European Journal of Endocrinology (2001) **145** 485–496 ISSN 0804-4643

EXPERIMENTAL STUDY

Genomic organization of the human thyroglobulin gene: the complete intron–exon structure

Fernando M Mendive[1], Carina M Rivolta[1], Christian M Moya[1], Gilbert Vassart[2] and Héctor M Targovnik[1,2]

[1]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, 1120-Buenos Aires, Argentina and [2]Service de Génétique Médicale, Hopital Erasme and IRIBHN, Université Libre de Bruxelles, 1070 Bruxelles, Belgique

(Correspondence should be addressed to Héctor 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, Avenida Córdoba 2351, 4to piso-sala 5, 1120-Buenos Aires, Argentina; Email: htargovn@huemul.ffyb.uba.ar)

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

Abstract

Objective: In order to complete the knowledge of the genomic organization of the human thyroglobulin gene, the present work was designed to establish the intron–exon organization from exon 24 to exon 35 and to construct a more complete physical map of the gene.
Design: Screening of two genomic libraries, and subsequent restriction mapping, hybridization and sequencing were used to characterize the recombinant phages.
Methods: Two human genomic DNA libraries were screened by *in situ* hybridization. Southern blotting experiments were performed to characterize the phage inserts. The Long PCR method was used to amplify the genomic DNA region containing exon 24. Intron–exon junction sequences were determined by using the *Taq* polymerase-based chain termination method.
Results: We isolated and characterized five λ phage clones that include nucleotides 4933 to 6262 of the thyroglobulin mRNA, encompassing exons 25–35 of the gene. The remaining exon 24 (nucleotides 4817–4932) was sequenced from the amplified fragment. In total, 8010 intronic bases were analyzed.
Conclusions: The present study shows that the five phages isolated and the amplified fragment include 59.4 kb genomic DNA, covering 1446 nucleotides of exonic sequence distributed over 12 exons, from exon 24 to exon 35. Using previous studies and our current data, 220 kb of the human thyroglobulin gene was analyzed, a physical map was constructed, and all exon–intron junctions were sequenced and correlated with the different domains of the protein. In summary, the thyroglobulin gene contains 48 exons ranging in size from 63 nucleotides to 1101 nucleotides.

European Journal of Endocrinology **145** 485–496

Introduction

Thyroglobulin (TG), the precursor of the thyroid hormones tri-iodothyronine (T_3) and thyroxine (T_4), is a homodimeric glycoprotein of 660 kDa synthesized and secreted by the thyroid cells into the follicular lumen (1, 2). TG functions as the matrix for T_3 and T_4 synthesis and in the storage of the inactive form of thyroid hormone and iodine. TG is synthesized as a 12S molecule, but forms 19S homodimers and even 27S tetramers. In humans, it is coded for by a large gene approximately 300 kb long (3), located on chromosome 8q24.2–8q24.3 (4–7) at 5.5 cR from the AFMA053XF1 marker (8). The number of exons has been estimated to be around 48 (9), each of which is separated by introns varying in size up to 64 kb (3, 8, 9). The 64 kb intron of the TG gene is an example of a large intron containing a small gene (10). This small gene codes for the human Src-like adaptor protein (hSLAP) and appears to be transcribed in the opposite direction relative to TG.

TG gene expresion is controlled positively by thyrotropin (TSH) through the modulation of the intracellular levels of cyclic adenosine monophosphate (cAMP) via its receptor (TSHr) located at the basal membrane of the cell (11, 12). Transcription of the TG gene is regulated by thyroid-specific transcription factors TTF-1, TTF-2 and Pax-8 (13, 14). It is mediated by binding to the TG promoter on their consensus sequences (15–17).

Human TG mRNA is 8.5 kb long (18). The general organization of the sequence showed a 41-nucleotide 5′-untranslated segment, followed by a single open reading frame of 8307 bases and a 3′-untranslated

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

RT-PCR fragment	Forward primer Nucleotide sequence (5′→3′)	Reverse primer Nucleotide sequence (5′→3′)
PCR 2.2	CTATCAGAGACGCCGCTTTTCCCC	CTAAACGCTCCCTGGTCAGACAACC
PCR 3.1	GCGGACGCTGGGAGTCACAGCTGC	CAGGACATGGGACACAGGCC
PCR 3.2	GGCCTGTGTCCCATGTCCTG	TGGCTCCTGAGGCTGAGAAC
PCR 3.3	GTTCTCAGCCTCAGGAGCCA	GGTCCGCATCGCACCGTCGTTCAC
PCR 4.1	ATCATGGAGTCCAATACCCAGGGGC	TTCCGACCCTGCAGACTATGTGTG
PCR 4.2	CACACATAGTCTGCAGGGTCGGAA	CCCCGCCAAATCCTCGGATGTGGG
PCR 5.1	GTTTGTATCTCAATGTGTTCATCC	GCCTTTGCTCTGTTGATGAG
PCR 5.2	CTCATCAACAGAGCAAAGGC	CACCGCATGTAGATACTTTATTGA

segment ranging from 101 to 120 bp (18–20). TG mRNA in human thyroid tissues is very heterogeneous because of 15 nucleotide polymorphisms, 10 of which result in amino acid changes (21–23), 11 alternatively spliced transcripts (22–27) and four polyadenylation cleavage-site variants (20). The preprotein monomer is composed of a 19-amino-acid signal peptide followed by a 2749-residue polypeptide (18). Eighty per cent of the monomer's primary structure is characterized by the presence of three types of repetitive units (18, 28). The remaining 20%, constituting the carboxy-terminal domain of the molecule, is not repetitive and shows striking homology with acetylcholinesterase (18, 29).

After translation, intensive post-translational processes take place in the endoplasmic reticulum, Golgi aparatus, apical membrane and follicular lumen, and include homodimer assembly, intrachain disulfide-bond formation, glycosylation, sialylation, sulfation, phosphorylation, iodination and multimerization (2, 30, 31). Several endoplasmic reticulum chaperones, such as calnexin, Grp94 and Bip, interact with TG during its maturation (32, 33) and may serve to prevent export of improperly folded TG proteins. This process is known as endoplasmic reticulum quality control (34).

Once TG has reached the follicular lumen, several tyrosine residues are iodinated and certain iodinated tyrosines are coupled to form T_3 and T_4. Five hormonogenic acceptor tyrosines have been identified and localized at positions 5, 1291, 2554, 2568 and 2747 in human TG and several tyrosines localized at positions 130, 847 and 1448 have been proposed as donor sites (35). The iodination and coupling reactions are mediated by thyroperoxidase (TPO) with a source of hydrogen peroxide. They take place in the follicular space in contact with the apical surface of thyrocytes (36). Hydrogen peroxide is generated by a metabolic pathway involving a flavoprotein enzyme. Recently, two cDNAs, ThOX1 and ThOX2 (37), encoding NADPH oxidases have been cloned. It was suggested that they constitute the thyroid hydrogen peroxide-generating system. ThOX1 and ThOX2 proteins are co-localized with TPO at the apical membrane.

We previously reported the partial genomic organization of the 5′ and 3′ regions of the human TG gene (8, 9, 38–40). Using Southern blotting, PCRs and sequencing analysis, we identified the first 23 (8, 39) and the last 13 (9) intron–exon junctions in the gene. In order to complete the knowledge of the genomic organization of the human TG gene, we report the establishment of the intron–exon borders of exons 24–35, including sequence data from splicing signals and the flanking intronic regions. On the basis of these results, we demonstrated that TG gene coding sequences are split into 48 exons. A more complete physical map of the gene was constructed.

Materials and methods

Probe amplification by the reverse transcriptase/polymerase chain reaction

Eight reverse transcriptase/polymerase chain reaction (RT-PCR) fragments: PCR 2.2, PCR 3.1, PCR 3.2, PCR 3.3, PCR 4.1, PCR 4.2, PCR 5.1 and PCR 5.2, were used in the screening of human genomic libraries and in Southern blot analysis (8, 9, 41). These eight RT-PCR probes map in the central and 3′ regions of the TG mRNA and together encompass nucleotides 3012–8410 (according to new cDNA numbering (19)). The DNA sequences of each of the oligonucleotides are shown in Table 1.

Total RNA was prepared from human thyroid tissue by the method of Chomczynski and Sacchi (42). Two micrograms total RNA were first reverse-transcribed with 200 U Moloney murine leukemia virus RT (Gibco BRL, Life Technology, Gaithersburg, MD, USA) and 20 U RNase inhibitor (Rnasin, Promega, Madison, WI, USA) in a 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 reverse primer for 1 h at 42 °C. The RT 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 $MgCl_2$, 4% dimethylsulfoxide, 2 U *Taq* polymerase (Gibco BRL) and 50 pmol

of each reverse and forward primers. No dNTP was added in the PCR reaction, so, as the RT reaction is diluted 1/5, the final concentration of the 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 amplified products (PCR 2.2, 1076 bp; PCR 3.1, 681 bp; PCR 3.2, 683 bp; PCR 3.3, 884 bp; PCR 4.1, 722 bp; PCR 4.2, 638 bp; PCR 5.1, 724 bp; PCR 5.2, 865 bp) were analyzed in a 1.5% agarose gel.

Screening of human genomic libraries

Aliquots of human genomic libraries constructed with λ Dash II (Stratagene, La Jolla, CA, USA) or λ charon 4 A (kindly provided by Dr T Maniatis, California Institute of Technology) recombinant phages were used to infect *Escherichia coli* XL1-Blue MRA, or Y1090 respectively. Infected bacteria were plated on Petri dishes and screened by the replica filter method using PCR 2.2, PCR 3.1, PCR 3.2, PCR 3.3, PCR 4.1, PCR 4.2, PCR 5.1 and PCR 5.2 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 M NaCl, 0.01 M NaH_2PO_4 (pH 7.7), and 0.001 M EDTA), 0.5% SDS and 100 μg/ml sonicated denatured salmon-sperm DNA.

The probes were labeled with [α-^{32}P]dATP by random priming (Gibco BRL). Filters were washed with 2× SSC (1× SSC is 0.15 M NaCl and 0.015 M 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.

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, the salt concentration was adjusted to 2 mol/l with amonium acetate, and the DNA was precipitated with ethanol.

Southern blot analysis

Restriction and blotting experiments were carried out using standard procedures (43–45). One microgram λ phage DNA was digested with 10 U EcoRI endonuclease (Gibco BRL). Prehybridization, hybridization and washing of the filters were performed as described for the screening of genomic libraries, except that washing was performed using 0.5% SDS.

In order to map the recombinant clones, the membranes were hybridized with the same probes used for the screening of genomic libraries, in separate experiments. Recombinant phage DNAs were also used as probes to check the possibility of overlap of contiguous phages.

Long PCR

The Long PCR technique is suitable for the amplification of long DNA templates. This approach was used to amplify the region containing exon 23/intron 23/exon 24/intron 24/exon 25 by primers situated in exons 23 and 25.

In addition, introns 11 and 18 were amplified by Long PCR with primers located in intron 11–exon 12 and exons 18–19 respectively. The DNA sequences of each forward and reverse oligonucleotide used for Long PCR were as previously described (9) except for the primers located in intron 11 (5′-gagtgagtccatctgtgtt-3′), exon 23 (5′-AGAATCAAAGGTGATCTTCGACGCC-3′) and exon 25 (shown in Table 2).

Long PCR was performed in 50 μl, using a standard elongase buffer (Gibco BRL) containing 100 ng DNA, 1.3 mmol/l $MgCl_2$, 200 μmol/l of each dNTP, 1 μl elongase enzyme mix (Gibco BRL) and 10 pmol each of the forward and reverse primers.

The samples were subjected to 35 cycles of amplification; each cycle consisted of denaturation at 94 °C for 30 s, primer anealing at 60 °C for 30 s, and primer extension at 68 °C for 15 min.

The amplified fragments were analyzed in a 1% agarose gel.

DNA sequencing

The exon and intron–exon-junction sequences were determined by the *Taq* polymerase-based chain termination method (*f*mol; Promega) from λ phage clone DNA. Primers were specially designed for each intron–exon junction. Oligonucleotide sequences and the positions of their 5′ ends are shown in Table 2. Exon–intron borders were characterized by alignment between the cDNA (19) and genomic sequences, using the PC GENE computer program (Intelligenetics, Inc., Geneva, Switzerland).

Results

Approximately 3.4×10^6 phages from human genomic libraries were screened by filter hybridization with eight human Tg cDNA probes (PCR 2.2, PCR 3.1, PCR 3.2, PCR 3.3, PCR 4.1, PCR 4.2, PCR 5.1 and PCR 5.2) corresponding to 5.4 kb Tg mRNA. One hundred and thirty-three plaques scored positive and four of them were randomly selected and purified to homogeneity. The DNA was prepared from the corresponding phages and digested with EcoRI. Four different restriction

Table 2 Oligonucleotides used as primers in sequence reactions. Exon sequences are in capital letters and introns sequences are in bold lower-case letters. The positions of exonic primers are indicated according to cDNA numbering. The positions of intronic primers are indicated from the beginning of the closed exon. Primers located upstream have negative numbering, while those located downstream have positive numbering.

	Forward primer		Reverse primer	
Exon no.	Position of 5′ end	Nucleotide sequence (5′→3′)	Position of 5′ end	Nucleotide sequence (5′→3′)
24	4817	ATTGCACAGAGGACGAGGCC	4932	CTGGTCAGAAGTCATGCAGGC
25	4933	AAACGAGATGCACTGGGG	5027	ACAGCTGGAGAATCTTGACCATGGC
26	5049	ATCCACCACAACACTTCAGAAACGC	5146	TGGCTCCTGAGGCTGAGAAC
27	5239	ATCATCTGTGGGTTGCTGAGC	5397	GATGAACTCCTGGTCTCCAAGACGC
28	−51	**tacccagtgacagcacactcaagc**	5463	CCAGAGATAAACCTGCTGAAAATTGG
29	5468	ATTCTGACATGGGGTCTCGGCC	5547	TGCTTCTGTTGGAGATGCTGG
30	5552	TGACAACAGAACTTTTCTCC	5686	**ac**GAGAAAGGCACCATAGG
	5640	CAAGCACCTGTTTTCAGCCC	−14	**agggagtcaacttcctgcacc**
	−210	**gaactattcctgtctgaccc**	−288	**atctggtccacctttcacc**
			+126	**ccacagtgatccatgagttatgacac**
			+638	**tgggtgttctctttcttctcc**
			+1014	**aaatgcctggtagagatccc**
31	5691	TGTGCAGGAGCACTCTTTCT	5789	TCATCACACACCTGTGCCTCT
	−996	**gtgtcataactcatggatcactgtgg**		
32	5882	TGAAGAACTTTTACACTCGCCTGCCG	+27	**gtgagacacaagtagcttaccc**
33	5981	TTGAATGTGAACGACGGTGCGATGCG	+49	**aactgcagagtactgtgcg**
34	6056	GAGGAGAGGTGACATGTCTCAC	6199	TGGGCTTCTGGTACCATCCG
35	−65	**gccttctgaactcgatgatacc**	6262	CTTTCTCTGTGAGGGAAGGC

patterns were obtained, leading to the identification of four recombinant phages termed λ dash 12, λ dash 54, λ dash 72 and λ ch 2. Southern blot experiments with the same probes described for the screening of genomic libraries were used to localize the phage inserts with respect to the mRNA sequence, and to determine their relative orientation. Figure 1 shows the localization of λ dash 12, λ dash 54, λ dash 72 and λ ch 2 clones according to their positive hybridization fragments.

Clone λ dash 72 is located 5′ to λ dash 12, since the exonic fragments of λ dash 72 hybridized with cDNA probes PCR 3.2 and PCR 3.3 and did not hybridize with PCR 4.1; the λ dash 12 exonic fragments, however, hybridized with PCR 3.3 and PCR 4.1 and

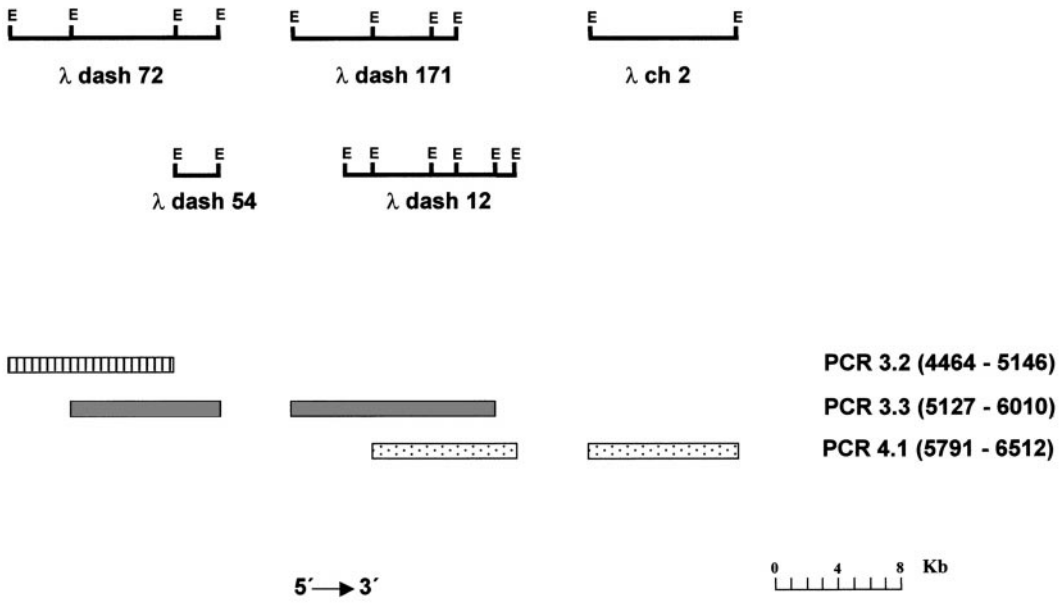

Figure 1 Relative positions and hybridization patterns of the λ phages used in this study. The EcoRI (E) fragments scoring positive for hybridization are shown, drawn to scale in the 5′-to-3′ direction. The lower bars represent the hybridization patterns with cDNA specific probes obtained by RT-PCR. The positions of the first and last nucleotides of each probe are shown in parentheses.

did not hybridize with PCR 3.2. Clone λ dash 54 was tentatively located in the same region of λ dash 72, on the basis that both phages contained the same 2.8 kb exonic fragment hybridizing only to PCR 3.3. Cross-hybridization analysis showed an overlap of approximately 5.3 kb between λ dash 72 and λ dash 54.

We previously reported the identification of the intron 30/exon 30/intron 31 junction sequences in the clone named λ dash 171 (46), obtained in the same initial screening of the present study. As shown in Fig. 1, λ dash 12 and λ dash 171 displayed three exonic fragments with similar hybridization patterns. The presence in λ ch 2 of a fragment of 9.5 kb which was negative for PCR 3.3 and positive for PCR 4.1 allowed localization of the λ ch 2 3′ relative to λ dash 12.

The possible overlap between the pairs λ dash 54/λ dash 171, λ dash 171/λ dash 12 and λ dash 12/λ ch 2 was confirmed by cross-hybridization using each phage as probe. Altogether, the five phages contain 52.1 kb contiguous genomic DNA.

In order to identify the intron–exon boundaries of the TG gene and to analyze the regions responsible for pre-mRNA proccessing, 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). The intron–exon junctions and splicing sites were sequenced from exonic and intronic primers (Table 2). Our sequencing results show that λ dash 72 contains exons 25–27, λ dash 54 exon 27, λ dash 171 exons 28–32, λ dash 12 exons 29–34 and λ ch 2 exon 35. The DNA sequences surrounding the intron–exon junctions are shown in Fig. 2.

Unfortunately, the characterization of recombinant phages showed that the exon 24 was not included in the positive phages obtained in the initial screening. Consequently, the genomic DNA fragment between exons 23 and 25 was generated by Long PCR. The 7.2 kb amplified fragment was sequenced with exonic primers and the intron 23–exon 24–intron 24 junctions were determined (Fig. 2).

The five phages and the amplified fragment included 59.4 kb genomic DNA covering 1446 nucleotides of exonic sequence (from nucleotides 4817 to 6262 of the mRNA, according to the new cDNA numbering system (19)) distributed in 12 exons, from exon 24 to 35, whose sizes range between 63 and 192 nucleotides (Table 3). The number of nucleotides of intronic sequence obtained was 8010 (Fig. 2).

We established a general picture of the complete gene by integrating EcoRI restriction, hybridization, Long PCRs and sequence experiments from previous studies (8, 9, 38–40) with our present data (Fig. 3). Twenty-nine different recombinant phage clones were isolated and characterized. In total, 220 kb of the TG gene were analyzed. These results demonstrated that the number of exons in the human TG gene is definitely 48, the exon size ranging between 63 and 1101 nucleotides (Table 3). We previously reported two intronic gaps in introns 11 and 18 (8); these were amplified by Long PCR in the present study. Another gap in intron 40 was considered (9), since the cross-hybridization between λ dash 31 and λ dash 56 revealed the existence of an intronic overlapping region of approximately 1 kb, indicating that the complete intron 40 had been cloned. However, there are still three intronic gaps in introns 35, 41 and 43. Our restriction analysis showed that intron 35 contains more than 11 kb. Intron 41 corresponds to large (64 kb) intron containing the human Src-like adaptor protein gene (10); the size of intron 43 has been determined by van Ommen *et al.* as 17 kb (47). Donor and acceptor splicing-site sequences of the 48 exons are shown in Table 3. When we compared our data with general splicing consensus sequences (48), we found that the GT–AG rule is maintained in all introns.

Discussion

In addition to the first 23 and the last 13 exons intron–exon boundaries available from our previous studies, we now report the sequencing and characterization of the intron–exon organization of exons 24–35 of the human TG gene. All exon borders and intron–exon junctions were localized precisely and also sequenced (Table 3). A more complete EcoRI restriction map of the human TG gene was constructed, and the relative positions of the 48 exons was established (Fig. 3).

Knowledge of the structural organization of the human TG gene will help to elucidate the functions of the different domains of the protein. The highly organized internal protein structure of the TG includes cysteine-rich repetitive units (8, 18, 28), hormonogenic sites (18, 49) and receptor-binding domains (50–54).

We analyzed the relationship between the three families of cysteine-rich repetitive units (8, 18, 28) and the intron–exon junction organization (Fig. 4). The monomer contains 11 type-1, 3 type-2 and 5 type-3 repeat motifs.

Detailed analysis of the repeats shows the following distribution. (i) Type-1 -2, -4, -7, -10 and -11 repeats are each encoded by a single exon (exons 4, 8, 10, 16 and 22 respectively), repeats 1 and 9 are each encoded by two exons (exons 2 and 3, and 14 and 15 respectively), repeats 3 and 8 are each encoded by three exons (exons 5, 6 and 7, and exons 11, 12 and 13 respectively), and repeats 5 and 6 are a fraction of exon 9. (ii) The three type-2 repetitive elements map between exons 20 and 21. (iii) The type-3 domain includes two subtypes, 3a and 3b, and maps between exons 23 and 37 (3a-1, between exons 23 and 26; 3b-1, between exons 26 and 30; 3a-2, between exons 30 and

```
.............................tgtgtgtgtgtgtgtgtgtgtgtgtttgggatgtggggna
ggggcaggggatgtgtctgctctgccctcagcatcccccgggaaagccaggtgagtgaccgtcccatggtgcttgca
g ATTGCACAGA---EXON 24 (4817-4932)---TTCTGACCAG gtgaggtggggcagccacgtgtggttctgc
tcctcatccgctgtggatcag atgtgctctgaggagcgggcagcaggctggccaggcgatggaatgaggttgtcgag
agctgatcctctgttacctccatttcttcaaagcaatgggactttggcaaattaaagctgtgatagtccagagacaga
cccaagatgaagctggagtaaggggcgtctgcttnagccttgtggcttcctgacgagtcagtcaggagagaaccacct
.....................atatatctcagaacctcactcagactcactttctggaactctctagctctcttct
gaatggaatcaattgctccttcctctctaattacctcttctttaattatatgtatctttgcagctaacactccgggct
gcaagttactgcttacacgtctgtctcctccaatagaccaggagcaactaacttacctttgtagccctggtgcctagc
catggttagggttggatgaatgtttgttggattgaattataaagtattgcaggataataatgcagcatctttccatct
ccag AAACGAGATG---EXON 25 (4933-5041)---TTGAAAAAGG gtaggttggtcaggctggttggcttag
gcccgcggtggcttcacacggtcatgtttggagctggtttccagcaggtgcatgtgaggaggagccagactgtcccgc
tccccactccacaccctcagtctggcgtctctgagctccatcctttggcagactcagcacagcacgctcctcaggcag
gcactgatctcccctgggactgcaaagatatttccacatcattctcctttgtgccttcaggtttgatcccattgcaat
ctagcaaggattga.............................tttctacacatacccttggacccttgaag
acgtgttgggcatttgtcccgatctcctgggtttggggccatggccttggctggccccacccctcttcatgcacaggca
catccagaagctttgtggcacagagagcatctcatcttggcccacacgctgcctggagcactgttgcagctaagtcaa
caaactgtcctgtgagatcagtaagctctgtccaactctgccatgttttgaggtcttttaaaatttgttctgccttt
cccccag GCCAAGGATC---EXON 26 (5042-5233)---GTTCAAGGAG gtaatgttggcagtgagggccagg
gcctaacaagggatggggagcacacagggatgcagaagccaggacacacaacatggagctgcatggggagtacagtgt
gagtgcctacacccaaatagaaggtgacagtcaagtacacaatacacacatgatactcacattcacattccacatgtt
gcacactgtgctg.....................................................aggcaggat
tgtctcaatttaaaaccaggctcttgaattctagtgtctatgtcttctcctgagacgctgtcacctatctttatttgc
aaatgctctcaggggacagagaagagtctctaaaggtccccctccatcacattgacctctcctacctcttgcgattct
cag GTGCCATCAT---EXON 27 (5234-5401)---TTCATCAAGA gtaagtctttgccatttgtccatattct
ttcaaaattactcttcacaccagcaaggccctctgctactttcttatgagccctccttgtggtctgaggagcttatac
ttctgaaaaaaaaaaaaactttaccaatcatactggatggaaacaaccctagcttttcagcagaactctatgggaagc
atccatttagctcctcagattgcaggtggact.....................................cagta
ttcccattctcaaaatcgctgtccacatttaccaatgaggcaaagagtctctgaggagttaagtgatgtgggaaactc
acacagccctcgagtgtcccaggaggtttactttcttcaagccaatctccagttctttaagtgactctcccgagtcat
aggttcacaacccctcatggaagtagtccctttggtgccctgagtagtacagccatgtctgaccagctctggttttgt
actagggagctggaatagagctgcaagaagcttctgggttacatcttacgtgggaaacagagttttcaggcctaggaa
gagcctgtgtcaagggagatggggctattgcagtaatggtgggtacccagtgacagcacactcaagctcataaaaata
aacatcttcctttgcag GTCTGACACC---EXON 28 (5402-5467)---CTCTGGAAAG gtgagctccgtggt
ggaagaggggttagcaggacgctgattacatgagattgtctggcaccacctctcattcacagggcactctatttctg
tagaggaataggtagagagtcactttccaaatgcatactttctgtggaatagaagggctccaactcatcagataagat
gccaaagtctaaaacaaaaatgaatactgtggaaacccaagaagggatggatagaggaagcactaggttcccgcgaga
taaagcttgctgaccacaagacacccttgttcctaggaggngagttcagggccattgcaaagcacaggaggcatttgt
ccaagtggactcagcccnatgtgccctttca.....................................gttaaa
ttctaattcatttactggccacttgagtggtctgggataaattattttgcctctttaagtctcagttttctcaagagt
aaaatggaaatcttgttccctacctacacagggagttataaacagcaatatttatgagcatattttgttaactcgaaa
gggctttgcgtatgttaataatttcaatatcacagataatgatgttctatcactgagaggttggatgatgttccatat
aaaaaagagtgcagatccatgtcagggcatctgttgcagatcttctctaagtcacctggctttagggcctgatttttc
tcacattgctactaccattttgttgaccagtggagtactacccattctccccaacattgcaacaactctttttttcc
tcctag ATTCTGACAT---EXON 29 (5468-5548)---ACAGAAGCAG gtactgacccccaaaccttcttaca
cacttgggtgtctgaatgcagagactaaatgggtgcaccaagagtttaatcaatgaacggatgtattgacatcactct
attctgtatccatggactctcctttaatcttttaacccaattatccagctcataaatatgggaagctcctcagatggg
caattgtcacagaaagtaaggcaaaatcactgcaaattcataagtgttactggaaatgaagacaaccataataataaa
acctgaaaaaagaaacaagagcaaggttcacctgactgtggccaatgggcttcacttatgtcagccctgaaaagtata
aacaaagtgtaaatattctg..............gggactgcaaacagtgtggtgatgagctagagttagcagggct
gttgtgttgctgaagtgcaaatcaagacaatccgagttgaggctcaaaatgcaagtagagcccagattaagaaaagaa
tctggacttattttctgtaggcagtggagactccaggaagagctgggatcatggcatgtaaatacctgaagacagagt
cagggaatggttgagctaaaactttcaagtttgggttaagtggagctgaaagctgactttgtagatgaagactaccat
tttggagagaggctggacttccatgtggcttcagaagacaaaaagaggatcagtggttcaaagttacaaaaagaggac
ttgactatacaacttcctaatgatggagtgggctgctgtggttagcacaatgatgtcatgtggagctatagaggcatt
agacagaagtctccaggggaaaacttctgcattaggtgaaaaggtggaccagataacttccaaggctttatccaaacc
cagagtctaagattttgccatcccataatttttaggacttcctggggccttctgaactattcctgtctgacccaactc
ccagctctactgacntcaatcctgctctggtttcacacagcataaaaatgtctctgtctctatctctgacactttctc
```

Figure 2 Sequence data for intron–exon boundaries 24–35 and their flanking intronic regions. The first and last 10 nucleotides of each exon are indicated by capital letters; their flanking intronic sequences are indicated by lower-case letters. Numbers indicate the first and last nucleotides of each exon. The gaps (……) represent intron regions whose sequences were not determined.

tctctgtctctctctgtgtgtgtgtgtgtgtgtttctttcttgtgtttttctcatattcactagagttaaaggtgc
aggaagttgactccctgcttctttttcag **GTTTGACAAC---Exon 30 (5549-5686)---TGCCTTTCTC** gt
aagtatccttagaactcattcttcttcttccagacactgtagtcaggcatcacaggccaagtctggcaactcctgaga
cacagataaggccaaattttgtgtcataactcatggatcactgtggatgatataaaagaaagatagcacataacccat
atcttcaagaaggttacagtgtagttaatctctactctgcttccataacctatcttcatccacctgtctctccatcta
tccgtccattcacccatccatccatccatctacccatccatcccattcatccatccatctacccatttatccat
cccatccatccttccatcccatccacccatccatccatccatccatccatccatcccattcatccgtccatctttgca
tcccatccatccatccattcgatccatccatccatccatccatccatccatccatccatccatccaccccatccaacca
tccatccatccatgtatccatgtatccatcccatccaaccatccatccatccatccatccatccatccatccatccat
ccatctatccatccatccatttggcagttattaagtatgcattatgttcaaagctttgtgtttgattttggagaagaa
agagaacacccaaaatgaatcaaatcaaccatgttccttgcccttggggagctcatatataccaggggcacaggctga
cttgggcacaggtgcagggagaccattcacaaagttcacagttttacagtcctgtgacctctggtctgcagatcctg
acttcctctgagatggccatgggtagatccaatgcaatgttactcttttgcctcttgtattgctgagtttgaggtttg
tcccttatgccttcaactctgcctttattcaggagcacacggccccctccccagctagcagaaccccagtgtttttgc
atgggccttccagactggatgatcatgacagtctcaggcctctgccctaactaggagggatctctaccaggcatttcc
ccacccagagaatcctgtagagattattctcccctgtagaccccacaaaaactaaaatcacactactctcttgcctgt
ag **GTTGTGTGCA---EXON 31 (5687-5863)---CGGAAGAAAG** gtgagcacttggagagatctgcataaact
gtatttccaatgttctgctggctctgtggttttagaaagatccacattagtttcttatttaaaaaacattttctttat
acttttattggtttcaagaagatgggaagatggatccaaactttcacggttagaatcaataaactactacgtaattgt
ctcagtttaaacaaaaaatatgaaaagggttatttgggggcagtatt..............................
.....................ccacagaaaagttgagtagttttcttggtctgttaggcctatttttatttctccata
tgctttctagcttccaacattttgttgctgtgtctccacttttttaaaaaaaatctttgtccctgtggacttgccatc
attttagtgagatttgagttaggaaatgaggtttaatatgaatttaatatgtcatctttaattggaaactttcagtct
gtattttactcatttgtcctcatattttcatgactacagcaaatctctaagtaatgtattttctttcttcctctatga
ag **TTATACTGGA---EXON 32 (5864-5975)---TTTCTAATGG** gtaagctacttgtgtctcacccctaatgt
ttattatgaataatacttcctcaaagaagatgacacaatcacagctaaaaacagaagcatacccagcactttgggagg
ccgagctgggctgatcacgaggtcaagagatcaagactatcctgaccaacatggtgaagccccgtctctactaaaaat
acaaaaattagctgggcatggtggcggcagatgctagtagtcccagttactcaggaggctgaggcagaagaattgctt
ganctcaggcggtggaggttgcagtgagctgagattgcgccactgcactccagcctagcaacagagcgagactctgtc
aaaaaaaaaa.........................gatcctgggttagcttggttggagcctggcacataataggca
cagaatatatgcatgatgaatgaatgggtgaataaatgattaatggcaggacagtattcctgagaggagtagaatttt
gttttcagaaagaacctcaaagctaagaatccactcatgcatattgaccaaagcacccccagtttaagtaggggtaa
aaaaataagctagttccccaaagcaagaatgactaccatcagcccacccagtaggtcctgggtcaccagtgaaaacct
tcaggcctgctctttctcttcctatgccag **GTTCTTTGAA---EXON 33 (5976-6055)---CAGTTAAAAG** g
taataatggtaacaacttcctctcccctgcgcacagtactctgcagtttagaaaacacattcacatctatgcttttac
aaatcctcagcatctcctatcctagttcacagataaggaaatgggggcttggagagggcaactggctaataaatgtgg
tagctgagacatggctcctagtccagtgatcatcactttacatgtgcagttctacctggccaaacttccttccttcta
ctcttggcttggatatcttttgttcatctttcaacacttaggccaagcagcttggcgaactttggatccatttgcatg
atgtgtacttaatgcttggttgattcccttactagtgtgtgattctggaaaagtcacttaatctcc............
............aaagctaccttcctcgtaaggttatgtaggaccaagtgaagtgataccaacagagacctagcacag
ccctaagcacatgttaggtgcttgataagtagctctatccctttgctcttagatcagaggctccttgagttcaggtac
caggtttttcagcagtggctgagatagtatctgaaatatagtggtgtatcccaaatgtctgctgaacaatgtacttgc
agaatttttctcatttattagactcttccattgtactcagtttcctgattgtggtttttgttttttttttttccacc
ccag **GAGGAGAGGT---EXON 34 (6056-6199)---CAGAAGCCCA** gtaagtaccctcctcatgacagctata
tggacgtctttagttaactatttcccagccatgcattaggacaatattcttggattggtgagtagattaatgaagact
cattgggttcaagcaacagaaatagattctgtttaacttaagcagacaagaaacttattgaaagtttgctcattgact
caaaggaaggatgagcaactaggtctcagtaaagacggaaatggaccatttccatggatgtatgtaacaagacaaata
ggcaacttcttcagtgtgccatgatcatcattaatctgctccatagggtttagtggttatattc..............
....tcgcattcgctttgccctgcatggca
attattgaaacttgaggcaagaagctgccttttcagcttggtccttgagagacgataacatagagcaagccccagc
ctctctgcagtaggcaggtggtacaaggagagggatcctgaactgttctaagccacttatgttgagattgtttgttag
ctcagcatcatccagcccaccctgaccaatacaggttgggacaatctgtctgccttccaagtccaattttacaatcac
catcttatacttatattaatgccttctgaactcgatgataccatgggggtagaaaagaactgaaagactgctttttcc
ttttcag **TTGCTCAAAA---EXON 35 (6200-6262)---ACAGAGAAAG** gtaagttcattgtctttttatctc
ttggatgagagtacccccctcattcatcttctttctcctcttcccccttgcttttgtacagaagttgatagcttgcata
tcactcgcattaattccagctcctatgaattaaacacatgtataagtggggtgagtaccagcaacatgcccactttgc
agatgagaaaattgactcacgggaatgaaatgatttgtggaaattatgtggctgattgaagcagagtaagacctaagt
cgcgttctcccagttt...

Figure 2 continued

Table 3 Intron–exon organization of the TG gene. Exons 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 and Senaphaty (48) are indicated at the bottom.

3′-end intronic sequences	Exon 5′ end	Exon no.	Exon size (bp)	Exon 3′ end	5′-end intronic sequences
5′ flanking region	5′ untranslated segment	1	108	ATC TTC G	gtaagttctg
actttttcttttcctag	AG TAC CAG	2	109	AGC TTC CA	gtaaggctta
ctgtgtctcctcctag	G ACT GTC	3	98	GTG GCT T	gtaagtggga
ctccttgtacccacag	GT CTG TCA	4	204	AAG CGA T	gtagtttcac
ttgtgaaaatgtttag	GT CCA AGG	5	160	TAC AAC AG	gtaaggggag
tctcattctctccaag	G TTT CCA	6	107	GAG ACA G	gtgagtgata
tgctgtctttgctcag	GT TTG GAG	7	144	TTC CGA T	gtaagtaata
tgtggatttcctctag	GC CCC ACA	8	186	TCT TGT G	gtgggtttcc
actttgtctcatgcag	CT GAA GGC	9	1101	AAG AAA T	gtaagtctgt
cattgttcctccccag	GC CCC ACG	10	585	CCA ACA T	gtgagctaac
agttttattcccctag	GT CCT GGC	11	240	CAG TCT A	gtagtggtgc
ccttccctgactccag	CC TTA AGC	12	138	GGG ACT G	gtaaggaggg
cttggctcttttccag	GG CAC TGC	13	78	CCA CAG T	gtaagcgaag
cctctctctcccacag	GC CCG ACA	14	113	TGC CTA GAA	gtaagggtct
cccggctttgtctcag	ACA GGA GAG	15	103	GCC CAG T	gtgagtagca
cttgtctctgtgtcag	GC CCA AGC	16	201	TGT GAG A	gtaagtcatg
actttccttctcccag	GC CCG CGG	17	213	TGC CAA C	gtgagtggca
tggtgcttgcctgcag	GG CCC CAG	18	155	CAG ATC CAG	gtacatgcct
gtctgtcttcttgtag	GTG AAG ACT	19	157	GAC ATT G	gtatgttttt
atcctgtgtcttacag	AG AGA GCC	20	219	GGA TGC G	gtaggtccac
ctctgtttttttctag	TT AAG TGT	21	150	ACT CAC T	gtaagttctg
aatctattggttctag	GT GTC ACT	22	171	TGT TTG A	gtaggtgctg
tctgctttatttttag	TG ATG CAG	23	117	TTG ACA G	gtgaggagtg
cccatggtgcttgcag	AT TGC ACA	24	116	TCT GAC CAG	gtgaggtggg
atctttccatctccag	AAA CGA GAT	25	109	AAA AAG G	gtaggttggt
ttctgcctttccccag	GC CAA GGA	26	192	CAA GGA G	gtaatgttgg
ctcttgcgattctcag	GT GCC ATC	27	168	ATC AAG A	gtaagtcttt
acatcttcctttgcag	GT CTG ACA	28	66	TGG AAA G	gtgagctccg
ttttttttcctcctag	AT TCT GAC	29	81	GAA GCA G	gtactgaccc
cctgcttctttttcag	GT TTG ACA	30	138	CTT TCT C	gtaagtatcc
actctcttgcctgtag	GT TGT GTG	31	177	AAG AAA G	gtgagcactt
ttcttcctctatgaag	TT ATA CTG	32	112	TCT AAT GG	gtaagctact
tctcttcctatgccag	G TTC TTT	33	80	TTA AAA G	gtaataatgg
ttttttttccaccccag	GA GGA GAG	34	144	AAG CCC A	gtaagtaccc
gctttttcctttttcag	TT GCT CAA	35	63	GAG AAA G	gtaagttcat
ttgccttctctcctag	TG TCT CTG	36	135	TTG TCG G	gtaaggggag
tccctcttttctgcag	AA TGT TCC	37	165	AAG CCA G	gtaagcccaa
tttctccttccaacag	GA ATC TCT	38	220	AAG CCA AG	gtatgggttg
cagtctgtatctgcag	G GCC AGC	39	94	CAG AAT GTG	gtgagttcaa
tctccaatacccacag	GCC CCT AAC	40	160	AGT TCT G	gtgagttgct
ttcctcttttctgaag	GG TCC GGA	41	203	GTG CTG ATG	gtaagtggtg
ctgctttctcttccag	GGA GGC TCC	42	165	CAG ACC AAG	gtgagcactt
gttgcatccaatgcag	CTC TTG GCC	43	168	GCT GTG AAG	gtaagcaggg
ctttttttttttctag	CAA TTT GAG	44	182	GCC ACC CG	gtaagctaag
gttctcttttcaccag	G GAC TAC	45	108	CAT GGC AG	gtaagacgct
tcctctgttttctcag	C CTG GAG	46	135	AGA TCA GG	gtaattttgg
cttctcatttgcccag	A AAT CCC	47	191	TCT GCA G	gtagcaaagc
tgcccctctgtttcag	AT GGA GCC	48	222	3′ untranslated segment	3′ flanking region
....ttttttttttncag	G........	Consensus sequences		........AG	gtaagt.....
cc ccccccc					g

33; 3b-2, between exons 33 and 36; and 3a-3, between exons 36 and 37). Type-1 repeats could function as binders and reversible inhibitors of the protease in the lysosomal pathway (28).

The five hormonogenic acceptor sites (8, 18, 49) are located at positions 5, 1291, 2554, 2568 and 2747 of the Tg monomer within exons 2, 18, 44, 45 and 48 respectively (Fig. 4), while three potential outer ring donors were identified at tyrosine residues 130, 847 and 1488 (35), which correspond to exons 4, 10 and 21 (Fig. 4). Tyrosine 5 is the most likely acceptor site for the donated iodotyrosyl from positions 130 (49).

On the other hand, TG interacts with several proteins during their intracellular transport to the surface of the cell, or with components of the apical membrane in the

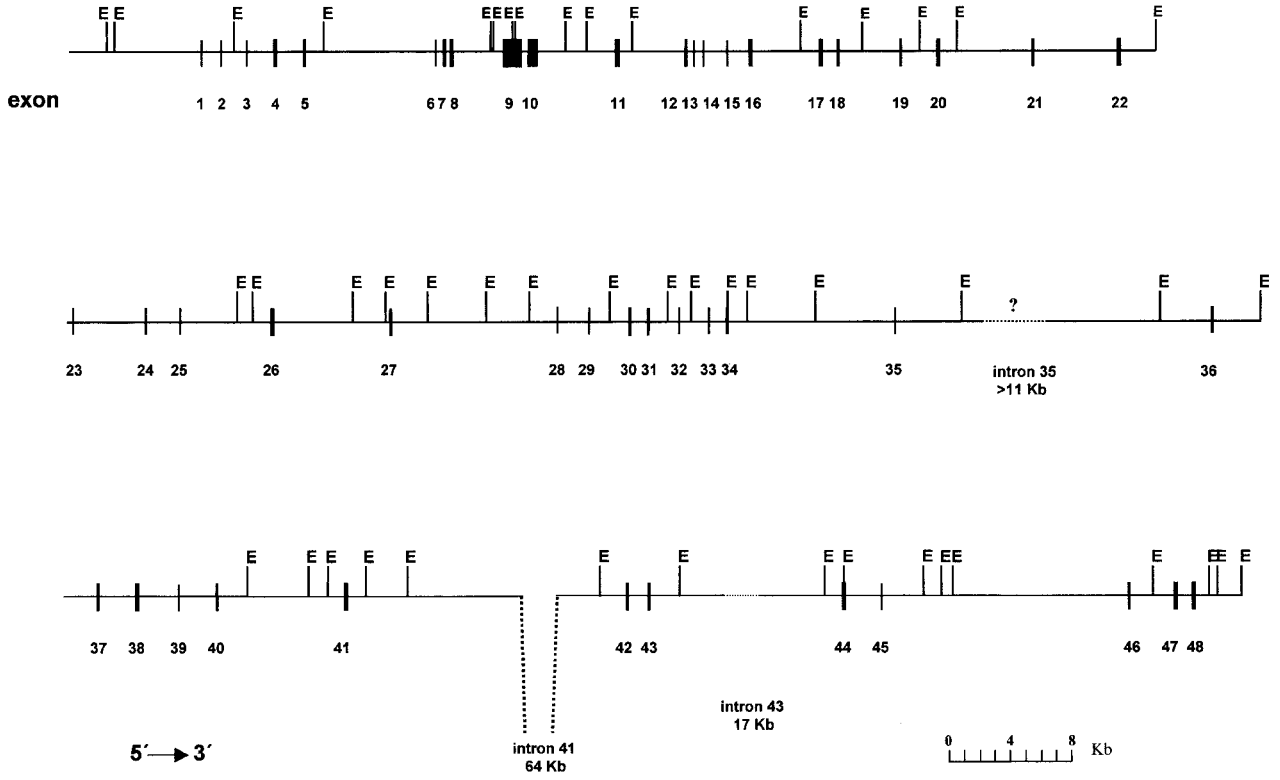

Figure 3 Schematic representation of the the physical EcoRI restriction map of the human TG gene and the distribution of these 48 exons. The exons are indicated by vertical black boxes and the introns are indicated by continuous lines. The gene is drawn to scale in the 5′-to-3′ direction. EcoRI (E) restriction sites are shown. The dotted line denotes the segment of introns not included in our phages.

exocytosis and endocytosis pathways of thyrocytes, such as the apical membrane asialoglycoprotein receptor (ASGPR) (50), megalin (51, 52) and protein disulfide isomerase (PDI) (53, 54). The ASGPR transports newly synthesized TG to the follicular lumen. It is hypothesized that the ASGPR is also indirectly involved in the endocytosis and proteolytic cleavage of highly iodinated TG by binding and sequestering immature TG. The region of TG that interacts with the receptor is unknown. It has recently been shown that the TG regulation of thyroid gene expression is mediated by the ASGPR (50). It is interesting to note that the follicular TG acts as a feedback suppressor of thyroid function, by suppressing the expression of TTF-1, TTF-2 and Pax-8 and, consequently, reducing the expression of the TG, TPO, sodium/iodide symporter (NIS) and TSHr genes. These findings support the idea that TG is not only the substrate for the biosynthesis of the thyroid hormones but also a regulator of thyroid function, playing a role in transcriptional signaling or being involved in some unknown mechanisms that remain to be determined.

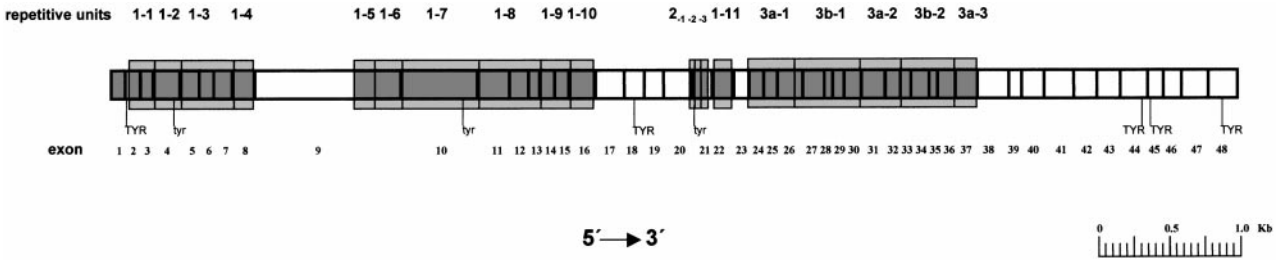

Figure 4 Exon organization and correlation with repetitive and hormonogenic domains. The exons are indicated by white boxes and the repetitive units by shaded boxes. Tyrosine residues, involved as acceptor (TYR) and donor (tyr) sites in thyroid-hormone synthesis, are shown.

Highly iodinated TG is removed from the follicular lumen by internalization via pseudopod ingestion and micropinocytosis, followed by fusion of the endosome with a lysosome and its proteolytic cleavage. It has recently been reported that megalin, a member of the low-density-lipoprotein receptor family, participates in the internalization of mature TG as a high-affinity receptor for TG (51). Megalin interacts with a heparin-binding region (SRRLKRP) in the carboxyl-terminal portion of rat TG (51). However, this domain was not detected when we searched the complete human TG protein for heparin-binding consensus sequences, using the PC GENE computer program (Intelligenetics, Inc.). Megalin plays a role in intact TG transcytosis from the apical surface to the basolateral surface of the thyrocyte (52). Subsequently, the endocytosis for proteolytic cleavage in the lysosomal pathway occurs via other mechanisms such as fluid-phase uptake or uptake by other affinity receptors.

In addition, it has been suggested that there is, at the apical surface of the thyroid cell, a quality control mechanism that prevents premature lysosomal transfer and degradation of immature TG (53). The immature molecules are internalized and recycled through the *trans*-Golgi compartments. PDI is thought to be a candidate for the receptor that mediates the internalization (54). The domain of TG responsible for the binding to the membrane is located between exons 10 (Ser^{789}) and 16 (Met^{1173}). This region contains a stretch of 385 amino acid residues that includes the cysteine-rich type I-7, I-8, I-9 and I-10 motifs and two *N*-linked glycan moieties. Cleavage of the glycan moieties reduces the binding affinity, suggesting that these complex-type oligosaccharide units are involved in the interaction between this domain and PDI.

During the completion of the present study and after our sequences had appeared in GenBank, the International Human Genome Project reported a draft of chromosome 8 that included the TG gene (http://www.ncbi.nlm.nih.gov/genome/guide/human). As expected, a comparison of the genomic sequences from both sources, performed using the BLAST version 2.1 computer program (National Center for Biotechnology Information, http://www.ncbi.nlm.nih.gov/BLAST/index.compat.html), revealed perfect sequence homology and some nucleotide differences that could be due to single nucleotide polymorphisms (SNPs).

In summary, on the basis of the previous studies and our present data, a physical map of the human TG gene was constructed and all intron–exon junctions were sequenced and correlated with the different domains of the protein. The TG gene was found to contain 48 exons, the exon sizes ranging between 63 and 1101 nucleotides.

The identification of the complete genomic organization of the human TG gene is of potential interest in terms of our undertanding of the structure–function relationship. It also opens up new perspectives in the study of the pathogenesis of hereditary thyroid diseases involving defective TG synthesis, facilitating the rapid detection of new mutations in the TG gene.

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 the Universidad de Buenos Aires (FA 124/95, TB 80/98), CONICET (7260/96, 0853/98), FONCYT (05-00000-01591/98) and the Fundación Alberto J. Roemmers (1997, 1998). The nucleotide sequence data reported in this paper have been submitted to the EMBL/GenBank/DDJ databases under the accession numbers AF169654, AF169655, AF169656, AF169657, AF169658, AF169659, AF169660, AF169661, AF169662, AF169663, AF169664 and AF255396.

References

1 Medeiros-Neto G, Targovnik HM & Vassart G. Defective thyroglobulin synthesis and secretion causing goiter and hypothyroidism. *Endocrine Reviews* 1993 **14** 165–183.
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 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 Avvedimento VE, Di Lauro R, Monticelli A, Bernardi F, Patracchini P, Calzolari E *et al.* Mapping of human thyroglobulin gene on the long arm of chromosome 8 by *in situ* hybridization. *Human Genetics* 1985 **71** 163–166.
5 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.
6 Bergé-Lefranc J-L, Cartouzou G, Mattei M-G, Passage E, Malezet-Desmoulins C & Lissitzky S. Localization of the thyroglobulin gene by *in situ* hybridization to human chromosomes. *Human Genetics* 1985 **69** 28–31.
7 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.
8 Moya CM, Mendive FM, Rivolta CM, Vassart G & Targovnik HM. Genomic organization of the 5′ region of the human thyroglobulin gene. *European Journal of Endocrinology* 2000 **143** 789–798.
9 Mendive FM, Rivolta CM, Vassart G & Targovnik HM. Genomic organization of the 3′ region of the human thyroglobulin gene. *Thyroid* 1999 **9** 903–912.
10 Meijerink PHS, Yanakiev P, Zorn I, Grierson AJ, Bikker H, Dye D *et al.* The gene for the human Src-like adaptor protein (hSLAP) is located within the 64-Kb intron of the thyroglobulin gene. *European Journal of Biochemistry* 1998 **254** 297–303.
11 Vassart G & Dumont JE. The thyrotropin receptor and the regulation of thyrocyte function and growth. *Endocrine Reviews* 1992 **13** 596–611.

12 Tonacchera M, Van Sande J, Parma J, Duprez L, Cetani F, Costagliola S *et al.* TSH receptor and disease. *Clinical Endocrinology* 1996 **44** 621–633.
13 Damante G & Di Lauro R. Thyroid-specific gene expression. *Biochimica et Biophysica Acta* 1994 **1218** 255–266.
14 Macchia PE, Mattei MG, Lapi P, Fenzi G & Di Lauro R. Cloning, chromosomal localization and identification of polymorphisms in the human thyroid transcription factor 2 gene (TTTF2). *Biochimie* 1999 **81** 433–440.
15 Sinclair AJ, Lonigro R, Civitareale D, Ghibelli L & Di Lauro R. The tissue-specific expression of the thyroglobulin gene requires interaction between thyroid-specific and ubiquitous factors. *European Journal of Biochemistry* 1990 **193** 311–318.
16 Zannini M, Francis-Lang H, Plachov D & Di Lauro R. Pax-8, a paired domain-containing protein, binds to a sequence overlapping the recognition site of a homeodomain and activates transcription from two thyroid-specific promoters. *Molecular and Cellular Biology* 1992 **12** 4230–4241.
17 Berg V, Vassart G & Christophe D. A zinc-dependent DNA-binding activity co-operates with cAMP-responsive-element-binding protein to active the human thyroglobulin enhancer. *Biochemical Journal* 1997 **323** 349–357.
18 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.
19 van de Graaf SAR, Pauws E, de Vjilder 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.
20 van de Graaf S. New insights in the human thyroglobulin structure. PhD Thesis. University of Amsterdam, 2000.
21 Mendive FM, Rossetti LC, Vassart G & Targovnik HM. Identification of a new thyroglobulin variant: a guanine-to-arginine transition resulting in the substitution of arginine 2510 by glutamine. *Thyroid* 1997 **7** 587–591.
22 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.
23 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.
24 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.
25 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.
26 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.
27 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.
28 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.
29 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.
30 Kim PS & Arvan P. Folding and assembly of newly synthesized thyroglobulin occurs in a pre-golgi compartment. *Journal of Biological Chemistry* 1991 **266** 12412–12418.
31 Desphande V & Venkatesh SG. Thyroglobulin, the prothyroid hormone: chemistry, synthesis and degradation. *Biochimica et Biophysica Acta* 1999 **1430** 157–178.
32 Kim PS & Arvan P. Calnexin and BiP act as sequential molecular chaperones during thyroglobulin folding in the endoplasmic reticulum. *Journal of Cell Biology* 1995 **128** 29–38.
33 Medeiros-Neto G, Kim PS, Yoo SE, Vono J, Targovnik HM, Camargo R *et al.* Congenital hypothyroid goiter with deficient thyroglobulin. Identification of an endoplasmic reticulum storage disease with induction of molecular chaperones. *Journal of Clinical Investigation* 1996 **98** 2838–2844.
34 Hammond C & Helenius A. Quality control in the secretory pathway. *Current Opinon in Cell Biology* 1995 **7** 523–529.
35 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.
36 Taurog A. Hormone synthesis. In *Werner's & Ingbar's The Thyroid*, pp 47–84. Eds LE Braverman & RD Utiger. Philadelphia: J.B. Lippincott Company, 1996.
37 De Deken X, Wang D, Many MC, Costagliola S, Libert F, Vassart G *et al.* Cloning of two human thyroid cDNAs encoding new members of the NADPH oxidase family. *Journal of Biological Chemistry* 2000 **275** 23227–23233.
38 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.
39 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.
40 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.
41 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.
42 Chomczynski P & Sacchi N. Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. *Analytical Biochemistry* 1987 **162** 156–159.
43 Southern EM. Detection of specific sequences among DNA fragments separated by gel electrophoresis. *Journal of Molecular Biology* 1975 **98** 503–517.
44 Reed KC & Mann DA. Rapid transfer of DNA from agarose gels to nylon membranes. *Nucleic Acids Research* 1985 **13** 7207–7221.
45 Sambrook J, Fritsch EF & Maniatis T. *Molecular Cloning – a Laboratory Manual*. Cold Spring Harbor, New York: Cold Spring Harbor Laboratory Press, 1989.
46 Targovnik HM, Rivolta CM, Mendive FM, Moya CM & Medeiros-Neto G. Congenital goiter with hypothyroidism due to a 5′ splice site mutation in the thyroglobulin gene. *Thyroid* 2001 **11** 685–690.
47 van Ommen GJB, Arnberg AC, Baas F, Brocas H, Sterk A, Tegelaers WHH *et al.* The human thyroglobulin gene contains two 15–17 kb introns nears its 3′-end. *Nucleic Acids Research* 1983 **11** 2273–2285.
48 Shapiro BM & Senaphaty P. RNA splice junctions of different classes of eukaryotes: sequence statistics and functional implications in gene expression. *Nucleic Acids Research* 1987 **15** 7155–7174.
49 Dunn AD, Corsi CM, Myers HE & Dunn JT. Tyrosine 130 is an important outer ring donor for thyroxine formation in thyroglobulin. *Journal of Biological Chemistry* 1998 **273** 25223–25229.
50 Ulianich L, Suzuki K, Mori A, Nakazato M, Pietrarelli M, Goldsmith P *et al.* Follicular thyroglobulin (TG) suppression

of thyroid-restricted genes involves the apical membrane asialoglycoprotein receptor and TG phosphorylation. *Journal of Biological Chemistry* 1999 **274** 25099–25107.

51 Marino M, Friedlander JA, McCluskey RT & Andrews D. Identification of a heparin-binding region of rat thyroglobulin involved in megalin binding. *Journal of Biological Chemistry* 1999 **274** 30377–30386.

52 Marino M, Zheng G, Chiovato L, Pinchera A, Brown D, Andrews D *et al.* Role of megalin (gp330) in transcytosis of thyroglobulin by thyroid cells. A novel function in the control of thyroid hormone release. *Journal of Biological Chemistry* 2000 **275** 7125–7137.

53 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.

54 Metzghrani A, Courageot J, Mani JC, Pugniere M, Bastiani P & Miquelis R. Protein-disulfide isomerase (PDI) in FRTL5 cells. PH-dependent thyroglobulin/PDI interactions determine a novel PDI function in the post-endoplasmic reticulum of thyrocytes. *Journal of Biological Chemistry* 2000 **275** 1920–1929.

Received 21 March 2001
Accepted 23 May 2001