The human thyroglobulin gene contains two 15 – 17 kb introns near its 3'-end

Gert-Jan B.van Ommen*, Annika C.Arnberg§, Frank Baas, Huguette Brocas+, André Sterk, Wil H.H.Tegelaers, Gilbert Vassart+ and Jan J.M.de Vijlder

Department of Pediatric Endocrinology, Academical Medical Center, GV2-131, Meibergdreef 9, 1105 AZ Amsterdam, §Biochemical Laboratory, The University, Groningen, The Netherlands and +IRIBHN, Campus Hôpital Erasme, Route de Lennik 808, B-1070 Brussels, Belgium

Received 1 March 1983; Accepted 25 March 1983

ABSTRACT

We have cloned overlapping segments of the human thyroglobulin gene from a genomic cosmid library. Restriction mapping and electron microscopy show that a region of 38 kb at or near the 3'-end of this gene encodes only 850 nucleotides or 10% of the messenger RNA (mRNA) sequence. The region contains five exons of 130-210 nucleotides, split by introns of 1 to 15-17 kb. This represents the lowest ratio of coding to non-coding DNA (2.2%) found thusfar in any eukaryotic gene. Blot hybridization under non-stringent conditions shows the presence of only one copy of this gene in the human genome and the absence of other closely related sequences.

INTRODUCTION

Several studies in western countries indicate that, amongst humans, about one in every 3000 newborns suffers from congenital hypothyroidism (1,2). About five percent of these cases are due to inborn errors in the synthesis of thyroglobulin (M.H. Gons and J.J.M. de Vijlder, unpublished observations). Thyroglobulin (Tg) is the 19S thyroid glycoprotein from which thyroxine (T_4) and 3,5,3'-triiodothyronine (T_3, the active thyroid hormone) are synthesized. Native Tg contains two identical 12S subunits of molecular weight 330 kD (3, 4,5) , encoded by a 33S mRNA, which corresponds to 8.5-8.7 kb (6,7,8,9).

With the final aim of elucidating defects in the synthesis of Tg we are studying the human Tg gene by molecular cloning. Here we describe the characterization of genomic clones containing segments from the 3'-region of the Tg gene, isolated from a human cosmid library. Our results show that this gene is extremely large. In the segment analysed, five exons which code together for only 850 nucleotides or 10% of the mRNA, lie dispersed over 38 kb of chromosomal DNA, separated by introns of 1 to 15-17 kb. Our further results suggest that the length of the entire human Tg gene may exceed 100 kb. Hybridization under conditions of different stringency indicates that this gene is a single copy gene and that there are no closely related (pseudo)genes present.

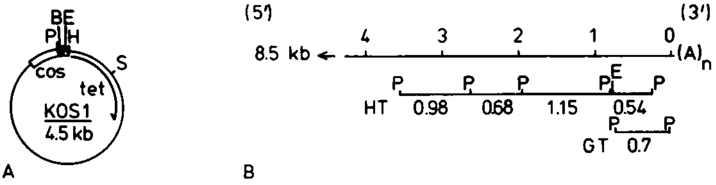

Fig.1. Cosmid vector and probes for thyroglobulin sequences.
A. Map of KOS I, a cosmid vector of 4.5 kb conferring tetracyclin resistance (constructed by Peter Little, Chester Beatty Research Institute, London). Cleavage sites for PstI(P), BglII(B), EcoRI(E), HindIII(H, see below) and SalI(S) are indicated. The segment from EcoRI clockwise to PstI is derived from MUA-5 (10), a cosmid vector constructed from pBR322. The 29-bp sequence between EcoRI and HindIII is present as a duplication-inversion on both sides of the EcoRI site (indicated with a connection line inside the circle). The fragment between the PstI site and the HindIII site left of the EcoRI site originates from pGA46 (11) and contains a BglII site, in which the partial MboI fragments from chromosomal DNA were cloned.
B. Map location of human Tg (HT) cDNA PstI probes with respect to the Tg coding sequence. The position of the PstI sites (P) and the EcoRI site in HT 0.54 (E) are indicated. This EcoRI site overlaps the PstI site for one nucleotide (A.R. Stuitje, recent results). GT 0.7 indicates the map position (relative to the goat Tg coding sequence) of a PstI fragment from a goat Tg cDNA clone (12). This fragment was used as an additional probe for the 3'-extremity of the gene.

MATERIALS AND METHODS

Construction of the cosmid library. Equal amounts of KOS I DNA (Fig.1A) were cleaved with either HindIII or PstI, subsequently with S_1-nuclease and finally with BglII, to obtain fragments with a functional BglII site either rightward or leftward of the cos sequence and a non-ligatable (blunt) end at the other side (13). The fragments were electrophoretically purified (14) and ligated in approximately fivefold molar excess to semi-randomly generated 30-50 kb fragments of human DNA. These were made by partial digestion of human fibroblast DNA with MboI and size fractionation on a sucrose gradient (the fragments were kindly provided by Dr. Ed Fritsch). In vitro packaging in phage lambda particles and transfection was carried out essentially according to Grosveld et al (15), with some modifications. After determining the optimal ratio of the packaging mix constituents, these were thawed batchwise, mixed together at 0°C, refrozen in 50 µl aliquots in liquid N_2 and stored at -70°C, without loss of packaging efficiency. Optimal efficiency was obtained by ligating 0.2-0.4 µg inserts with 0.2 µg of both left and right vector arms in 5 µl, packaging 1 µl aliquots of this ligation in 6 µl of thawed packaging mix (scaling up reduced the efficiency markedly, we used more reactions instead), diluting each packaging with 100 µl of phage buffer and using this to infect 400 µl of cells, diluted freshly 1:4 in 10 mM $MgSO_4$. The host cells E.coli 1046 (803 sup F, $recA^-$) were grown overnight in NZY medium containing 0.4% maltose, harvested and resuspended in the same amount of 10 mM $MgSO_4$. When stored at +4°, the host cells could be used for 1-2 weeks without loss of transfection efficiency. Dilution of the cells prior to infection was found to increase transfection efficiency. In more recent constructions of human, goat and bovine libraries, efficiencies of 0.7-1.8 x 10^6 clones/µg insert DNA were obtained following this protocol (G.J.B. van Ommen en G. de Martynoff, recent results). The infected cells were plated on 10 µg/ml tetracyclin NZY plates. The screening was done according to Ish-Horowitz and Burke (13), with a nick-translated mixture of four contiguous PstI fragments, electrophoretically purified from three different human Tg cDNA clones in pBR322 (16) (Figure 1B). To the (pre-)hybridization reactions of the screening steps (see below) 1-2 µg/ml of sheared, denatured vector DNA and 2-4 µg/ml sheared, denatured E.coli DNA were added. Cloned DNA was isolated as in ref. 13.

Restriction enzymes, gel electrophoresis and hybridization. Restriction enzymes were from Boehringer and New England Biolabs and were used as stated by the supplier. S_1-nuclease was from Sigma. Gel electrophoresis was through horizontal 0.6% agarose gels (Sigma, type I), 50 mM Tris-acetate, 1 mM EDTA (pH 7.7). Prehybridization was carried out in 0.9 M NaCl, 50 mM Na-phosphate (pH 7.4), 5 mM EDTA, 50% formamide, 10 x Denhardt's solution (17) and 100 µg/ml of sonicated denatured salmon sperm DNA, for 1-4 h at 42°C. Hybridization was done under similar conditions, with 10% dextran sulphate added, for 18 h, with a probe concentration of 10-20 ng/ml. The washes were 0.5-1 h each in 0.3 M NaCl, 0.03 M Na-citrate, pH 7.0 (2 x SSC), 0.1% Na-dodecylsulphate at 50°C, 55°C, 60°C and 68°C, successively, The heterologous hybridization with the GT 0.7 probe (Figure 1B) was washed only four times at 50°C.

Restriction mapping techniques. All the clones were mapped independently by cross-blot hybridization analysis of EcoRI-cleaved DNA versus both HindIII- and BamHI-cleaved DNA. The maps were confirmed by multiple enzyme digestions and hybridization with exon probes and sub-cloned fragments. The overlap of different clones was verified by two-by-two crossblot analysis of EcoRI-restricted clones. This will be reported in detail elsewhere (F. Baas and G.J.B. van Ommen, manuscript in preparation).

Preparation of human 33S Tg mRNA and electron microscopy. Human 33S Tg mRNA was purified from the goiter of a patient with hyperthyroidism, using the guanidinium extraction method (18), modified to contain two extraction steps with an equal volume of chloroform directly after the homogenization in guanidinium-HCl. All homogenizations were for 2 min at topspeed in a Sorvall Omnimixer, followed by centrifugation for 10 min at 10.000 rpm and recovery of the aqueous phase (19). Subsequently, the RNA was sedimented through CsCl (18), poly(A)-RNA was isolated by poly(U)-Sepharose chromatography (12) and 33S RNA was isolated by elution of the appropriate fraction from agarose gel (20). Electron microscopy was carried out as described by Arnberg et al (20). The precise conditions are given in the legends to Fig. 5.

RESULTS

Restriction analysis of cloned DNA.

We have constructed a human cosmid library in the vector KOS 1 (Figure 1A)and screened this library with four adjacent PstI fragments of human Tg cDNA. These fragments cover a total of 3.2 kb near the 3'-end of Tg mRNA (Figure 1B). From 160,000 colonies, 2-3 times the haploid human genome, we obtained several partially overlapping clones, with chromosomal insert sizes of 40.9 to 47.2 kb. Three of these clones, cHT10, cHT2 and cHT6, were detected by the 3'-proximal cDNA probe, HT 0.54. Of these, cHT10 and cHT2 also hybridized with the 5'-adjacent probe HT 1.15. A restriction map was constructed of these clones and their region of overlap is presented in Figure 2.Blot hybridization of the cDNA probes using various restriction digestions of cHT2 and two subcloned exon-containing EcoRI fragments is shown in Figure 3. The EcoRI restriction pattern of cHT2 is shown in lane 1. The HT 0.54 probe detects EcoRI fragments of 12.5, 5.3 and 3.8 kb (lane 2). The 5.3 kb fragment is 14.2 kb apart from the 3.8 kb fragment, with the 12.5 kb fragment mapping in between the two (Figure 2).

The subcloned 5.3 kb fragment (Figure 3; lane 3) is cleaved by HindIII into two fragments of 4.5 and 0.85 kb, both of which hybridize with HT 0.54 (lane 4). Since this cDNA probe itself lacks HindIII sites, the 5.3 kb fragment contains at least two exonic regions, separated by an intronic HindIII site. HT 0.54 contains an EcoRI site which overlaps the 5'-PstI site for one nucleotide(Figure 1B) and should thus overlap a genomic EcoRI site without being able to detect sequences 5' of it. The 5.3 kb EcoRI fragment is the most 5'-located one detected with HT 0.54 (Figure 2). Immediately 5' of this lies a 1.3 kb fragment which is only detected with the next cDNA probe, HT 1.15 (lane 18, 19; see below). The EcoRI site between the 5.3 and 1.3 kb genomic fragments thus corresponds with the EcoRI site in HT 0.54. In conclu-

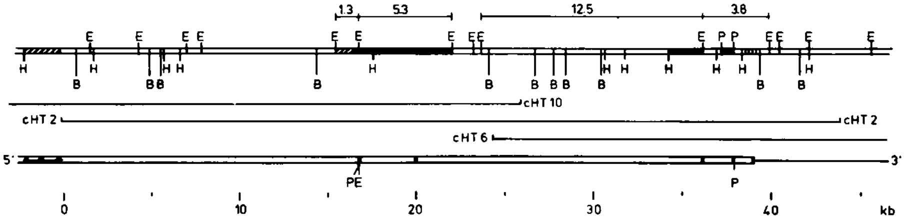

Fig.2. Map of the 3'-proximal region of the human Tg gene. The upper bar shows the restriction map for EcoRI(E), HindIII(H) and BamHI(B). The sizes of the EcoRI fragments detected with cDNA probes are indicated. The position of the (only) two sites for PstI (P) within the 3'-proximal 3.8 kb EcoRI exonic fragment are shown. The regions which are detected by the cDNA probes HT 1.15, HT 0.54 and GT 0.7 are hatched diagonally, horizontally and vertically, respectively. The segments of the chromosomal DNA contained within clones cHT10, cHT2 and cHT6 are underlined below the map. The restriction maps of the three clones were constructed independently and confirmed by cross-hybridization and hybridization of subcloned fragments to multiple restriction blots of all three clones (F. Baas and G.J.B. van Ommen, manuscript in preparation). The lower bar represents the map of introns (open) and exons (closed), derived from correlation of electron microscopy and hybridization data. The spacing of the five 3'-located exons, contained within cHT2, follows from the electron microscopy. The most 5'-located of these exons is aligned with the EcoRI site separating the 1.3 and 5.3 kb fragments (see text) and the 3'-penultimate exon is aligned with the 3'-end of the 0.75kb PstI fragment detected with the HT 0.54 probe within the 3.8 kb EcoRI fragment (see text). This alignment produces an exon map which is consistent with the homologous and heterologous hybridization data. The exonic 4.0 kb EcoHindIII fragment of cHT10 (Figure 3) has a 1.8 kb overlap with the 5'-end of cHT2 (F. Baas and G.J.B. van Ommen, see above). This limits the position of this exonic region to 2.2 kb immediately upstream of cHT2.

sion, the 5.3 kb EcoRI fragment begins within an exon at its 5'-end and contains at least one more exon located more than 0.85 kb downstream, beyond the HindIII site. The hybridization of the subcloned 3.8 kb EcoRI fragment (lane 5) is limited to a 1.45 kb HindIII fragment, not cleaved by BamHI (lanes 6,7) and further to a 0.75 kb PstI fragment, not cleaved by HindIII (lanes 8,9). This PstI fragment is the most 3'-located region detected with HT 0.54 (Figure 2). Since HT 0.54 is a PstI fragment itself, the 3'-end of the genomic 0.75 kb PstI fragment lies within an exon. Upon cross-hybridization (not shown), the HT 0.54 sequences were found to fall 200-300 bp short of the 3'-end of a goat cDNA clone GT 0.7 (Figure 1B), which extends to within 20 bp of the 3'-end of goat Tg mRNA (12). Thus, to search for exons downstream of those detected with HT 0.54, the GT 0.7 probe was hybridized with the 3.8 kb EcoRI fragment. One additional hybridizing region was found on a 1.55 kb HindIII-EcoRI fragment (Figure 3, lane 10). This could be cleaved with BamHI into a 0.95 kb HindIII-BamHI fragment (Figure 3, lane 11; Figure 2).

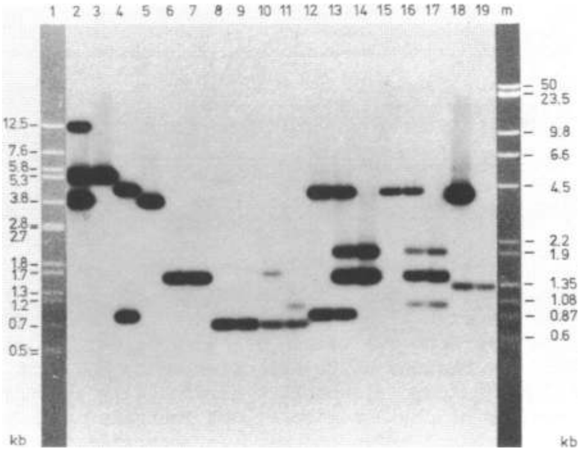

Fig.3. Restriction analysis of cosmid clones and subclones. Samples of 300 ng of cosmid DNA or 50 ng of subcloned DNA were restricted and electrophoresed in parallel, according to the following scheme (in which E stands for EcoRI,H for HindIII, B for BamHI, P for PstI and the subclones containing the 5.3 kb or 3.8 kb EcoRI fragments of cHT2 in pAT153 have been denoted p5.3 and p3.8, respectively): 1,2)cHT2 xE; 3) p5.3 xE; 4) p5.3 xE,H; 5) p3.8 x E; 6) p3.8 xH; 7) p3.8 xE,H,B;8)p3.8 xE,P; 9,10) p3.8 xE,H,P; 11) p3.8 xE,H,P,B; 12, 15,18) cHT10 xE,H; 13,19) cHT2 xE,H; 14) cHT6 xE,H; 16) cHT2 xE,H,B; 17) cHT6 xE,H,B. The gel was blotted onto nitrocellulose (21). The strips corresponding to lanes 2-9 and 12-14 were hybridized with HT 0.54 probe, those corresponding to lanes 10-11 and 15-17 with GT 0.7 probe and those corresponding to lanes 18-19 with HT 1.15 probe. After washing (see methods) the strips were mounted in the proper order and autoradiographed. Lane 1 shows the total EcoRI fragment pattern of cHT2, stained with ethidium bromide with the fragment sizes indicated. Lane m shows the marker fragments of λ, λ x HindIII and ΦX x HaeIII DNA. NB. Lane 1 was not run in the same gel, the fragments on the autoradiogram should be compared with the marker lane m.

Digestion of the overlapping clones cHT10, cHT2 and cHT6 with EcoRI plus HindIII, followed by hybridization with HT 0.54, shows that cHT10 and cHT2 share the 5'-exonic regions on 4.5 and 0.85 kb fragments (Figure 3,lanes 12,13; cf.lane 4) and that cHT2 and cHT6 share an exon fragment of 2.0 kb and the 1.45 kb HindIII fragment discussed above (lanes 13,14; cf.lanes 6, 7). The 2.0 kb fragment is cleaved by HindIII from the 3'-end of the 12.5 kb EcoRI fragment (Figure 2). Hybridization of the overlapping clones with GT 0.7 produces a very similar pattern with the following differences.The goat probe apparently does not see the most 5'-located exon (see above) on the 0.85 kb EcoRI-HindIII fragment (lanes 15,16; cf lanes 4,12 and 13) and it shows the most 3'-located exon (see above) on the 0.95 kb HindIII-BamHI fragment (lanes

15,16; cf lanes 11,13 and 14). (Cleavage of the samples 16 and 17 with BamHI, in addition to EcoRI and HindIII, allows resolution of the two 3'-proximal exonic segments without altering the length of any other exon-containing Eco-HindIII fragment, see Figure 2). Lane 17 shows that cHT6, which extends 24 kb 3'-ward of cHT2, lacks additional regions of homology whith GT 0.7.

The HT 1.15 cDNA probe which is 5'-adjacent to HT 0.54, does detect additional exonic regions upstream of cHT2 on cHT10. The exonic regions common to cHT2 and cHT10 map within the 5.3 kb EcoRI fragment and at the junction of the 1.3 kb and 5.3 kb EcoRI fragments (see above). Of this latter, HT 1.15 detects the part on the 1.3 kb fragment (Figure 3; lane 18,19). An additional exonic region on cHT10 is found on a 4.0 kb fragment (lane 18). The complete identity of cHT10 and cHT2 over their region of overlap indicates that the 5'-end of cHT2 consists entirely of intronic sequences. This is confirmed by the electron microscopy results (see below). These sequences span at least 15 kb. Of the 4.0 kb EcoRI-HindIII fragment visible in lane 18, 1.8 kb are present at the 5'-end of cHT2 (Figure 2). The exonic sequences on the 4.0 kb fragment thus map within 2.2 kb from the 5'-end of cHT2, limiting the size of the 5'-intron to 15-17 kb. Further study of the sequences upstream of cHT2 has thusfar been hampered by the lack of overlapping clones from this segment of the Tg gene. A more upstream region of at least 35 kb is present in overlapping clones as well and is currently under study (see Discussion).

Restriction analysis of chromosomal DNA; number of Tg genes.

Hybridization of human chromosomal DNA with the HT 0.54 probe shows the 12.5, 5.3 and 3.8 kb bands (Figure 4, strip 1). Low stringency wash gives a background smear but no additional bands (strip 2).The GT 0.7 probe detects the same bands (strips 3, 4). Although the heterologous hybridization has a lower signal, it is well discernible after low stringency wash (strip 4). The subcloned 3.8 kb EcoRI fragment detects its genomic counterpart, with very faint additional bands showing up after low stringency wash (strips 5, 6). The subcloned 5.3 kb EcoRI fragment contains repeated sequences and produces a smear (strip 7). By raising the stringency, only the 5.3 kb fragment becomes detectable (strip 8). Other subcloned regions of cHT2, insofar lacking repeated sequences, also show unique genomic localizations (F. Baas and H. Bikker, unpublished data). From these results we conclude that the Tg gene is present only once in the haploid genome. Related or pseudogenes, if present at all, have a lower homology than the human-goat heterologous hybrids. In support of this conclusion, we consistently find that 10 µg of chromosomal DNA give a hybridization signal intermediate to that of 0.1 and 0.2 ng of

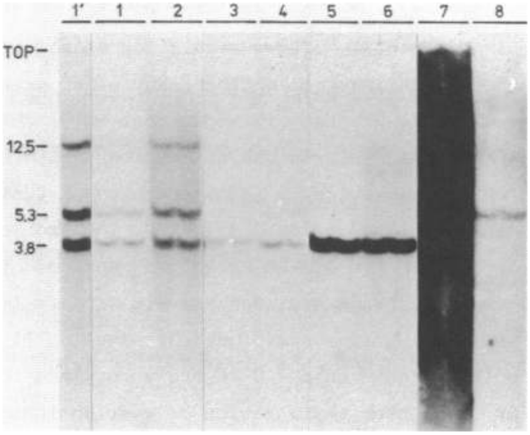

Fig.4. Restriction analysis of human chromosomal DNA. Human placental DNA was digested with EcoRI to completion. Samples of 10 µg were electrophoresed in parallel through 0.6% agarose gel in tris-acetate buffer (see Methods). The gel was blotted onto a nitrocellulose filter and double-lane strips were hybridized with the following probes (see text): HT 0.54 (1,2); GT 0.7 (3,4); genomic 3.8 kb EcoRI fragment (5,6) and genomic 5.3 kb EcoRI fragment (7,8). Strips 1,3,5,7 and 8 were washed four times at 65° for 30 min in 2 x SSC, 1 x SSC, 0.3 x SSC and 0.1 SSC, consecutively; strip 8 was washed for another 30 min in 0.1 x SSC at 75°. Strips 2,4 and 6 were washed once in 2 x SSC and three times in 1 x SSC for 30 min at 65°. The strips were autoradiographed for 24 h except lane 1', which is a 4 d exposure of the left lane of strip 1.

cloned DNA (not shown). This yields a complexity difference between 0.5 and 1.0 x 10^5, consistent with a single copy gene present either in a cosmid clone (50 kb) or in chromosomal DNA (3 x 10^6 kb).

Electron microscopy

We have carried out an electron microscopy analysis of hybrids of SalI-linearized cHT2 and human 33S Tg mRNA. A typical example of the hybrids observed is shown in Figure 5. Five short exons can be detected,with sizes, from left to right, of 161 ± 44(n=6), 133 ± 36(n=9), 136 ± 19 (n=2), 203 ± 27(n=20) and 211 ± 33(n=17) nucleotides. We interpret the short single-stranded RNA tail, adjacent to the most rightward exon, as the poly(A) tail. Its size varies between 33 and 213 nucleotides (n=12). The size of the introns, from left to right, is respectively 2820 ± 210(n=6), 14,100 ± 660(n=9), 1580 ± 105(n=22) and 1050 ± 105(n=18) nucleotides. These results are in good agreement with the restriction-hybridization results and provide a more precise map of position and number of exons in cHT2 (Figure 2, lower bar). In the restriction map the leftmost and rightmost exons of cHT2 are about 22 kb a-

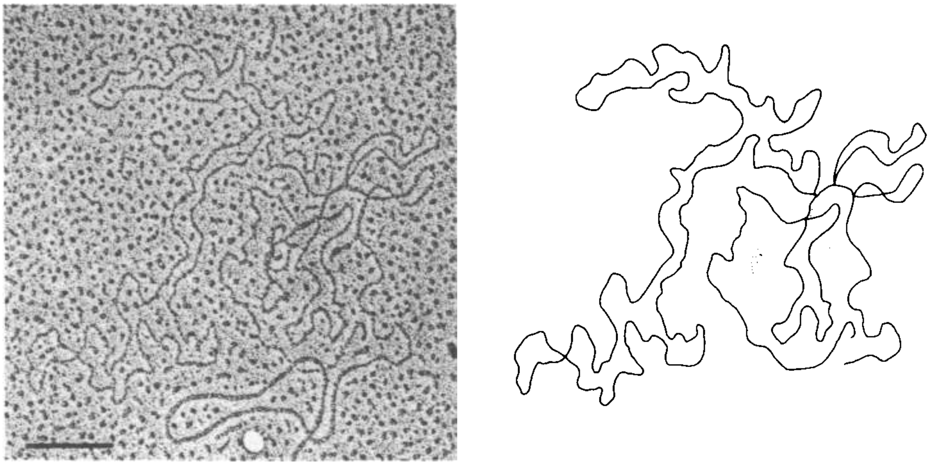

Fig.5. Electron micrograph of a hybrid between cHT2 and Tg mRNA. A mixture of SalI-linearized cHT2 DNA and human 33S Tg mRNA in 70% formamide, 0.1 M PIPES (pH 7.8), 0.4 M NaCl and 10 mM EDTA was subjected to a linear temperature gradient of 63-53°C over a period of 2 h and spread from 64% formamide, 20 mM PIPES, 80 mM NaCl, 8 mM EDTA, 60 mM Tris (pH 8.0), 0.01% cytochrome c. All further treatments were as in Arnberg et al. (20). The bar represents 0.2 μm. In the schematical drawing the continuous line represents DNA and the dotted line RNA. Following our interpretation (see text), the hybrid sequences are depicted with the 5' to 3' order from left to right. The duplex region at the lower edge of the hybrid is a renatured or incompletely denatured segment of cHT2 and is not included in the drawing.

part and in the electron microscopy 20.5 kb. We attribute this to a calibration difference between the two methods. In the map, when correcting for this difference, the large intron becomes 15.3 kb.

DISCUSSION

We have constructed and isolated several cosmid clones overlapping the 3'-proximal region of the human Tg gene. A segment of 44 kb, present in clone cHT2, is covered by two other clones, cHT10 and cHT6, which overlap each other for 1.5 kb in the middle of this region. This multiple overlap, coupled with the identity of exonic restriction fragments in cloned and chromosomal DNA, shows that the cloned region is authentic. Both restriction analysis and electron microscopy independently arrive at a remarkably dispersed map of the exons in this part of the gene: Five exons of 130-210 nucleotides, in total 850 nucleotides or 10% of the entire Tg mRNA, are encoded in a 38 kb region at or near the 3'-end of the gene, which contains two giant introns of 15 - 17 kb and three smaller introns of 1 to 3 kb. This exon content of 2.2% is the lowest reported thusfar in any eukaryotic gene. Other genes have been found to contain very large introns, such as the mouse dihydrofolate reductase gene (22) or the immunoglobulin heavy chain locus (23). In the first case, the gene as a whole is much smaller, containing one 16.5 kb and four smaller introns, which implies a lower overall complexity of posttranscriptional processing. In the second case, the region actually consists of many alternative gene segments, the organisation of which plays a functional role in antibody class switching. This does not apply to the Tg gene. There is no significant sequence reiteration within the Tg-coding region (4, 24, 25) and cDNA segments, when hybridized with chromosomal or cloned DNA, show only singly encoded positions (see above). Our current knowledge however does not completely rule out the existence of minor variant 33S Tg mRNA species, which may have followed alternative processing pathways and contain exons, alternative to those found in the present study. This will be studied using _in vitro_ labelled RNA.

Recent studies of the rat Tg gene have yielded very similar results. An internal region was shown to contain approximately 1 kb of coding sequences, also split up into exons of 180-200 bp, and spread over 14 kb of genomic DNA (26). Others (27,28) have shown that the exons of the rat Tg gene are homogeneously sized (ca 200 bp) and more widely spread in the 3'-region than in more internal segments. In all these cases, however, the lack of overlap between the λ-clones has interfered with a precise positioning of the exons. The human genomic clones we have found using more upstream Tg cDNA probes, show an overlap over at least 35 kb, containing several exonic regions with restriction maps identical to chromosomal DNA. These exonic regions once more are widely separated (F. Baas, H. Bikker and G.J.B. vanOmmen, recent results). This segment has not yet been linked with the presently described segment by overlapping clones. Between the two segments, approximately 30 kb of exon-containing EcoRI fragments are detected in chromosomal DNA, using the cDNA probes HT 0.68 and HT 1.15 (data not shown and G. de Martynoff pers. comm.). According to these findings, the portion of the gene which codes for about 3.2 kb at the 3'-end of the 8.5 kb mRNA, is already in excess of 100 kb. Depending upon whether the 5'-half of the Tg gene has a similar or different genomic organization, the entire human Tg gene might thus span 100-300 kb of chromosomal DNA. The Tg-coding sequence in the 3'-region and, in view of the results from the rat gene,possibly more upstream as well, is split up into many small exons of 130-210 bp. This remarkably dispersed gene organization adds yet another level of complexity to thyroid hormone synthesis. A process which already appears to be surprisingly uneconomic: complete degradation of an entire 19S glycoprotein, consisting of about 5600 aminoacids, to yield at most four molecules of thyroid hormone (29,30).

It remains to be shown conclusively whether the most 3'-located exon in cHT2 actually represents the 3'-end of the Tg gene. The following evidence is in support of this. First, the GT 0.7 probe, which extends to within 20 bp from the poly(A) tail of goat mRNA (12) and contains ca 200-300 bp of Tg mRNA-coding sequences 3' of the HT 0.54 probe, does not detect additional exons on clone cHT6, which extends 24 kb 3'-ward of cHT2. This result per se would not rule out the presence of 3'-exons not detected with a heterologous probe. A second argument, however, is that the electron micrographs consistently show an RNA-tail of variable length (30-210 nucleotides),protruding from the last exon of 210 bp in cHT2. This is precisely as can be expected for the poly(A) tail of a cellular mRNA. The interpretation that the last exon in cHT2 is the end of the Tg gene, coupled with the absence thusfar of introns found in 3'-untranslated sequences on eukaryotic mRNA's, suggests that the Tg protein-coding region could extend to within ca.200 nucleotides from the poly(A) tail. In support of this, the size of the unglycosylated Tg subunit (300 kD or about 2800 aminoacids, requiring 8400 nucleotides of coding sequence) indicates that there is little space for untranslated regions on the 8.5-8.7 kb Tg mRNA.

Finally, it should be interesting to study the expression of giant transcription units like the Tg gene in detail. For this, the availability of animal model systems with Tg synthesis defects is of great help. We are studying an inbred Dutch goat strain with congenital hereditary goiter, in which Tg is absent due to a non-translatable Tg mRNA of normal size (12). At present one can only wonder how (and why) nature performs the enormous task of correctly composing these bits of information from such a vast excess of non-coding sequences. For instance, does the splicing of the giant introns follow the same mechanism as that of the smaller ones, or is there an early compaction process, which removes large non-coding sequences from nascent transcripts(31), generating intermediately-sizedhnRNAs for further processing? A second question is that of transcriptional fidelity. On the basis of nucleotide tautomerism, transcription should have no higher fidelity than one error per 10^4 to 10^5 nucleotides (32). If so, it would be lethal for a gene to have significantly over 10^4 nucleotides transcriptionally involved (in a cis-fashion) with the emergence of the final gene product, whether protein-coding or containing signals for processing, transport, translation or mRNA halflife.

The extreme scarcity of polypeptides with sizes in excess of 300-400 kD (requiring around 10^4 nucleotides of protein-coding RNA) is at least consistent with this. The chances of introducing false processing signals, not at the DNA level (33), but at the RNA level, should increase proportionally with total primary transcript size. The existence of giant transcription units therefore suggests the presence of processing failsafe systems we are currently unaware of.

ACKNOWLEDGEMENTS

We are indebted to Drs. Tom Maniatis, Ed Fritsch, Doug Engel and staff for helpful instructions and preparations, to Mr. Stephen Goodbourn for his collaboration, to Dr. P. Borst for critical advice, to Drs. L.A. Grivell and A.Y.C. Liu for valuable comments on the manuscript, to Dr. G.S.P. Groot for allowing us to work in the CII facility of the Free University of Amsterdam and to Dr. E.F.J. Van Bruggen for his interest in this study. We thank Hennie Bikker for technical assistance, Klaas Gilissen for printing of the electron micrograph and Corrie Van den Berg for typing the manuscript. This work was supported in part by the Foundation for Medical Research (FUNGO), subsidized by the Netherlands Organisation for the Advancement of Pure Research (ZWO), the Jan Dekker Stichting and the Ludgardine Bouwman Stichting.

*To whom correspondence should be sent

REFERENCES

1. DeLange, F., Beckers, C., Höfer, R., König, M.P., Monaco, F. and Varonne, S. (1980) In: Neonatal Thyroid Screening (Burrow, G.N. and Dussault, J.H., eds.) Raven Press, New York, pp 107-131.
2. Van Herle, A.J., Vassart, G. and Dumont, J.E. (1979) New Engl. J. Med. 301, 239-249 and 307-314.
3. Edelhoch, H. (1965) Rec. Progr. Horm. Res. 21, 1-24.
4. Vassart, G. and Brocas, H. (1980) Biochem.Biophys. Acta 610, 189-194.
5. Marriq, C., Rolland, M. and Lissitzky, S. (1980) Eur. J. Biochem. 79, 143-149.
6. Vassart, G., Verstreken, L. and Dinsart, C. (1977) FEBS Lett.79,143-149.
7. Chebath, J., Chabaud, O., Bekarevic, A., Cartouzou, G. and Lissitzky, S. (1977) Biochem. Biophys. Res. Commun. 79, 267-273.
8. Van Ommen, G.J.B., Baas, F., Sterk, A. and De Vijlder, J.J.M. (1981) Ann. d'Endocrinologie 42, 11A.
9. Bergé-Lefranc, J.L., Cartouzou, G., Malthiery, Y., Perrin, F., Jarry, B. and Lissitsky, S. (1981) Eur. J. Biochem. 120, 1-7.
10. Meyerowitz, E.M., Guild, G.M., Prestidge, L.S. and Hogness, D.S. (1980) Gene 11, 271-282.
11. An, G. and Friesen, J.D. (1979). J. Bacteriol 140, 400-407.
12. De Vijlder, J.J.M., Van Ommen, G.J.B., Van Voorthuizen, W.F., Koch, C.A.M., Arnberg, A.C., Vassart, G., Dinsart, C. and Flavell, R.A. (1981) J. Mol. and Appl. Genet. 1, 51-59.
13. Ish-Horowitz, D. and Burke, J.F. (1981) Nucl. Acids Res. 9, 2989-2998.
14. Girvitz, S.R., Bacchetti, S., Rainbow, A.J. and Graham, F.L. (1980) Anal. Biochem. 106 , 492-496.

15. Grosveld, F.G., Dahl, H.H.M., De Boer, E. and Flavell, R.A. (1981) Gene 13, 227-237.
16. Brocas, H., Christophe, D., Pohl, V. and Vassart, G. (1982) FEBS Lett. 137 , 189-192.
17. Denhardt, D.T. (1966) Biochem. Biophys. Res. Commun. 23, 641-646.
18. Goodman, H.M. and MacDonald, R.J. (1979) In: Meth. in Enzymol. 68 (R.Wu, ed.) Academic Press, New York, pp 75-90.
19. Alvino, C.G., Tassi, V., Paterson, B.M. and DiLauro, R. (1982) FEBS Lett. 137, 307-313.
20. Arnberg, A.C., Van Ommen, G.J.B., Grivell, L.A., van Bruggen, E.F.J. and Borst, P. (1980) Cell 18, 313-319.
21. Southern, E.M. (1975) J. Mol. Biol. 98 , 503-517.
22. Schilling, J., Beverley, S., Simonsen, C., Course,G., Setzer, D., Feagin, J., McGrogan, M., Kohlmiller, N. and Schimke, R.T. (1982). In:Gene Amplification (R.T. Schimke, ed.) Cold Spring Harbor Laboratory Publications, New York, USA,pp 149-153.
23. Yaoita, Y., Kamagai, Y., Okumura, K. and Honjo, T. (1982) Nature 297, 697-699.
24. Mercken, L., Christophe, D. and Vassart, G. (1982) Ann. d'Endocrinologie 43, 31A.
25. Malthiery, Y., Cartouzou, C. and Lissitzky, S. (1982) Ann.d'Endocrinologie 43 , 32A.
26. Christophe, D., Pohl, V., Van Heuverswyn, B., De Martynoff, G., Dumont, J.E., Pasteels, L.J. and Vassart, G. (1982) Biochem. Biophys.Res.Commun. 105 , 1166-1175.
27. Avvedimento, V.E., Cocozza, S. and DiLauro, R. (1981) Ann.d'Endocrinologie 43 , 12A.
28. Cocozza,S., Obici, S., Condliffe, D., Musti, M. and Lissitzky, S. (1981) FEBS Lett. 132 , 29-32.
29. Lamas, L., Taurog, A., Salvatore, G. and Edelhoch, H. (1974) J. Biol. Chem. 249 , 2732-2737.
30. Maurizis, J.C., Marriq, C., Rolland , M. and Lissitzky, S. (1981) FEBS Lett. 132, 29-32.
31. Beyer, A.L., Banton, A.H. and Miller jr., O.L. (1981) Cell 26, 155-165.
32. Topal, M.D. and Fresco, J.R. (1976) Nature 263, 285-289 and 289-293.
33. Busslinger, M., Moschonas, N. and Flavell, R.A. (1981) Cell 27,289-298.