaaaaaaataccagaaatgggaagtagtttggccaaggtaattaca
 tagcagcagaggcagaatccataccatctcatcattggccatctttctacataagatgtcaacagttactgagcact
 taatatgggcatcctt.....atgctaaggttgtttcaccagtgga
 agtatggtgactggctttagatgcatgagcctagctcttccctccacttatactctgtacaccatgtgaccaggcaggtt
 atttaacttctctgaaactgttttccctcagatgtgaaaaggggtcactattccttgaccccctgaagtaggggttgg
 ggaggattgaaggaagtaacataagccaagttctggttccctgggtgtggaccaacaaaactagctttctactctgtgt
 tcttgagctgaaaggtgtctgtcttctgttag **GTGAAGACTT---EXON 19 (4003-4159)---CATGACATTG**
 gtagttttctgtggttagtacctagaataagctatgcagcctctctgagctcagttttctcatctgccaatggag
 ctggtgaagaagcactgatgtcttcttctggaagcattgagggagagcctaattatttaaaacataaataacata
 gacttcttgccaatcgtttactagagttaaataatgtcaacaaagagcctggttctgtgtcgtgaacatctgctaata
 cttctcttgatttctgcgtgcagagaggtgaacagatgtgtgaaggatagcagggtggctgggag.....
caggaatctgggtgagctccctactgtgtctctgattcctgggtgacctgaaccaggttc
 atcccttctctggcctcatgttccctccctctgtaaaatgagacaccacatgatgcctatgtctcatcttccacagct
 attctaggtcatctctgcccctgcttaactcctccctggccctagcagagtacagtggggggtgacctctacctatcctgt
 gtcttacag **AGAGAGCCTT---EXON 20 (4160-4378)---CTGGGATGCG** gtaggtccactctctccctgga
 tatctctgtggagcctatgtgaggtcttaggaaagccacctcagcactggagagaggtactctggacatccagggc
 ttgaaacctccatttagtaacgagcaaaagtctcaatcactctactatgcaatgggtgtatgtctcatcttctggtag
 catggcagaaggagcccatatttgacgtgatcttctctgttggttattgtaaggtgacctggctggctgctgtcaa
 tt.....aagc
 tactctctggagattggcatctcagacacacctgacctgaaccaagtaaaacaccccctcttatttacctngaaa
 ccctgtctgaccctctcatctccctacaagcattgtcactcccgacagcaggaagctcgctgacttctcaccagct
 gcttccctgcttcttagagctctgacacacagtaggttctcagtgggacatttgttaaatgaatgagctcttgaactggt
 ttgaggtatttcttaccctgtgtgttctgggattgtaactgtgaagcattgtcttgattttaaactcatttctct
 gttttttctag **TTAAGTGTC---EXON 21 (4379-4528)---CAGACTCACT** gtaagttctgtggaggtct
 tcttgaagaagcaacctgcaatgaaatggaactcaacaggggttcccaatctgcaactattgcaacttctgggaca
 gataattcttctgtcaggggtggcctgtgcttggtaggataatttagtagcattggtagctctcagatgccagtagcac
 tcctccctgaagttgtgacagccaacatgtctccagacattgccaatgtctcttgggtggaaatggcagggagagc
 aaattgacctacctgtgagaacctattgggtacacaacataatacaagggccgacctt.....
aaagactggggagctgtggccagagagccagatgagcaagctgtgttggagcaaatggtatc
 agagagattgggtgtattcagggcatggaatgaagcaggtggtgtgtgaagcattgggtatgtgtgcagaacttagggt
 ggggtggcggcaaatcacagggccttattgatcagaacagtgaggacactgaagagtttaagctggaatgtctgag
 ttagatgtgtgacttgacaccttgtcaatgaccgagagcctgggagtaggagtcaggggattccagaggccattat
 tgacggctatgtcaatctattgggtctag **GTGTCACTGA---EXON 22 (4529-4699)---CAGTGTTTGA** gt
 aggtgctgggggtgaaatcagtcattggttccctggggactggggagtagtctcaagggctttttagaaaggagagaag
 ctggaggtgcatcttctcacaacttttggtttggaaaattttaatctgtgtagaagttggaagaaccgcaaatgaa
 caaataccatcacctagatccaaaaattgtaacatttcaactctctcttttttggaaatcatttggtagttaaattcc
 agatatcaggaaccttacttctaagtagtctcagcagcagagagatgaaattttatattaagatctgtggaggcct
 cattcttctagggcagcttgacttactaatctagttagcatttctct.....
ttcggaaagaattcaagtggtttacagaaagaaaatgccagtaaaata
 gttatagtagtaattagcactgaaggataggacagagaataaagtttatattatcaatggttaatgagaagaaaa
 tagttgtaaggactaacattaaatttccgccaacttctctgtgacaatgaggtagaagcagggagagctggtaagttgtaa
 ctctcagttttagatagggagcagtagtggtagctaccttacaaggaattgcttttctctgggacatttgtgttacc
 aagagctacgaatgggtattgactgcttgacctaacagctcttattggcttctctcagatgcccacaccttctgag
 tcagatttactaaatctgctttatttttag **TGATGCAGAA---EXON 23 (4700-4816)---TGCTTGACAG** g
 tgaggagtggtgggagatagcactcagaagaaggtgtgaaataagctctgctgagagttgctcagagcaaaccaa
 tcaaacattaaaaactataatttaaaacatactttatcaaaacactataaatgaaatgatttccagaatgcatatt
 caattcataacttaattgcagagttgacagtttatctactgtatgctatccacatggcacattgtaaggagtagaaat
 gaacaaaattagttggcata.....

Figure 2 (Part II) Sequence data of intron–exon boundaries 12 to 23 and their flanking intronic regions. The first and last ten nucleotides of each exon are indicated by capital letters and their flanking intronic sequences by lower-case letters. Numbers indicate the first and last nucleotides of each exon. The gaps (.....) represent intron regions whose sequences were not determined.

repeats was based on the limited available intron–exon junctions identified by Parma *et al.* (9) – exons 1 to 11 for the human and exons 12 to 16 for the bovine TG gene. In this study, we established an alignment of type-1 repeats according to our results about the first 23 intron–exon junctions of human TG (Fig. 3b) and based on the criteria proposed by Molina *et al.* (16): (i) repeats 2, 4, 7, 10 and 11 are each encoded by a single exon (exon 4, 8, 10, 16 and 22 respectively), repeats 1 and 9 by two exons (exons 2 and 3, and 14 and 15 respectively) and repeats 3 and 8 by three exons (exons

5, 6 and 7, and 11, 12 and 13 respectively); (ii) repeats 5 and 6 are a fraction of exon 9, the N-terminal limit of repeat 5 is ambiguous (15, 16); (iii) in repeats 1, 2, 3, 6, 7, 8, and 9 the conserved cysteine located in the C-terminal segment of the unit is interrupted by an intron (intron class 1); (iv) in repeats 4, 10 and 11 mapping in exons 8, 16 and 22 respectively, the conserved C-terminal cysteines are not interrupted by any intron.

Yamashita and Konagaya (37) reported the isolation from chum salmon eggs of a potent protease inhibitor with high sequence similarity with the TG type-1

Table 3 Exon/intron organization of the 5' region of the Tg gene.

3' End intronic sequences	Exon 5' end	Exon number	Exon size (bp)	Exon 3' end	5' End intronic sequences	Intron class	Amino acid interrupted
5' flanking region	5' untranslated segment						
		1	108	ATC TTC G	gtaagttctg	1	glu
acttttcttttctctag	AG TAC CAG	2	109	AGC TTC CA	gtaaggctta	2	glu
ctgtgtctcctctctag	G ACT GTC	3	98	GTG GCT T	gtaagtggga	1	cys
ctccttgaccacacag	GT CTG TCA	4	204	AAG CGA T	gtagtttcac	1	cys
ttgtgaaaatgttttag	GT CCA AGG	5	160	TAC AAC AG	gtaaggggag	2	arg
tctcattctctccaag	G TTT CCA	6	107	GAG ACA G	gtgagtgata	1	gly
tgctgtctttgtctcag	GT TTG GAG	7	144	TTC CGA T	gtaagtaata	1	cys
tgtggatttcctctctag	GC CCC ACA	8	186	TCT TGT G	gtgggtttcc	1	ala
actttgtctcatgcag	CT GAA GGC	9	1101	AAG AAA T	gtaagtctgt	1	cys
cattgttctctcccag	GC CCC ACG	10	585	CCA ACA T	gtgagctaac	1	cys
agttttatctcccctag	GT CCT GGC	11	240	CAG TCT A	gtagtggtgc	1	thr
ccttccctgactccag	CC TTA AGC	12	138	GGG ACT G	gtaaggaggg	1	gly
cttggctcttttccag	GG CAC TGC	13	78	CCA CAG T	gtaagcgaag	1	cys
cctctctctcccacag	GC CCG ACA	14	113	TGC CTA GAA	gtaagggtct	0	-
cccggctttgtctcag	ACA GGA GAG	15	103	GCC CAG T	gtgagtagca	1	cys
cttgtctctgtgtcag	GC CCA AGC	16	201	TGT GAG A	gtaagtcctg	1	ser
actttccttctcccag	GC CCG CGG	17	213	TGC CAA C	gtgagtggtca	1	arg
tggtgcttgcctgcag	GG CCC CAG	18	155	CAG ATC CAG	gtacatgcct	0	-
gtctgtctctctgttag	GTG AAG ACT	19	157	GAC ATT G	gtatgttttt	1	glu
atcctgtgtctctacag	AG AGA GCC	20	219	GGA TGC G	gtaggtccac	1	val
ctctgttttttctctag	TT AAG TGT	21	150	ACT CAC T	gtaagttctg	1	cys
aatctattgttctctag	GT GTC ACT	22	171	TGT TTG A	gtaggtgctg	1	met
tctgctttatttttag	TG ATG CAG	23	117	TTG ACA G	gtgaggagtg	1	asp
...ttttttttttncag cc ccccc	G.....	Consensus sequences	AG	gtaagt..... g		

Exon sequences are in capital letters, intron sequences are in lower-case letters. Intron class indicates which nucleotide in the codon is split by the intron. The consensus sequences of Shapiro & Senapaty (35) are indicated at the bottom.

module. This cysteine-rich motif is also found in 32 different proteins (16, 37, 38), suggesting that they arose from an early duplication event followed by additional gene duplications and divergence. However, the role that this repeat sequence plays in these genes and its relationship to proteolytic activity remains unclear. Molina *et al.* (16) proposed that the 11 type-1 repeats from TG would control the sequential proteolytic events of TG degradation, first generating the selective release of thyroid hormones and then complete TG degradation. These findings support the idea that type-1 motifs selectively and reversibly inhibit the lysosomal proteases. It is possible that the 11 copies present in TG may have specificities for different proteases activities.

The three type-2 repetitive elements map between exons 20 and 21. It is interesting to note that exons 17 to 20 disrupt the repetitive organization of TG (Fig. 3b), connecting the first ten type-1 modules with the three type-2 modules. Curiously, this region is the subject of four alternative splicing events (25) (Fig. 3c). Within the TG locus, the conservation of nonrepetitive sequences between two regions that are highly repetitive was hypothesized to be important in the structural organization and expression of TG. It has been speculated that the skipping of exons 17 to 20 by alternative splicing may result in a population of

alternative TG molecules with a role in the control of thyroid hormonogenesis. It is conceivable that partial or total excision of this segment of the TG would affect the protein structure and subsequently the coupling mechanism, resulting in a decrease of hormone production. Alternatively, it is possible that elimination by alternative splicing of exons containing type-1 motifs would result in an increase in hormone production secondary to a loss of protease inhibitor activity. In this context, three alternative splicing events were localized in exons 3, 4 + 6 and 22 (14, 23, 24), thus eliminating type-1 repetitive units 1, 2 + 3 and 11 from TG transcripts. In a previous study, we identified a heterozygous nonsense mutation in exon 22 in two siblings with hereditary goiter and markedly decreased tissue levels of TG (34, 39). The point mutation was removed from a fraction of transcripts by preferential accumulation of an alternatively spliced product missing exon 22.

In conclusion, the present study shows that the first 4857 bases of TG mRNA (including a 41-nucleotide 5'-untranslated segment) are divided into 23 exons, whose sizes range between 78 and 1101 base pairs. The GT-AG splicing consensus sequences are rigorously respected in all introns analyzed. The 5' region contains eleven type-1 and three type-2 repeat motifs and is subjected to seven alternative splicing events,

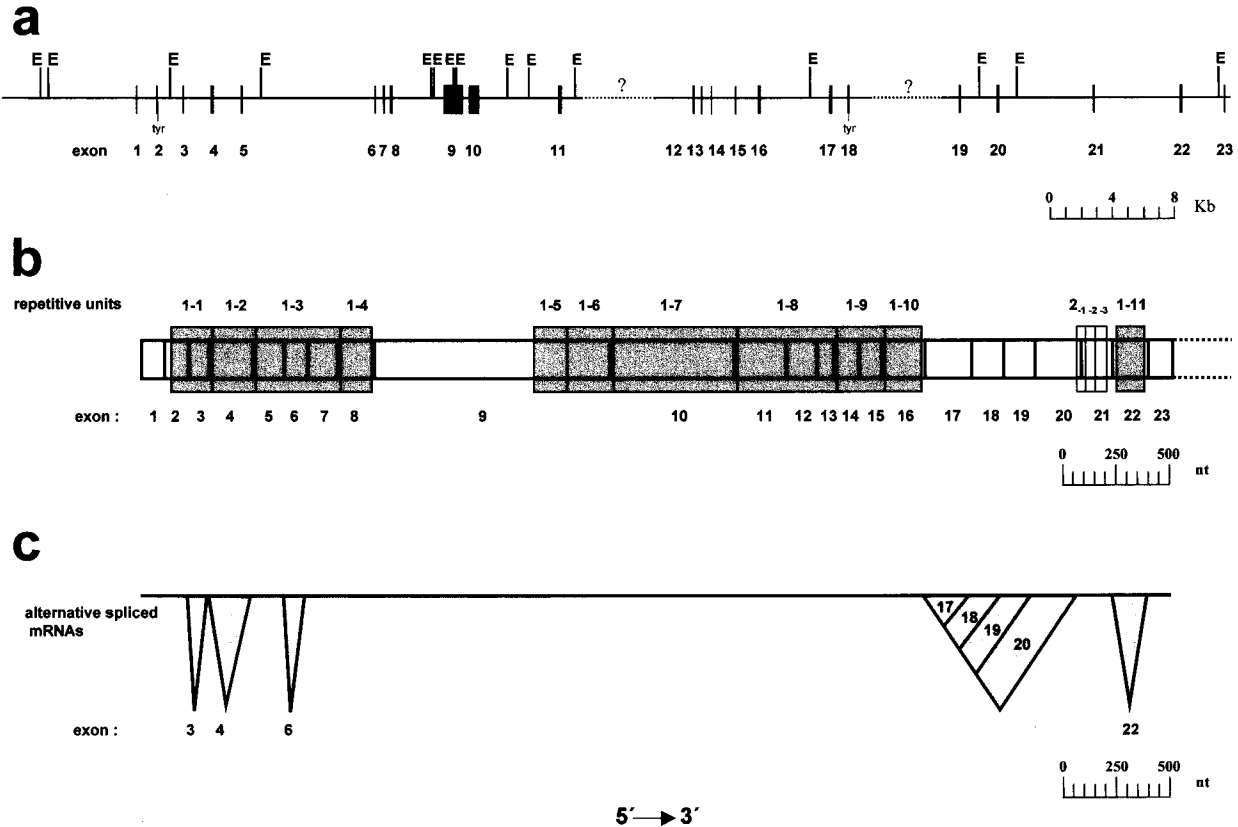


Figure 3 Schematic representation of the 5' region of the human TG gene. (a) EcoRI restriction map. The exons are indicated by vertical black boxes and the introns by a continuous line. EcoRI (E) restriction sites and tyrosine residues (tyr), involved as acceptor sites in thyroid hormone synthesis are shown. The dotted line denotes the segment of introns 11 and 18 that are not included in our phages. (b) Exon organization and their correlation with type-1 (shaded boxes) and type-2 (open box) repetitive units. (c) Map of exon skipping of TG pre-mRNA in the 5' region. Shaded areas indicate identified exon skipping in the alternative splicing events. Exons 3, 4 + 6, 17, 17 + 18, 17-19, 17-20 and 22 are alternatively spliced exons. The scale is drawn below each diagram.

which could be involved in the regulation of biosynthesis and release of thyroid hormones.

Detailed knowledge of the organization of the TG gene and the availability of the sequence of each of the intron/exon boundaries provide the basis for the investigation of families affected with defective TG synthesis.

Acknowledgements

H M Targovnik is an established investigator of the Argentine National Research Council (CONICET). F M Mendive is a research fellow of the Universidad de Buenos Aires. C M Rivolta is a research fellow of CONICET. This work was supported by grants from Universidad de Buenos Aires (EA 124 / 95, TB 80 / 98), CONICET (7260/96, 0853/98), FONCYT (05-00000-01591/ 98), Fundación Alberto J Roemmers (1997, 1998).

The nucleotide sequence data reported in this paper have been submitted to the GenBank, EMBL and DDJ databases under the accession numbers AF105681,

AF105682, AF105683, AF105684, AF105685, AF105686, AF105687, AF170486, AF170487, AF170488, AF170489, and AF237421.

References

- 1 Van Herle AJ, Vassart G & Dumont JE. Control of thyroglobulin synthesis and secretion. *New England Journal of Medicine* 1979 **301** 239-249 (Part I) & 307-314 (Part II).
- 2 Medeiros-Neto G, Targovnik HM & Vassart G. Defective thyroglobulin synthesis and secretion causing goiter and hypothyroidism. *Endocrine Reviews* 1993 **14** 165-183.
- 3 Baas F, van Ommen G-JB, 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.
- 4 Baas F, Bikker H, Geurts van Kessel A, Melsert R, Pearson PL, de Vijlder JJM & van Ommen G-JB. The human thyroglobulin gene: a polymorphic marker localized distal to c-myc on chromosome 8 band q24. *Human Genetics* 1985 **69** 138-145.
- 5 Bergé-Lefranc J-L, Cartouzou G, Mattei M-G, 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.

- 6 Rabin M, Barker PE, Ruddle FH, Brocas H, Targovnik H & Vassart G. Proximity of thyroglobulin and c-myc genes on human chromosome 8. *Somatic Cell and Molecular Genetics* 1985 **11** 397–402.
- 7 Mendive FM, Rivolta CM, Vassart G & Targovnik HM. Genomic organization of the 3' region of the human thyroglobulin gene. *Thyroid* 1999 **9** 903–912.
- 8 Targovnik HM, Pohl V, Christophe D, Cabrer B, Brocas H & Vassart G. Structural organization of the 5' region of the human thyroglobulin gene. *European Journal of Biochemistry* 1984 **141** 271–277.
- 9 Parma J, Christophe D, Pohl V & Vassart G. Structural organization of the 5' region of the thyroglobulin gene. Evidence for intron loss and 'exonization' during evolution. *Journal of Molecular Biology* 1987 **196** 769–779.
- 10 Targovnik H, Paz C, Corach D & Christophe D. The 5' region of the human thyroglobulin gene contains members of the Alu family. *Thyroid* 1992 **2** 321–324.
- 11 van Ommen G-JB, Arnberg AC, Baas F, Brocas H, Sterk A, Tegelaers WHH. *et al.* The human thyroglobulin gene contains two 15–17 kb introns near its 3' end. *Nucleic Acids Research* 1983 **11** 2273–2285.
- 12 Malthièry 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.
- 13 van de Graaf SAR, Pauws E, de Vrijlder JJM & Ris-Stalpers C. The revised 8307 base pair coding sequence of human thyroglobulin transiently expressed in eukaryotic cells. *European Journal of Endocrinology* 1997 **136** 508–515.
- 14 van de Graaf SAR, Cammenga M, Ponne NJ, Veenboer GJM, Gons MH, Orgiazzi J. *et al.* The screening for mutations in the thyroglobulin cDNA from six patients with congenital hypothyroidism. *Biochimie* 1999 **81** 425–432.
- 15 Gentile F & Salvatore G. 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. *European Journal of Biochemistry* 1993 **218** 603–621.
- 16 Molina F, Bouanani M, Pau B & Granier C. Characterization of the type-1 repeat from thyroglobulin, a cysteine-rich module found in proteins from different families. *European Journal of Biochemistry* 1996 **240** 125–133.
- 17 Swillens S, Ludgate M, Mercken L, Dumont JE & Vassart G. Analysis and structure homologies between thyroglobulin and acetylcholinesterase: possible functional and clinical significance. *Biochemical and Biophysical Research Communications* 1986 **137** 142–148.
- 18 Dunn JT. Thyroglobulin: chemistry and biosynthesis. In *Werner's & Ingbar's The Thyroid*, pp. 85–95. Ed LE Braverman & RD Utiger. Philadelphia: J.B. Lippencott Company, 1996.
- 19 Taugog A. Hormone synthesis. In *Werner's & Ingbar's The Thyroid*, pp. 47–84. Ed LE Braverman & RD Utiger. Philadelphia: J.B. Lippencott Company, 1996.
- 20 Lamas L, Anderson PC, Fox JW & Dunn JT. Consensus sequences for early iodination and hormonogenesis in human thyroglobulin. *Journal of Biological Chemistry* 1989 **264** 13541–13545.
- 21 Metzghrani H, Mziaut H, Courageot J, Oughideni R, Bastiani P & Miquelis R. Identification of the membrane receptor binding domain of thyroglobulin. Insights into quality control of thyroglobulin biosynthesis. *Journal of Biological Chemistry* 1997 **272** 23340–23346.
- 22 Molina F, Pau B & Granier C. The type-1 repeats of thyroglobulin regulate thyroglobulin degradation and T₃, T₄ release in thyrocytes. *FEBS Letters* 1996 **391** 229–231.
- 23 Bertaux F, Noël M, Malthièry Y & Fragu P. Demonstration of a heterogeneous transcription pattern of thyroglobulin mRNA in human thyroid tissues. *Biochemical and Biophysical Research Communications* 1991 **178** 586–592.
- 24 Targovnik HM, Cochaux P, Corach D & Vassart G. Identification of a minor Tg mRNA transcript in RNA from normal and goitrous thyroids. *Molecular and Cellular Endocrinology* 1992 **84** R23–R26.
- 25 Bertaux F, Noël M, Lasmoles F & Fragu P. Identification of the exon structure and four alternative transcripts of the thyroglobulin-encoding gene. *Gene* 1995 **156** 297–301.
- 26 Mason ME, Dunn AD, Wortsman J, Day RN, Day KH, Hobavk 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.
- 27 Hishinuma A, Takamatsu J, Ohyama Y, Yokozawa T, Kanno Y, Kuma K. *et al.* Two novel cysteine substitutions (C1263R and C1995S) of thyroglobulin cause a defect in intracellular transport of thyroglobulin in patients with congenital goiter and the variant type of adenomatous goiter. *Journal of Clinical Endocrinology and Metabolism* 1999 **84** 1438–1444.
- 28 Targovnik HM, Varela V, Frechtel GD, Cerrone GE, Copelli SB, Propato FV. *et al.* Molecular genetics of hereditary thyroid diseases due to a defect in the thyroglobulin or thyroperoxidase synthesis. *Brazilian Journal of Medical and Biological Research* 1994 **27** 2745–2757.
- 29 Chomczynski P & Sacchi N. Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. *Analytical Biochemistry* 1987 **162** 156–159.
- 30 Southern EM. Detection of specific sequences among DNA fragments separated by gel electrophoresis. *Journal of Molecular Biology* 1975 **98** 503–517.
- 31 Reed KC & Mann DA. Rapid transfer of DNA from agarose gels to nylon membranes. *Nucleic Acids Research* 1985 **13** 7207–7221.
- 32 Sambrook J, Fritsch EF & Maniatis T. *Molecular Cloning – A Laboratory Manual*. New York: Cold Spring Harbour Press, 1989.
- 33 Christophe D, Cabrer B, Bacolla A, Targovnik H, Pohl V & Vassart G. An unusually long poly(purine)-poly(pyrimidine) sequence is located upstream from the human thyroglobulin gene. *Nucleic Acids Research* 1985 **13** 5127–5144.
- 34 Targovnik HM, Frechtel GD, Mendive FM, Vono J, Cochaux P, Vassart G. *et al.* Evidence for the segregation of three different mutated alleles of the thyroglobulin gene in a Brazilian family with congenital goiter and hypothyroidism. *Thyroid* 1998 **8** 291–297.
- 35 Shapiro BM & Senapaty P. RNA splice junctions of different classes of eukaryotes: sequence statistics and functional implications in gene expression. *Nucleic Acids Research* 1987 **15** 7155–7174.
- 36 Deloukas P. NCBI. Gene Map'99. <http://www.ncbi.nlm.nih.gov/genemap99>.
- 37 Yamashita M & Konagaya S. A novel cysteine protease inhibitor of the egg of chum salmon containing a cysteine-rich thyroglobulin-like motif. *Journal of Biological Chemistry* 1996 **271** 1282–1284.
- 38 Lenarčić B & Bevec T. Thyropins – new structurally related proteinase inhibitors. *Biological Chemistry* 1998 **379** 105–111.
- 39 Targovnik HM, Medeiros-Neto G, Varela V, Cochaux P, Wajchenberg BL & Vassart G. A nonsense mutation causes human hereditary congenital goiter with preferential production of a 171-nucleotide-deleted thyroglobulin ribonucleic acid messenger. *Journal of Clinical Endocrinology and Metabolism* 1993 **77** 210–215.

Received 10 March 2000

Accepted 24 August 2000