STRUCTURE UPDATE

Up to date with human thyroglobulin

S A R van de Graaf, C Ris-Stalpers, E Pauws, F M Mendive[1], H M Targovnik[1] and J J M de Vijlder

Academic Medical Center, University of Amsterdam, Emma Children's Hospital AMC, Laboratory of Pediatric Endocrinology, Amsterdam, The Netherlands

[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

(Requests for offprints should be addressed to C Ris-Stalpers, Academic Medical Center H2–255, Laboratory of Pediatric Endocrinology, PO Box 22700, 1100 DE Amsterdam, The Netherlands; Email: C.Ris@amc.uva.nl)

Abstract

The coding region of the human thyroglobulin (TG) mRNA has been resequenced, and comparison with the TG sequence originally published in 1987 showed many variations. All of the variations were validated in 20–40 other alleles, and this resulted in the revision of 41 nucleotide positions. This review presents the revised wild-type human TG sequence, including all known exon/exon boundaries and additional data on the TG mRNA population, concerning alternative splicing and variability of the polyadenylation cleavage site. The amino acid sequence derived shows one additional, 12 changed, and 10 polymorphic residues. Protein characteristics, such as acceptor and donor tyrosine residues, *N*-glycosylation sites, cysteine-rich repeats, the proposed receptor domain, and antigenic epitopes, are included, and their relationship to the revised sequence is discussed. Furthermore, all reported TG mutations causing dyshormonogenesis in humans and animals are designated in the nucleotide and amino acid sequences. This up-to-date profile of the human TG molecule presents the features of importance for its complex role in thyroid hormonogenesis, and is the basis for future studies on the structure–function relationship.

Journal of Endocrinology (2001) **170,** 307–321

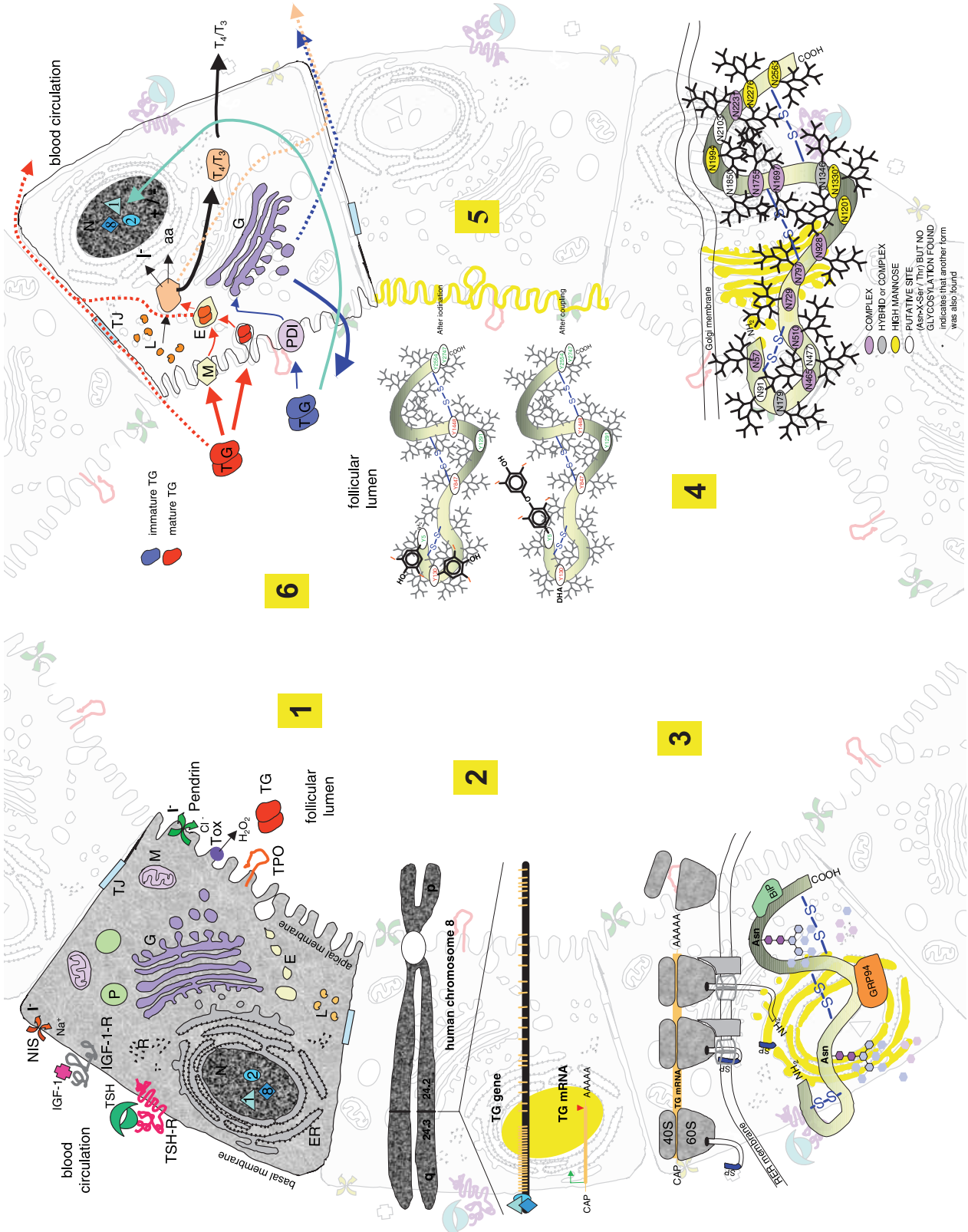
Introduction

Thyroid gland

The thyroid gland is responsible for thyroid hormone production and mainly produces the prohormone thyroxine (T_4), which contains four iodide molecules (Dunn 1999). The gland consists of follicular structures – epithelial cells that border a lumen with their apical membranes and are in contact with the blood circulation through their basal membranes (Fig 1). The most highly expressed protein in the thyroid gland is thyroglobulin (TG), which functions as a scaffold protein for thyroid hormonogenesis and as storage protein for thyroid hormones and iodide. Other thyroidal proteins involved in thyroid hormone synthesis are the receptor for thyrotropin (TSH-R), a seven-transmembrane receptor located in the basal membrane passing the signal that thyroid hormone is needed. Upon thyrotropin stimulation, several processes, including TG synthesis, are upregulated in the thyrocyte cells; all of these processes favor thyroid hormonogenesis. The sodium iodide symporter (NIS), located in the basal membrane, and pendrin, located in the apical membrane, are responsible for the iodide supply and transport within the gland. Iodination of specific tyrosine residues in TG and coupling of these iodinated residues to form T_4 depend on the thyroid peroxidase (TPO) anchored in the apical membrane and on one or more thyroid oxidases (Tox) that provide the H_2O_2 (Dupuy *et al.* 1999, De Deken *et al.* 2000) (Fig. 1, cell 1).

Thyroglobulin expression, synthesis and maturation

Initial transcription of the TG gene (>300 kb) is regulated by thyroid-specific transcription factors TTF-1, TTF-2 and Pax8 (Fig. 1, cell 2). The full-length 8·7 kb mRNA transcript is the most highly expressed mRNA transcript in normal thyrocytes, with an expression level of 2·7% (Pauws *et al.* 2000, 2001). After translation of the mRNA, the post-translational route of TG starts with the signal peptide directing the uptake in the endoplasmic reticulum (ER), where the first mannose and glucose residues are added and the TG protein is folded (Fig. 1, cell 3). At least seven molecular chaperones are known to guide this process to ensure TG quality (Kim & Arvan 1993, 1998).



When TG is misfolded at this stage, stable complexes are formed, and TG is retained in the ER and broken down. TG that 'passes' the ER quality control is directed to the Golgi apparatus, where glycosylation proceeds (Fig. 1, cell 4). At this point, TG molecules have a molecular weight of 300 000, contain 10% carbohydrate structures, and are routed further through the cell to the follicular lumen while homodimers with a molecular weight of 660 000 are formed.

In the follicular lumen, iodination takes place on tyrosine residues in TG, resulting in monoiodotyrosines and di-iodotyrosines (MITs and DITs, respectively). Subsequent coupling of two DIT residues (or DIT and MIT) takes place under oxidative conditions, resulting in iodinated TG molecules carrying T_4 (or tri-iodothronine) on specific hormonogenic sites (Lamas *et al.* 1989, de Vijlder & Den Hartog 1998) (Fig. 1, cell 5). The efficiency of iodination correlates with completion of carbohydrate structures (Bastiani *et al.* 1995), and poorly iodinated TG molecules (immature TG) can be internalized from the lumen and recycled through the Golgi apparatus (Fig. 1, cell 6). A receptor-binding domain involved in the binding of immature TG molecules to a membrane receptor has been identified (Mezgrhani *et al.* 1997). Protein disulphide isomerase (PDI) is thought to be a candidate receptor (Mezghrani *et al.* 2000).

The TG–T_4 complexes (mature TG) are stored in a concentrated compact protein form in the follicular lumen (Berndorfer *et al.* 1996). Hormonal secretion to the venous flow requires limited extracellular proteolysis of the stored TG–T_4 complexes, followed by vesicular internalization; subsequent fusion with lysosomes results in breakdown of the TG–T_4 complexes, thereby liberating thyroid hormones from the protein backbone (Brix *et al.* 1996). Vesicular internalization can take place both by non-selective fluid phase uptake and by receptor-mediated endocytosis. Two receptors, located in the apical membrane, have been proposed to participate in the endocytosis: megalin (Zheng *et al.* 1998), and the thyroid asialoglycoprotein receptor (Ulianich *et al.* 1999).

Thyroglobulin DNA sequence

The TG gene, mapped to human chromosome 8q24·2–8q24·3 (Baas *et al.* 1985, Bergé-Lefranc *et al.* 1985, Rabin *et al.* 1985), covers at least 300 kb of genomic DNA and contains 8·5 kb coding sequence divided over 48 exons (Mendive *et al.* 1999), separated by introns varying in size up to 64 kb (Baas *et al.* 1986). In 1987, the primary structure of human TG was deduced from the sequence of its 8448-base mRNA and corresponding coding DNA (cDNA) sequence of 8304 bp (Malthièry & Lissitzky 1987). This sequence has been used since to screen for mutations in patients with hypothyroidism due to putative TG-synthesis defects (Ieri *et al.* 1991, Targovnik *et al.* 1993, 1995). Recently we sequenced the TG cDNA of seven hypothyroid patients in which dyshormonogenesis was caused by a putative TG-synthesis defect (van de Graaf *et al.* 1999*a*,*b*). Multiple sequence variations were found relative to the original sequence; these were further investigated using TG mRNA from normal thyroid tissue (van de Graaf *et al.* 1997, 1999*a*, and this report).

Reverse transcriptase/PCR (RT-PCR) also demonstrated the existence of various wild-type splice variants (Bertaux *et al.* 1991, 1995, Targovnik *et al.* 1992, Mason *et al.* 1995, van de Graaf *et al.* 1999*a*, Hishinuma *et al.* 1999), which upon translation attributes to a heterogeneous population of TG molecules within one individual.

Materials and Methods

RNA isolation and complementary DNA preparation

Total RNA was isolated from up to 20 thyroid gland tissue samples (obtained with informed consent) by using TRIzol reagent (Life Technologies BV, Breda, The Netherlands). cDNA was synthesized using random hexamers and reverse transcriptase according to standard procedures.

DNA amplification

PCR amplification was performed using 100 ng cDNA as the template in a total reaction volume of 25 µl. For nucleotide sequencing, fragments of 500 bp (with 20–70 bp overlap) were amplified with 2·5 U AmpliTaq DNA polymerase (Perkin Elmer, Norwalk, CT, USA) using the following protocol: 2 min at 94 °C; 35 cycles of 15 s at 94 °C, 1 min at 60 °C, 1 min at 72 °C; 10 min at 72 °C.

Oligonucleotides (van de Graaf *et al.* 1999*a*) (synthesized using an Expedite nucleic acid synthesis system; Millipore, Bedford, MA, USA) were specific for human TG (for the sequence, see GenBank accession no. U93033) and were coupled to M13 tags.

Figure 1 Thyroglobulin synthesis and maturation processes within the thyroid follicle. A thyroid follicle is schematically depicted with a general thyrocyte (cell 1) in which different organelles (E, endosome; ER, endoplasmic reticulum; G, Golgi apparatus; L, lysosome; M, mitochondrium; N, nucleus; P, peroxisome; R, bound and free ribosome) and thyroid-specific proteins (IGF-I, insulin-like growth factor-I; IGF-I-R, receptor for IGF-I; NIS, sodium iodide symporter; TG, thyroglobulin; TJ, tight junction; Tox, thyroid oxidases; TPO, thyroid peroxidase; TSH, thyrotropin; TSH-R, receptor for TSH; 1, TTF-1; 2, TTF-2; 8, Pax8) are indicated. In cells 2 to 5, the relevant processes in the different organelles are presented (cell 2, nucleus; cell 3, endoplasmic reticulum; cell 4, Golgi apparatus; cell 5, apical membrane and follicular lumen). Cell 6 illustrates the pathways for TG internalization and the secretion of thyroid hormones and TG.

Nucleotide sequencing

Both the sense strand and the antisense strand were sequenced using the M13 tags linked to the PCR fragments. Reactions were performed using the Big Dye primer cycle sequencing kit (M13 forward and M13 reverse) and the Big Dyedeoxyterminator cycle sequencing kit (in combination with PCR primers) depending on the GC content of the fragment; both kits were from Applied Biosystems/Perkin Elmer, Foster City, CA, USA. After electrophoresis on a sequencing gel, the samples were analysed on an ABI Prism 377 DNA sequencer and aligned with the TG cDNA sequence by using AutoAssembler software (Applied Biosystems/Perkin Elmer).

Results and Discussion

In order to update the coding sequence of human TG, we partially repeated and extended our earlier studies. This resulted in the direct sequencing of total cDNA for human TG in a population of 20–40 alleles (8–24 normal alleles and 12–16 from patients with a putative TG-synthesis defect).

We showed that the open reading frame of human TG consists of 8307 bp because of an extra nucleotide triplet (CAG) after position 2952. The open reading frame codes for 2768 amino acid residues, the signal peptide included. The 23 revised nucleotide positions resulted in the change of 12 amino acid residues. One of these, at amino acid position 1024, reduces the original number of tyrosine residues in the TG monomer from 67 to 66. Furthermore, if the study of a Japanese population (Hishinuma *et al.* 1999) is also taken into account, 15 nucleotide positions can be considered polymorphic, 10 of them resulting in amino acid polymorphisms.

All of the changes detected in patients with a putative TG-synthesis defect were either wild type or polymorphic, except for nucleotide G229A, which was homozygously present only in three patients at an allele frequency of 15%.

After the initial screening of 6 alleles (van de Graaf *et al.* 1997), the A-to-G change at position 7589 seemed to be a revision; however, extension of the screening to 66 alleles (Mendive *et al.* 1997) showed that it is a polymorphic nucleotide position.

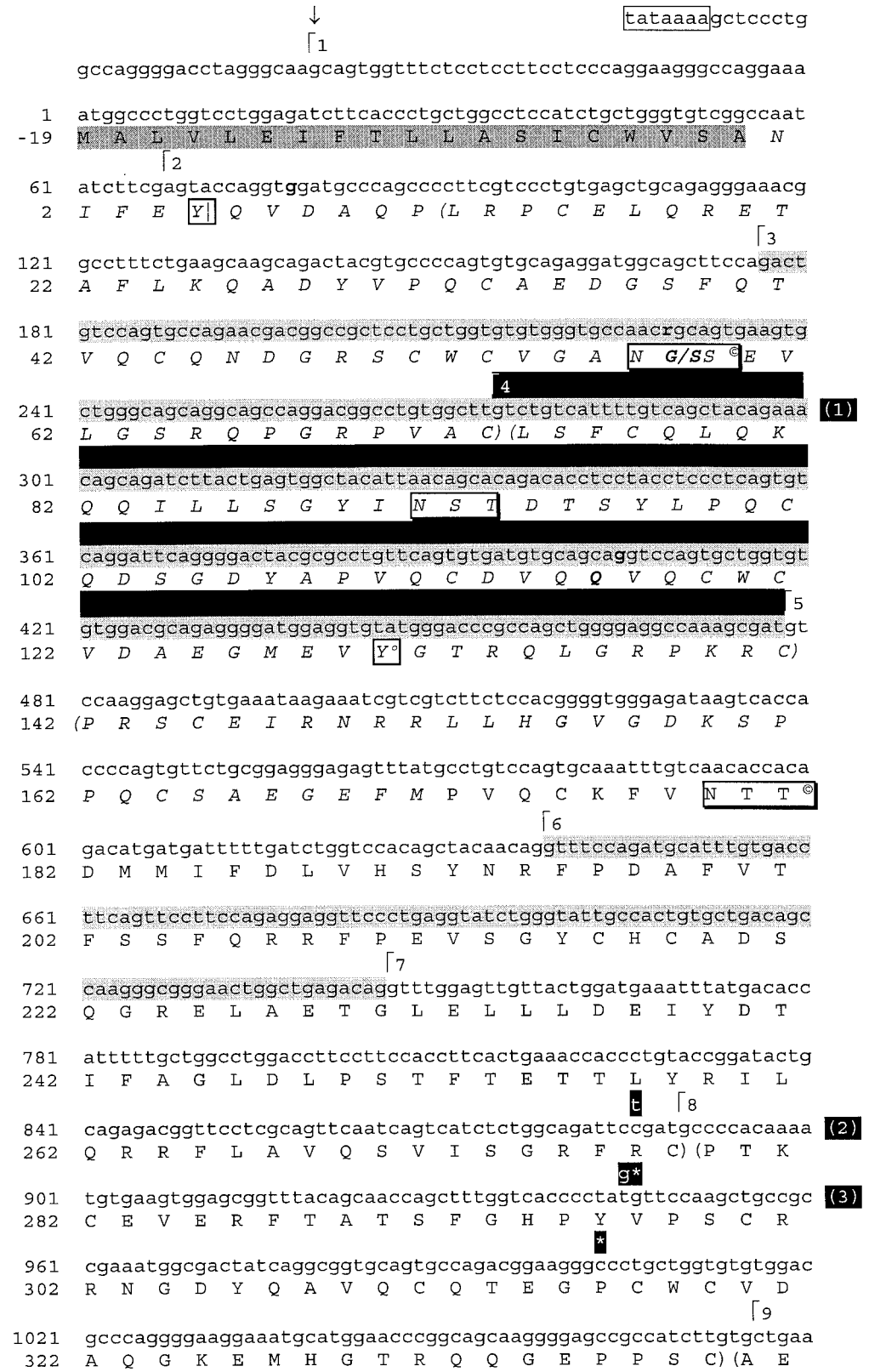
Figure 2 shows the revised human TG coding sequence of 8307 nucleotides according to the international ubiquity code. The 5′ untranslated region up to the TATA box (Parma *et al.* 1987) and the 3′ untranslated region up to the polyadenylation cleavage site that was positioned variably within one individual (Pauws *et al.* 2001) have been added to the cDNA sequence. The 41 nucleotide variations compared with the original sequence (insertions, revisions and polymorphisms) are indicated. The genomic organization of 48 exons, recently further unraveled by Targovnik and co-workers (Parma *et al.* 1987, Mendive *et al.* 1999, Rivolta *et al.* 1999, Moya *et al.* 2000) is presented in Fig. 2. The 11 alternative splice variants from normal alleles that were characterized are also indicated (Bertaux *et al.* 1991, 1995, Targovnik *et al.* 1992, Hishinuma *et al.* 1999, Mason *et al.* 1995, van de Graaf *et al.* 1999*a*).

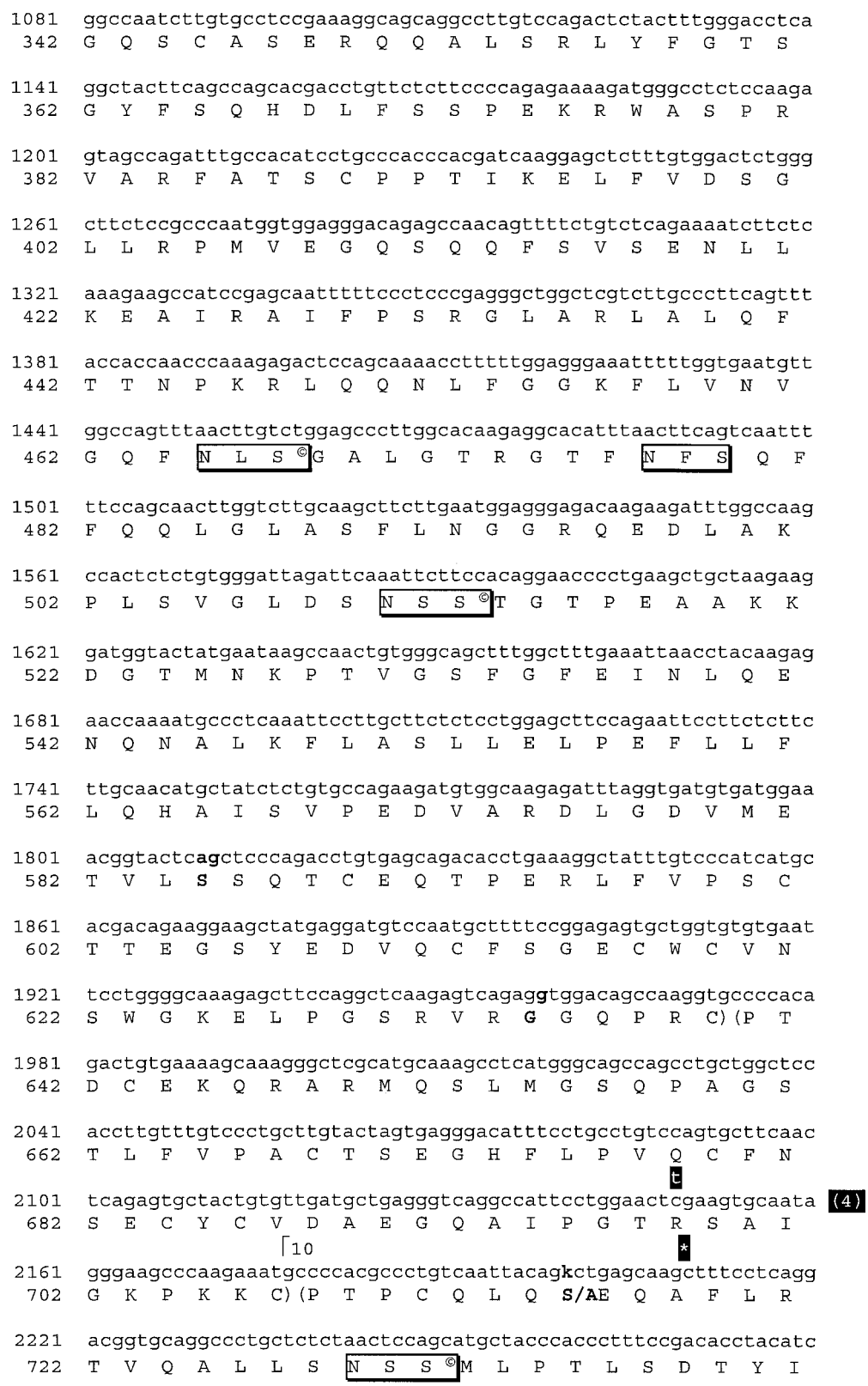
The encoded amino acid sequence (a 19-amino-acid signal peptide plus 2749 amino acids) is presented in one-letter code underneath the cDNA, and the 23 amino acid variations are indicated. To complete the complex profile of the TG protein, the human acceptor and the donor tyrosine residues (Lamas *et al.* 1989, Dunn 1999), *N*-glycosylation sites (Yang *et al.* 1996), the cysteine-residue-rich repeated domains (Malthièry & Lissitzky 1987, Gentile & Salvatore 1993, Molina *et al.* 1996), the acetylcholinesterase homologous domain (Swillens *et al.* 1986), the proposed binding domain to the *N*-acetyl-glucosamine receptor (Mezgrhani *et al.* 1997), the recently identified thioredoxin boxes (Klein *et al.* 2000), and the most prominent antigenic epitopes (Henry *et al.* 1992, Mallet *et al.* 1992, Kong *et al.* 1995, Saboori *et al.* 1995, Erregragui *et al.* 1997) have been included.

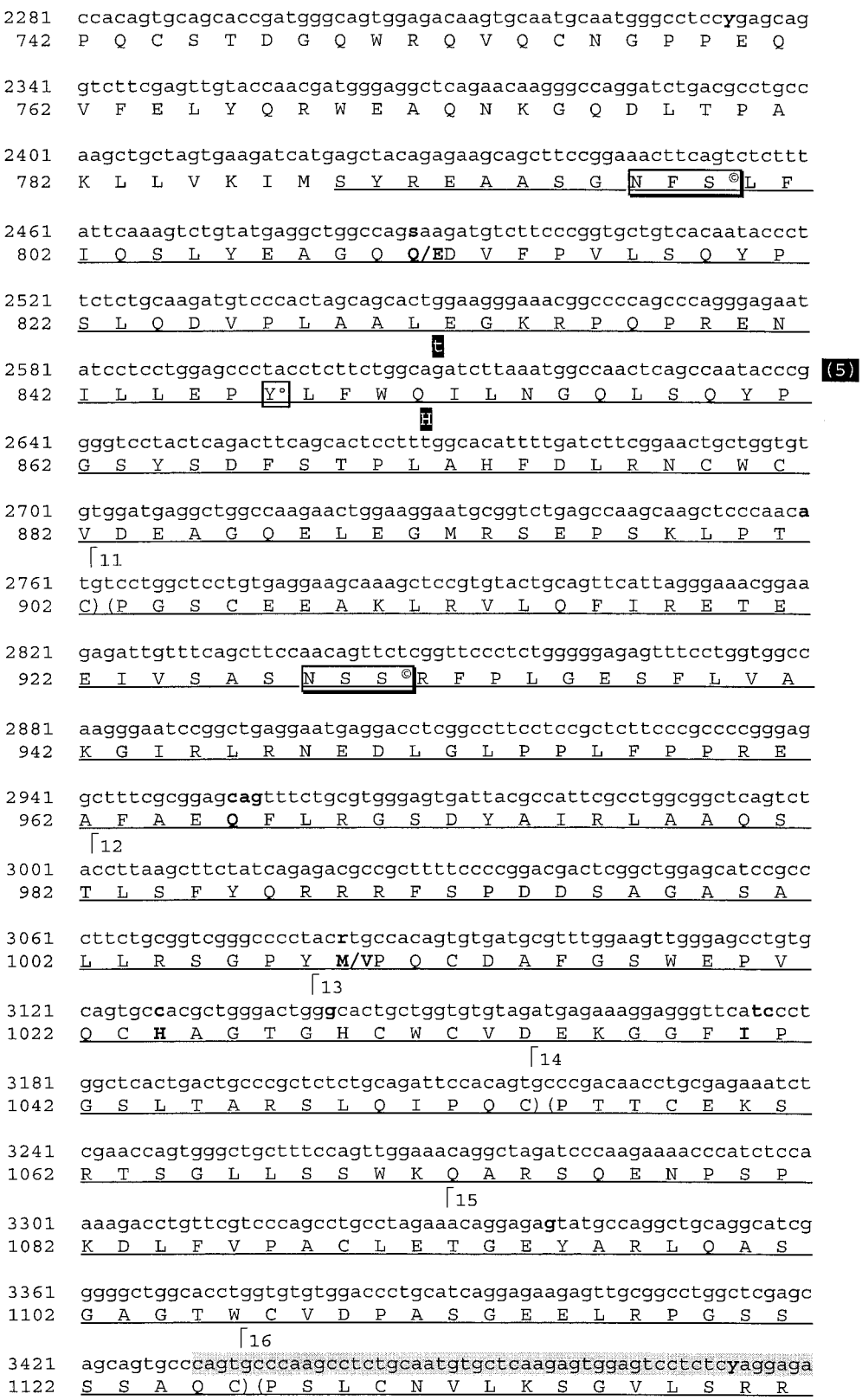
Previously identified mutations and deletions in human thyroid pathology (Ieri *et al.* 1991, Targovnik *et al.* 1993, 1995, Hishinuma *et al.* 1999, van de Graaf *et al.* 1999*b*) and homologous positions in animals with hereditary thyroid disorders linked to a TG defect (Dutch goats (Veenboer & de Vijlder 1993), Afrikander cattle (Ricketts *et al.* 1987), *rdw* rats (Hishinuma *et al.* 2000) and *cog/cog* mice (Kim *et al.* 1998)) are indicated in Fig. 2. Other variations in TG mRNA identified in patients with Pendred's syndrome (Mason *et al.* 1995), a variant type of adenomatous goiter (Hishinuma *et al.* 1999), and non-endemic simple goiter (Corral *et al.* 1993) are also indicated (Fig. 2). There are no data on a putative functional role for the TG variations in these types of diseases.

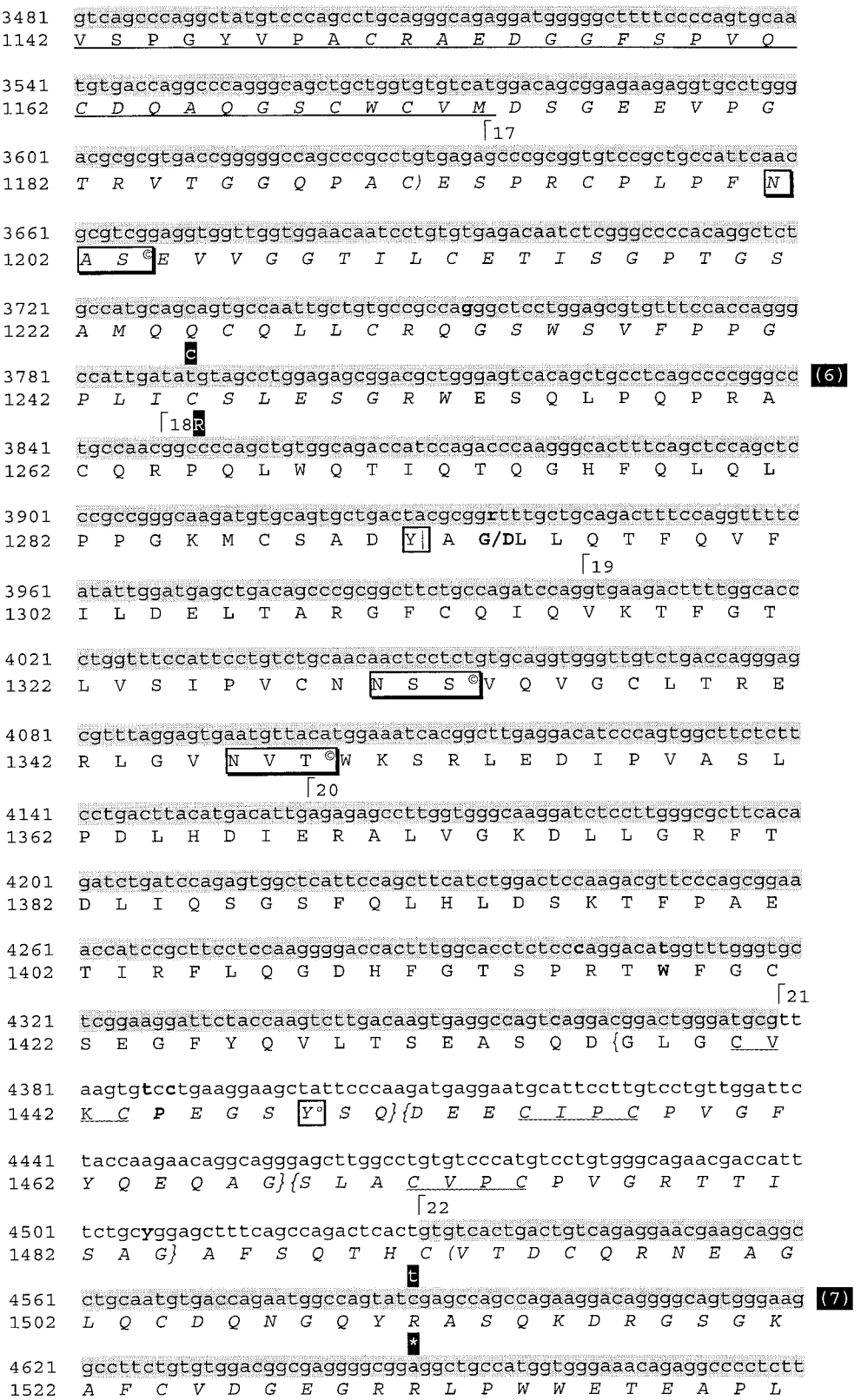
The 11 type-1 cysteine-rich repeats, first described in the bovine TG sequence (Mercken *et al.* 1985), are very well characterized and are even referred to as 'thyroglobulin type-1 domains' when identified in other proteins of different origin and function. It has been shown that these repeats are protein modules, i.e. self-folding structural units, and can act as proteinase inhibitors in other proteins (Lenarcic & Bevec 1998). It is hypothesized that the type-1 repeats in TG could act as pH-dependent binders and reversible inhibitors of the proteases implicated in TG degradation and T_4 release (Molina *et al.* 1996), but up till now no definite function for this type of repeat in TG has been established. No mutations that could shed more light on the role of the cysteine type-1 repeats in TG have been reported in any of the cysteine residues located in the type-1 domains.

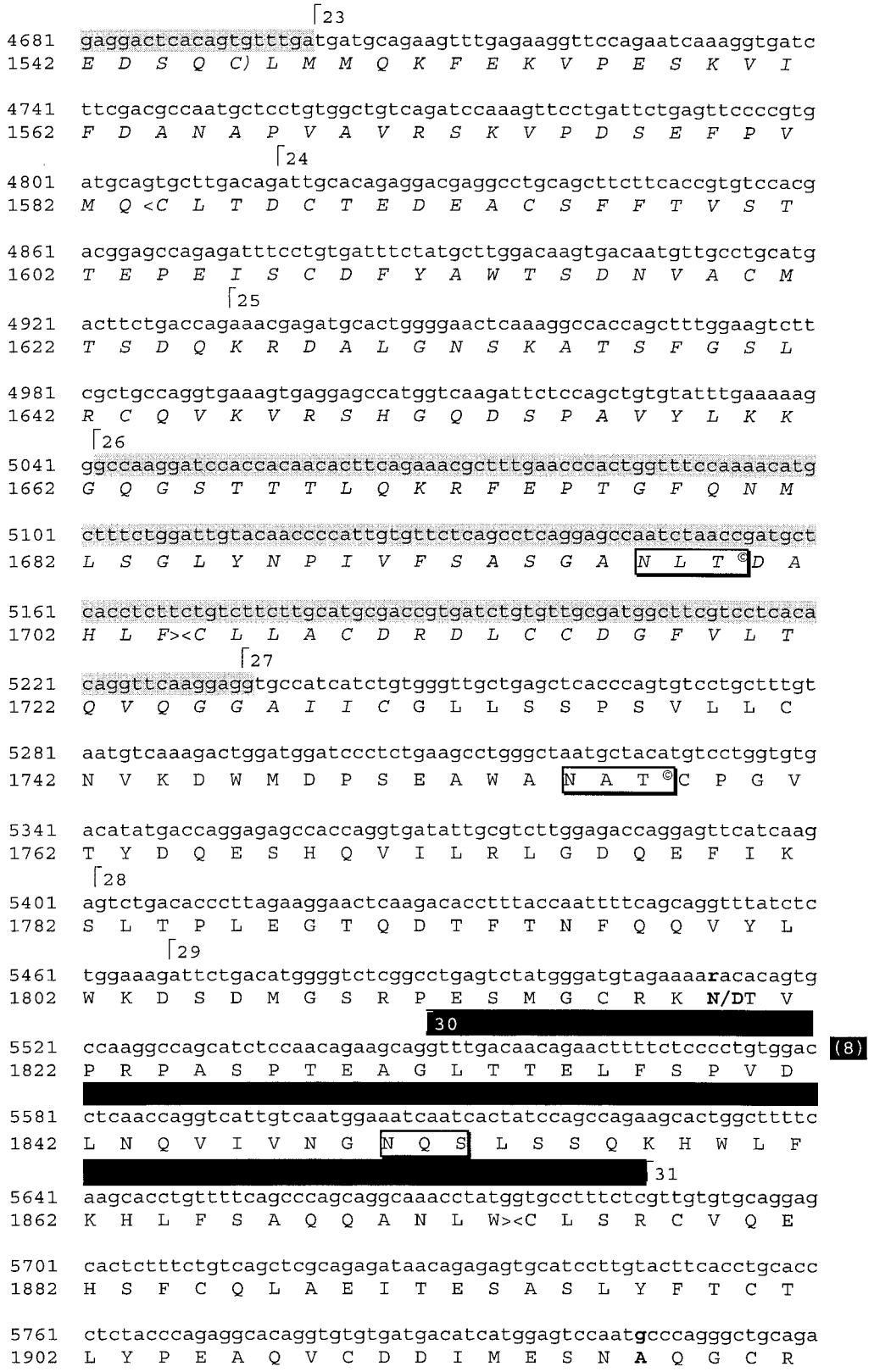
The putative function of the acetylcholinesterase homologous domain in TG is not yet clear, but as acetylcholinesterase interacts with cell membranes in the nervous system a similar role for the homologous domain at

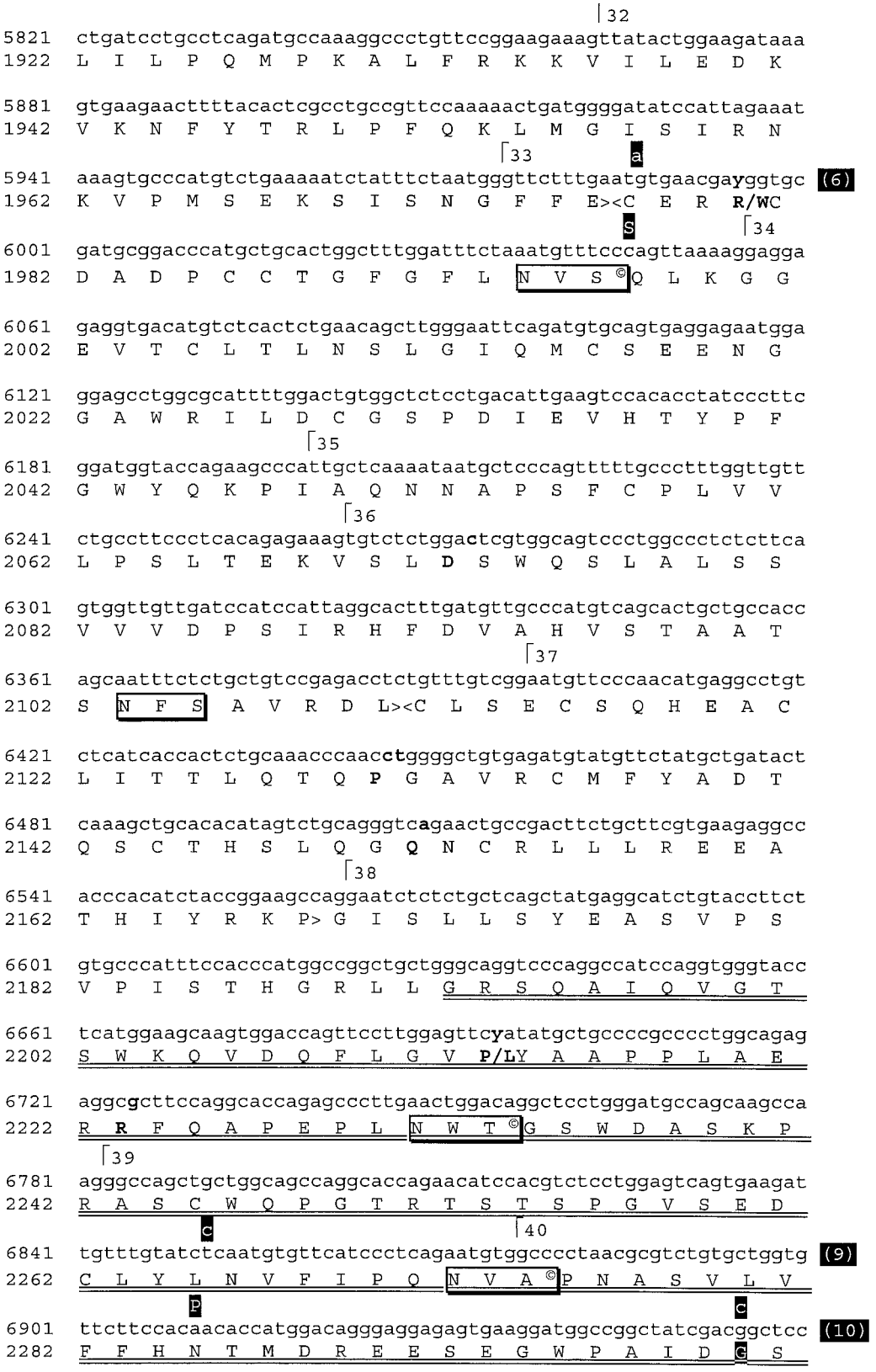


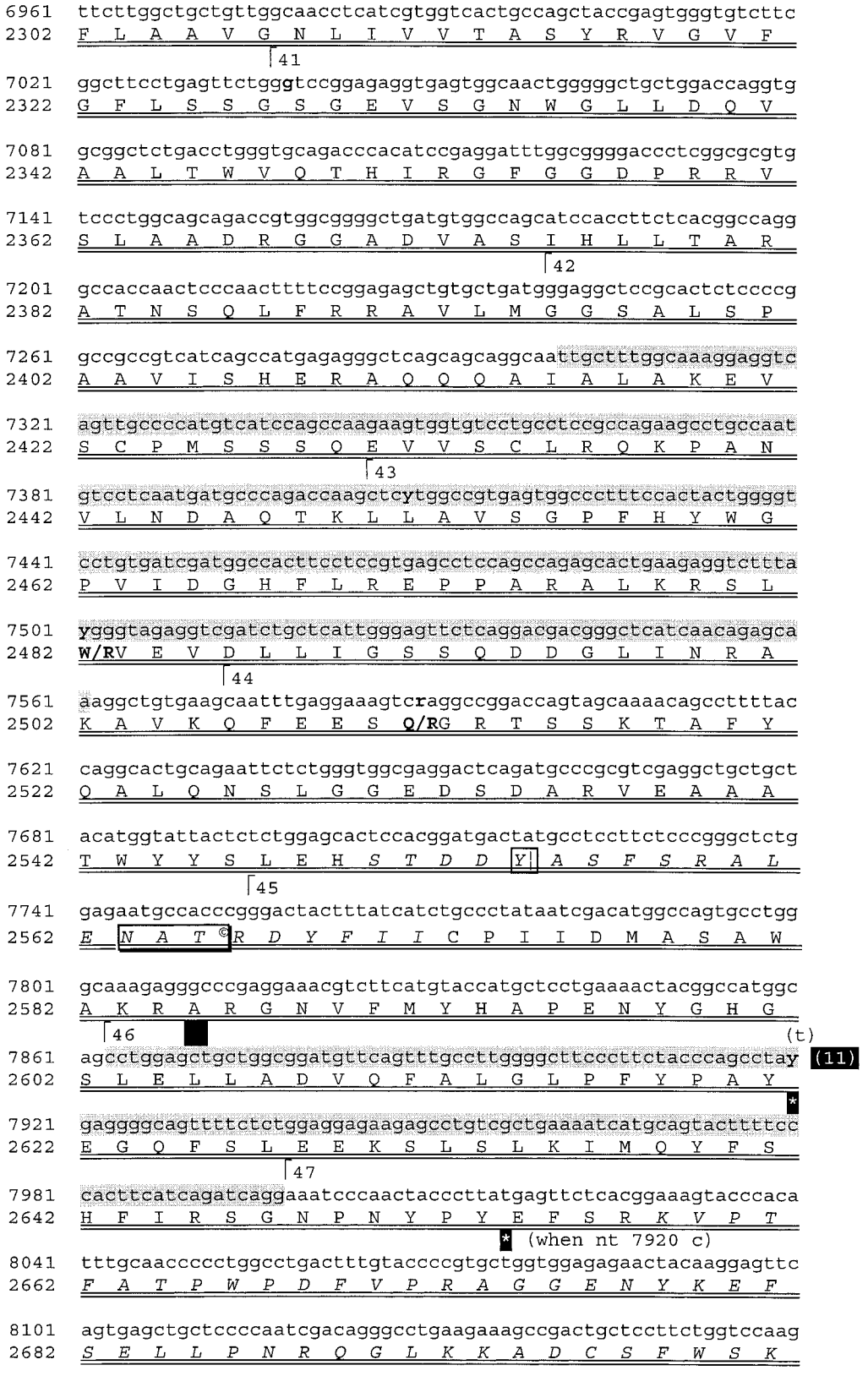


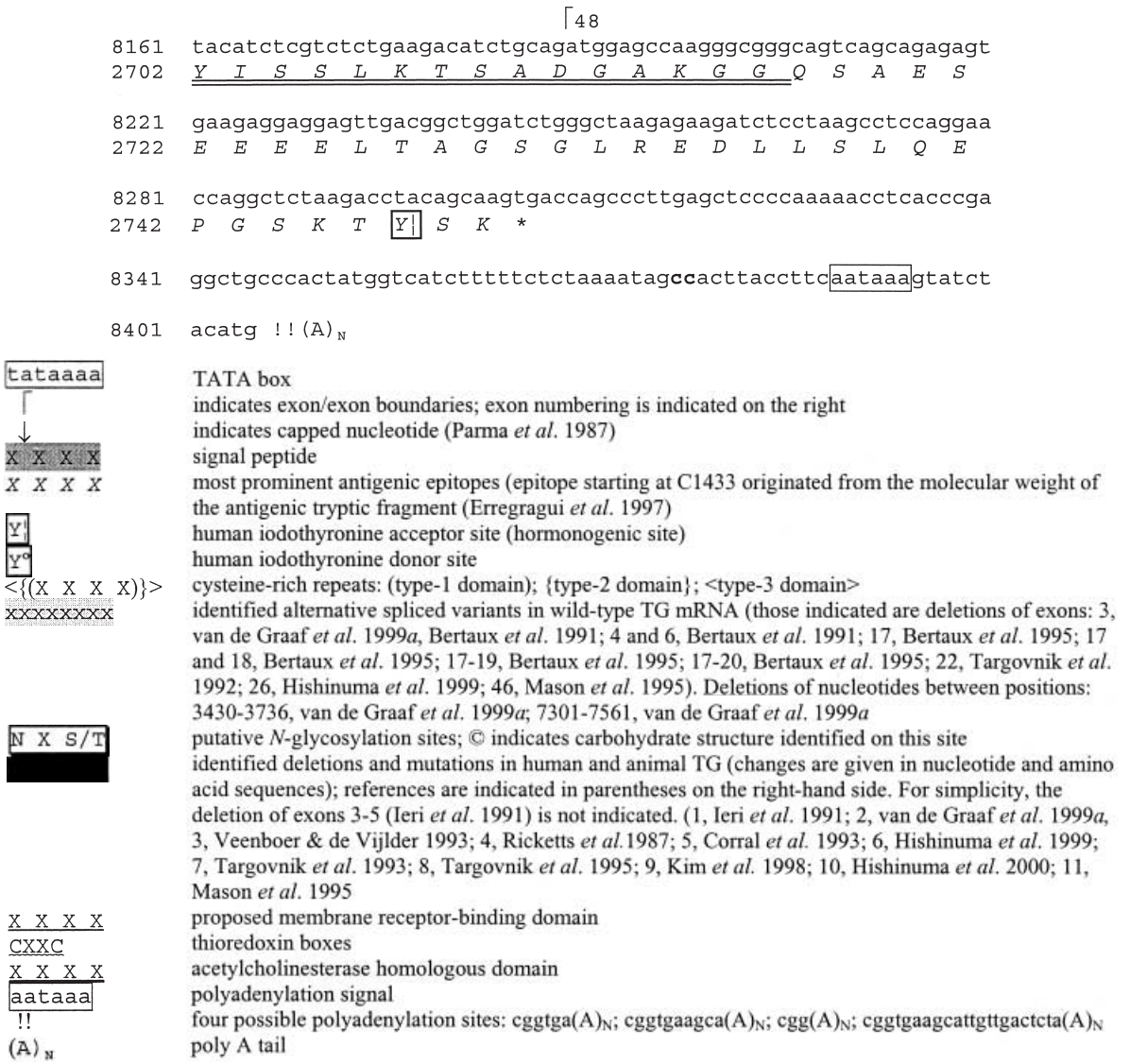












TATA box
indicates exon/exon boundaries; exon numbering is indicated on the right
indicates capped nucleotide (Parma *et al.* 1987)
signal peptide
most prominent antigenic epitopes (epitope starting at C1433 originated from the molecular weight of the antigenic tryptic fragment (Erregragui *et al.* 1997)
human iodothyronine acceptor site (hormonogenic site)
human iodothyronine donor site
cysteine-rich repeats: (type-1 domain); {type-2 domain}; <type-3 domain>
identified alternative spliced variants in wild-type TG mRNA (those indicated are deletions of exons: 3, van de Graaf *et al.* 1999*a*, Bertaux *et al.* 1991; 4 and 6, Bertaux *et al.* 1991; 17, Bertaux *et al.* 1995; 17 and 18, Bertaux *et al.* 1995; 17-19, Bertaux *et al.* 1995; 17-20, Bertaux *et al.* 1995; 22, Targovnik *et al.* 1992; 26, Hishinuma *et al.* 1999; 46, Mason *et al.* 1995). Deletions of nucleotides between positions: 3430-3736, van de Graaf *et al.* 1999*a*; 7301-7561, van de Graaf *et al.* 1999*a*
putative *N*-glycosylation sites; © indicates carbohydrate structure identified on this site
identified deletions and mutations in human and animal TG (changes are given in nucleotide and amino acid sequences); references are indicated in parentheses on the right-hand side. For simplicity, the deletion of exons 3-5 (Ieri *et al.* 1991) is not indicated. (1, Ieri *et al.* 1991; 2, van de Graaf *et al.* 1999*a*, 3, Veenboer & de Vijlder 1993; 4, Ricketts *et al.*1987; 5, Corral *et al.* 1993; 6, Hishinuma *et al.* 1999; 7, Targovnik *et al.* 1993; 8, Targovnik *et al.* 1995; 9, Kim *et al.* 1998; 10, Hishinuma *et al.* 2000; 11, Mason *et al.* 1995
proposed membrane receptor-binding domain
thioredoxin boxes
acetylcholinesterase homologous domain
polyadenylation signal
four possible polyadenylation sites: cggtga$(A)_N$; cggtgaagca$(A)_N$; cgg$(A)_N$; cggtgaagcattgttgactcta$(A)_N$
poly A tail

Figure 2 Sequence of the human TG gene. The cDNA nucleotide sequence is given in the upper line (nucleotides 1–8307) and the amino acid translation is given below the first nucleotide of each triplet (amino acids 19–2749). Nucleotide polymorphisms are indicated in the nucleotide ambiguity code, and encoded amino acid polymorphisms are separated by forward slashes (/); an asterisk (*) indicates a stop codon. All variations with respect to the original sequence (Malthièry & Lissitzky 1987), in nucleotides and amino acids, are indicated (bold).

the carboxy-terminal end of the TG molecule has been proposed for apical membrane interaction. Unfortunately, the mutation in *cog/cog* mice located in this domain leads to ER storage of apparently misfolded TG molecules (Kim *et al.* 1998) and therefore does not directly contribute to the elucidation of the function of this domain.

The heparin-binding domain consists of non-basic/non-acidic amino acids (X) alternating with basic amino acids (B) and is shown to have different formats, for example XBBXBBX or XBBXBX (Cardin & Weintraub 1989). In rat TG, the heparin-binding domain 693 RRLKRP 699 is proposed to bind to megalin that participates in the endocytosis of TG–T_4 complexes from the follicular lumen (Marino *et al.* 1999). However, in human TG, such a heparin-binding domain is not found in the homologous region, but a possible equivalent could be the 2582 AKRARG 2587 motif.

Recently, three highly conserved thioredoxin boxes have been identified in mammalian TG, and studies on a bovine TG fragment revealed a role for these boxes in self-assisted disulfide-bond formation leading to the intermolecular cross-linking of lumenal TG (Klein *et al.* 2000).

The existence of a wide variation of alternative RNA transcripts for human TG has been described in normal and pathological situations, but only 11 have been mapped exactly in the nucleotide sequence (Fig. 2). Most of these alternative transcripts are found by using sensitive RT-PCR techniques, and their concentration in normal situations is less than that of the wild-type spliced transcript. Only in two pathological situations described is the alternatively spliced transcript concentration increased relative to that of the wild type and linked to the mutation (Targovnik *et al.* 1993, Ricketts *et al.* 1987). The template used in RT-PCR reactions is RNA isolated from thyroid tissue, and usually no histological studies are performed to determine whether all thyrocytes contain these alternatively spliced transcripts or if they show a heterogeneous pattern.

Although nothing is known about the effect of alternative splicing on post-translational processes, one can speculate that these TG molecules are concluded 'misfolded' by the ER quality control and therefore never reach the lumen. On the other hand, these alternatively spliced transcripts may result in a population of alternative TG protein molecules. Such protein molecules could have alternative conformations, leading to a downregulating role in thyroid hormonogenesis (because the conformation is less optimal for T_4 formation) or to an increase in TG degradation and T_4 release in the case of spliced out cysteine-rich repeated domains. Alternative TG protein molecules might have a lower degree of iodination and could be involved in the recently proposed feedback suppressor function of TG (Suzuki *et al.* 1998, 1999). These studies describe a suppressive effect of poorly iodinated follicular TG protein molecules on the transcription of thyroid-specific transcription factors TTF-1, TTF-2 and Pax8, serving as a feedback autoregulator of TSH-stimulated thyroid function. This novel regulatory pathway awaits further confirmation.

In the normal human population, the existence of alternative spliced mRNAs lacking exon 4 of the TG gene has been described (Bertaux *et al.* 1991). The implications of this for TG function are difficult to assess, because excision of the important donor tyrosine residue 130 and the second type-1 cysteine-residue-rich repeated domain (both of which are encoded by exon 4) is linked to thyroid dyshormonogenesis (Ieri *et al.* 1991).

In addition to its biochemical role, TG serves as the target for an autoimmune response both in healthy subjects and in subjects with autoimmune diseases associated with thyroid dysfunction (Salvi *et al.* 1988). Serological studies have shown that there are at least 40 antigenic epitopes on human TG (Roitt *et al.* 1958). Recently, Tomer discussed the three general approaches used in these studies and their specific results (Tomer 1997). He concluded that while the epitopes recognized by anti-TG antibodies found in healthy individuals are polyclonal, the autoimmune thyroid-disease-related epitopes are different, and are restricted to specific epitopes that are mainly conformational and probably located in non-hormonogenic regions of the molecule. Another study, using 26 monoclonals in cross-inhibition and epitope-linkage experiments, resulted in a tentative map of human TG epitopes, providing information on the three-dimensional arrangement of functional domains (Erregragui *et al.* 1997). Since alternatively spliced transcripts are defined which, upon translation, lack part of an antigenic region (Fig. 2), one could be critical of the use of tryptic digests in combination with N-terminal amino acid sequencing and cross-inhibition in order to map antigenic epitopes.

In view of this, the update of the human TG coding sequence includes essential reference data necessary to screen for mutations and to map epitopes, and structure–function subsequently for further relationship studies on thyroglobulin.

Acknowledgements

We thank the Ludgardine Bouwman Foundation (The Netherlands) for financial support. This work was also partially supported by grants from Universidad de Buenos Aires (FA 124/95, TB 80/98), CONICET (7260/96, 0853/98), FONCYT (05–00000–01591/98) and Fundación Alberto J Roemmers (1997, 1998). H M T is an established investigator of the Argentine National Research Council (CONICET).

References

Baas F, Bikker H, Geurts van Kessel A, Melsert R, Pearson PL, de Vijlder JJM & van Ommen G-JB 1985 The human thyroglobulin gene: a polymorphic marker localized distal to c-myc on chromosome 8 band q24. *Human Genetics* **69** 138–145.

Baas F, van Ommen G-JB, Bikker H, Arnberg AC & de Vijlder JJM 1986 The human thyroglobulin gene is over 300 kb long and contains introns of up to 64 kb. *Nucleic Acids Research* **14** 5171–5186.

Bastiani P, Papandreou J, Blanck O, Fenouillet E, Thibault V & Miquelis R 1995 On the relationship between completion of N-acetyllactosamine oligosaccharide units and iodine content of thyroglobulin: a reinvestigation. *Endocrinology* **136** 4204–4209.

Bergé-Lefranc J-L, Cartouzou G, Mattei M-G, Passage E, Malezet-Desmoulins C & Lissitzky S 1985 Localization of the thyroglobulin gene by *in situ* hybridization to human chromosomes. *Human Genetics* **69** 28–31.

Berndorfer U, Wilms H & Herzog V 1996 Multimerization of thyroglobulin (TG) during extracellular storage: isolation of highly crosslinked TG from human thyroids. *Journal of Clinical Endocrinology and Metabolism* **81** 1918–1926.

Bertaux F, Noël M, Malthiéry Y & Fragu P 1991 Demonstration of a heterogeneous transcription pattern of thyroglobulin mRNA in human thyroid tissues. *Biochemical and Biophysical Research Communications* **178** 586–592.

Bertaux F, Noël M, Lasmoles F & Fragu P 1995 Identification of the exon structure and four alternative transcripts of the thyroglobulin-encoding gene. *Gene* **156** 297–301.

Brix K, Lemansky P & Herzog V 1996 Evidence for extracellularly acting cathepsins mediating thyroid hormone liberation in thyroid epithelial cells. *Endocrinology* **137** 1963–1974.

Cardin AD & Weintraub HJ 1989 Molecular modeling of protein–glycosaminoglycan interactions. *Atherosclerosis* **9** 21–32.

Corral J, Martín C, Pérez R, Sánchez I, Mories MT, San Millan JL, Miralles JM & González-Sarmiento R 1993 Thyroglobulin gene point mutation associated with non-endemic simple goitre. *Lancet* **341** 462–464.

De Deken X, Wang D, Many MC, Costagliola S, Libert F, Vassart G, Dumont JE & Miot F 2000 Cloning of two human thyroid cDNAs encoding new members of NADPH oxidase family. *Journal of Biological Chemistry* **275** 23227–23233.

Dunn JT 1999 Thyroglobulin: chemistry and biosynthesis. In *Werner's and Ingbar's The Thyroid*. Eds LE Braverman & RD Utiger. Philadelphia: JB Lippincott.

Dupuy C, Ohayon R, Valent A, Noël-Hudson M-S, Dème D & Vrion A 1999 Purification of a novel flavoprotein involved in the thyroid NADPH oxidase. *Journal of Biological Chemistry* **274** 37265–37269.

Erregragui K, Prato S, Miquelis R, Barrande C, Daniel C & Fert V 1997 Antigenic mapping of human thyroglobulin. *European Journal of Biochemistry* **244** 801–809.

Gentile F & Salvatore G 1993 Preferential sites of proteolytic cleavage of bovine, human and rat thyroglobulin. *European Journal of Biochemistry* **218** 603–621.

van de Graaf SAR, Pauws E, de Vijlder JJM & Ris-Stalpers C 1997 The revised 8307 base pair coding sequence of human thyroglobulin transiently expressed in eukaryotic cells. *European Journal of Endocrinology* **136** 508–515. GenBank locus U93033.

van de Graaf SAR, Cammenga M, Ponne NJ, Veenboer GJM, Gons MH, Orgiazzi J, de Vijlder JJM & Ris-Stalpers C 1999*a* The screening for mutations in the thyroglobulin cDNA from six patients with congenital hypothyroidism. *Biochimie* **81** 425–432. GenBank locus U93033.

van de Graaf SAR, Ris-Stalpers C, Veenboer GJM, Cammenga M, Santos C, Targovnik HM, de Vijlder JJM & Medeiros-Neto G 1999*b* A premature stop codon in thyroglobulin mRNA results in familial goiter and moderate hypothyroidism. *Journal of Clinical Endocrinology and Metabolism* **84** 2537–2542.

Henry M, Zanelli E, Piechaczyk M, Pau B & Malthièry Y 1992 A major human thyroglobulin epitope defined with monoclonal antibodies is mainly recognized by human autoantibodies. *European Journal of Immunology* **22** 315–319.

Hishinuma A, Takamatsu J, Ohyama Y, Yokozawa T, Kanno Y, Kuma K, Yoshida S, Nobuo M & Ieiri T 1999 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* **84** 1438–1444.

Hishinuma A, Furudate S-I, Oh-Ishi M, Nagakubo N, Namatame T & Ieri T 2000 A novel missense mutation (G2320R) in thyroglobulin causes hypothyroidism in *rdw* rats. *Endocrinology* **141** 4050–4055.

Ieri T, Cochaux P, Targovnik HM, Suzuki M, Shimoda S-I, Perret J & Vassart G 1991 A 3′ splice site mutation in the thyroglobulin gene responsible for congenital goiter with hypothyroidism. *Journal of Clinical Investigation* **88** 1901–1905.

Kim PS & Arvan P 1993 Hormonal regulation of thyroglobulin export from the endoplasmic reticulum of cultured thyrocytes. *Journal of Biological Chemistry* **268** 4873–4879.

Kim PS & Arvan P 1998 Endocrinopathies in the family of endoplasmic reticulum (ER) storage diseases: disorders of protein trafficking and the role of ER molecular chaperones. *Endocrine Reviews* **19** 173–202.

Kim PS, Hossain SA, Park Y-N, Lee I, Yoo S-E & Arvan P 1998 A single amino acid change in the acetylcholinesterase-like domain of thyroglobulin causes congenital goiter with hypothyroidism in the *cog/cog* mouse: a model of human endoplasmic reticulum storage diseases. *Cellular Biology* **95** 9909–9913.

Klein M, Gestmann I, Berndorfer U, Schmitz A & Herzog V 2000 The thioredoxin boxes of thyroglobulin: possible implications for intermolecular disulfide bond formation in the follicle lumen. *Biological Chemistry* **381** 593–601.

Kong YM, McCormick DJ, Wan Q, Motte RW, Fuller BE, Giraldo AA & David CS 1995 Primary hormonogenic sites as conserved autoepitopes on thyroglobulin in murine autoimmune thyroiditis: secondary role of iodination. *Journal of Immunology* **155** 5847–5854.

Lamas L, Anderson PC, Fox JW & Dunn JT 1989 Consensus sequences for early iodination and hormonogenesis in human thyroglobulin. *Journal of Biological Chemistry* **264** 13541–13545.

Lenarcic B & Bevec T 1998 Thyropins – new structurally related proteinase inhibitors. *Biological Chemistry* **379** 105–111.

Mallet B, Lejeune PJ, Ruf J, Piechaczyk M, Marriq C & Carayon P 1992 Tyrosine iodination and iodotyrosyl coupling of the N-terminal thyroid-hormone forming site of human thyroglobulin modulate its binding to auto- and monoclonal antibodies. *Molecular and Cellular Endocrinology* **88** 89–95.

Malthièry Y & Lissitzky S 1987 Primary structure of human thyroglobulin deduced from the sequence of its 8448-base complementary DNA. *European Journal of Biochemistry* **165** 491–498. GenBank locus HSTHYRR.

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

Mason ME, Dunn D, Wortsman J, Day RN, Day KH, Hobavk SJ, Myers HE, Shupnik MA & Dunn JT 1995 Thyroids from siblings with Pendred's syndrome contain thyroglobulin messenger ribonucleic acid variants. *Journal of Clinical Endocrinology and Metabolism* **80** 497–503.

Mendive FM, Vassart G & Targovnik HM 1997 Identification of a new thyroglobulin variant. The guanine-to-adenine transition resulting in the substitution of arginine 2510 by glutamine. *Thyroid* **7** 587–591.

Mendive FM, Rivolta CM, Vassart G & Targovnik HM 1999 Genomic organization of the 3′ region of the human thyroglobulin gene. *Thyroid* **9** 903–912. GenBank loci HSTHRG and HSTHGL.

Mercken L, Simons MJ, DeMartynoff G, Swillens S & Vassart G 1985 Presence of hormonogenic and repetitive domains in the first 930 amino acids of bovine thyroglobulin as deduced from the cDNA sequence. *European Journal of Biochemistry* **147** 59–64.

Mezgrhani H, Mziaut H, Courageot J, Oughideni R, Bastiani P & Miquelis R 1997 Identification of the membrane receptor binding domain of thyroglobulin. *Journal of Biological Chemistry* **272** 23340–23346.

Mezghrani A, Courageot J, Mani JC, Pugniere M, Bastiani P & Miquelis R 2000 Protein disulphide 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* **275** 1920–1929.

Molina F, Pau B & Granier C 1996 The type-1 repeats of thyroglobulin regulate thyroglobulin degradation and T3, T4 release in thyrocytes. *FEBS Letters* **391** 229–231.

Moya CM, Mendive FM, Rivolta CM, Vassart G & Targovnik HM 2000 Genomic organization of 5′ region of the human thyroglobulin gene. *European Journal of Endocrinology* **143** 789–798. GenBank accession numbers AF105681, AF105682, AF105683, AF105684, AF105685, AF105686, AF105687, AF170486, AF170487, AF170488, AF170489.

Parma J, Christophe D, Pohl V & Vassart G 1987 Structural organization of the 5′ region of the thyroglobulin gene. *Journal of Molecular Biology* **196** 769–779. GenBank locus HSTHYR.

Pauws E, Moreno JC, Tijssen M, Baas F, de Vijlder JJM & Ris-Stalpers C 2000 Serial analysis of gene expression as a tool to assess

the human thyroid expression profile and to identify novel thyroidal genes. *Journal of Clinical Endocrinology and Metabolism* **85** 1923–1927.

Pauws E, van Kampen AHC, van de Graaf SAR, de Vijlder JJM & Ris-Stalpers C 2001 Heterogeneity in polyadenylation cleavage sites in mammalian mRNA sequences: implications for SAGE analysis. *Nucleic Acids Research* **29** 1690–1694. GenBank locus U93033.

Rabin M, Barker PE, Ruddle FH, Brocas H, Targovnik H & Vassart G 1985 Proximity of thyroglobulin and c-myc genes on human chromosome 8. *Somatic Cell and Molecular Genetics* **11** 397–402.

Ricketts MH, Simons MJ, Parma J, Mercken L, Dong Q & Vassart G 1987 A nonsense mutation causes hereditary goitre in the Afrikander cattle and unmasks alternative splicing of thyroglobulin transcripts. *PNAS* **84** 3181–3184.

Rivolta CM, Mendive FM, Moya CM, Vassart G & Targovnik HM 1999 Genomic organization of the central region of the human thyroglobulin gene. GenBank accession numbers AF169654, AF169655, AF169656, AF169657, AF169658, AF169659, AF169660, AF169661, AF169662, AF169663, AF169664.

Roitt IM, Campbell PN & Doniach D 1958 The nature of the thyroid autoantibodies present in patients with Hashimoto's thyroiditis (lymphadenoid goitre). *Biochemical Journal* **69** 248–254.

Saboori AM, Rose NR & Burek CL 1995 Amino acid sequence of a tryptic peptide of human thyroglobulin reactive with sera of patients with thyroid diseases. *Autoimmunity* **22** 87–94.

Salvi M, Fukazawa H, Bernard N, Hiromatsu Y, How J & Wall JR 1988 Role of autoantibodies in the pathogenesis and association of endocrine autoimmune disorders. *Endocrine Reviews* **9** 450–466.

Suzuki K, Lavaroni S, Mori A, Ohta M, Saito J, Pietrarelli M, Singer DS, Kimura S, Katoh R, Kawaoi A & Kohn LD 1998 Autoregulation of thyroid-specific gene transcription by thyroglobulin. *PNAS* **95** 8251–8256.

Suzuki K, Mori A, Lavaroni S, Miyagi E, Ulianich L, Katoh R, Kawaoi A & Kohn LD 1999 *In vivo* expression of thyroid transcription factor-1 RNA and its relation to thyroid function and follicular heterogeneity: identification of follicular thyroglobulin as a feedback suppressor of thyroid transcription factor-1 RNA levels and thyroglobulin synthesis. *Thyroid* **9** 319–331.

Swillens S, Ludgate M, Mercken L, Dumont JE & Vassart G 1986 Analysis and structure homologies between thyroglobulin and acetylcholinesterase: possible functional and clinical significance. *Biochemical and Biophysical Research Communications* **137** 142–148.

Targovnik H, Cochaux P, Corach D & Vassart G 1992 Identification of a minor Tg mRNA transcript lacking 171 nucleotides in RNA from normal and goitrous thyroids. *Molecular and Cellular Endocrinology* **84** R23–R26.

Targovnik HM, Medeiros-Neto G, Varela V, Cochaux P, Wajchenberg BL & Vassart G 1993 A nonsense mutation causes human hereditary goiter with preferential production of a 171 nucleotide deleted thyroglobulin RNA messenger. *Journal of Clinical Endocrinology and Metabolism* **77** 210–215.

Targovnik HM, Vono J, Billerbeck AEC, Cerone GE, Varela V, Mendive F, Wajchenberg BL & Medeiros-Neto G 1995 A 138-nucleotide deletion in the thyroglobulin ribonucleic acid messenger in a congenital goiter with defective thyroglobulin synthesis. *Journal of Clinical Endocrinology and Metabolism* **80** 3356–3360.

Tomer Y 1997 Anti-thyroglobulin autoantibodies in autoimmune thyroid diseases: cross-reactive or pathogenic? *Clinical Immunology and Immunopathology* **82** 3–11.

Ulianich L, Suzuki K, Mori A, Nakazato M, Pietrarelli M, Goldsmith P, Pacifico F, Consiglio E, Formisano S & Kohn LD 1999 Follicular thyroglobulin (TG) suppression of thyroid-restricted genes involves the apical membrane sialoglycoprotein receptor and TG phosphorylation. *Journal of Biological Chemistry* **274** 25099–25107.

Veenboer GJM & de Vijlder JJM 1993 Molecular basis of the thyroglobulin synthesis defect in Dutch goats. *Endocrinology* **132** 377–381.

de Vijlder JJM & Den Hartog MT 1998 Anionic iodotyrosine residues are required for iodothyronine synthesis. *European Journal of Endocrinology* **138** 227–231.

Yang S-X, Pollock HG & Rawitch AB 1996 Glycosylation in human thyroglobulin: location of the N-linked oligosaccharide units and comparison with bovine thyroglobulin. *Archives of Biochemistry and Biophysics* **327** 61–70.

Zheng G, Marino M, Zhao J & McCluskey RT 1998 Megalin (gp330): a putative endocytic receptor for thyroglobulin. *Endocrinology* **139** 1462–1465.

Received in final form 14 March 2001
Accepted 12 April 2001