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

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, 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, 4th 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 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 (T3) and thyroxine (T4), 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 T3 and T4 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 expression 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
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 peptidase 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 apparatus, 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 as quality control (34). 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 T₃ and T₄. Five hormonal 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 via free access.
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 NaH₂PO₄ (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-³²P]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 ammonium 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′-gagtcgccctgtggt-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 MgCl₂, 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 annealing 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 (fmol: 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 3.4×10⁶ 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
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

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.

<table>
<thead>
<tr>
<th>Exon no.</th>
<th>Position of 5' end</th>
<th>Nucleotide sequence (5'→3')</th>
<th>Position of 5' end</th>
<th>Nucleotide sequence (5'→3')</th>
</tr>
</thead>
<tbody>
<tr>
<td>24</td>
<td>4817</td>
<td>AATTGACACAGAGGAGAGGCC</td>
<td>4932</td>
<td>CTGCTCAGAGGTGTGAGGCC</td>
</tr>
<tr>
<td>25</td>
<td>4933</td>
<td>AAACGAGATGCTGTGGGG</td>
<td>5027</td>
<td>ACAGCTGAGAAACTCTGGACATTGCC</td>
</tr>
<tr>
<td>26</td>
<td>5049</td>
<td>ATCCACCAACACTCTGAGAACGC</td>
<td>5146</td>
<td>TGCCCTCTGAGCTGAGACAC</td>
</tr>
<tr>
<td>27</td>
<td>5239</td>
<td>AATCATCTGCTGGGTGTGCTAGGCC</td>
<td>5397</td>
<td>GATGAAACTCTCTGCTCCAAAGAGCC</td>
</tr>
<tr>
<td>28</td>
<td>−51</td>
<td>taccaccgatgacagcaactcaagcc</td>
<td>5463</td>
<td>CCAAGAGAAACTCTGCTGAATGTTCC</td>
</tr>
<tr>
<td>29</td>
<td>5468</td>
<td>AATTCTGACATGGGTCTCTGAGCC</td>
<td>5547</td>
<td>TGCCCTCTGAGATGCTGGG</td>
</tr>
<tr>
<td>30</td>
<td>5552</td>
<td>TGACACACAGACACTTTTCTCC</td>
<td>5686</td>
<td>aAGAAAGAGCACCAGATGAG</td>
</tr>
<tr>
<td>31</td>
<td>5640</td>
<td>GAAGCATGTTTTTCAAGCC</td>
<td>5789</td>
<td>TCTACACACTGCCTTCCTCTCT</td>
</tr>
<tr>
<td>32</td>
<td>−996</td>
<td>ggtgctctctagctgatcactcgcttg</td>
<td>5822</td>
<td>TGAAGAGCTTCTTCATAGG</td>
</tr>
<tr>
<td>33</td>
<td>5981</td>
<td>TGAAGAGCTTCTTCATAGG</td>
<td>5934</td>
<td>TGAAGAGCTTCTTCATAGG</td>
</tr>
<tr>
<td>34</td>
<td>6056</td>
<td>GAGGAGAGGTGACATGTCTCAC</td>
<td>6199</td>
<td>TGGGCTTCTGAGCTGAGACCC</td>
</tr>
<tr>
<td>35</td>
<td>−65</td>
<td>gocctctctgagctgatgatacc</td>
<td>6262</td>
<td>TCTCTTCTGAGGGAGGCC</td>
</tr>
</tbody>
</table>

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 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). 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 37).
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.
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...
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 3** Schematic representation of 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.

**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 (SRRLKR) in the carboxy-terminal portion of rat TG (51). However, this domain was not detected when we examined 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 (Ser989) and 16 (Met1173). 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 understanding 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

Thyroglobulin gene structure

495

EUROPEAN JOURNAL OF ENDOCRINOLOGY (2001) 145

16 Zannini M, Francis-Lang H, Plachov D & Di Lauro R. Pax-8, a heterogeneous transcription pattern of thyroglobulin mRNA in
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.
31 Desphande V & Venkatesh SG. Thyroglobulin, the prothyroid hormone: chemistry, synthesis and degradation. Biochimica et Biophysica Acta 1999 1430 157–178.
43 Reed KC & Mann DA. Rapid transfer of DNA from agarose gels to nylon membranes. Nucleic Acids Research 1985 13 7207–7221.
45 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.

www.eje.org

Downloaded from Bioscientifica.com at 02/07/2019 08:11:53PM via free access


Received 21 March 2001

Accepted 23 May 2001