Multiple Messenger Ribonucleic Acid Transcripts and Revised Gene Organization of the Human TSH Receptor

AYUMU KAKINUMA AND YUJI NAGAYAMA*

Department of Internal Medicine 1, University of Occupational and Environmental Health, 1-1 Iseigaoka, Yahatanishi-ku, Kitakyushu 807-0804, Japan

*Department of Pharmacology 1, Nagasaki University School of Medicine, 1-12-4 Sakamoto, Nagasaki 852-8523, Japan

Abstract. Northern blot analysis of human TSH receptor (hTSHR) messenger ribonucleic acid (mRNA) expression has previously demonstrated multiple species of transcripts in the thyroid gland, suggesting the presence of multiple transcription initiation sites, alternatively spliced forms or alternate polyadenylation (poly(A)) sites. The first two have already been reported elsewhere. To clarify alternate poly(A) sites in the hTSHR gene, the present study was designed to characterize three full-length hTSHR cDNAs with distinct poly(A) signals that we have previously cloned. The comparison of the nucleotide sequencing data on the 3'UTR of these three clones to the Draft Human Genome in NCBI database revealed that the 3' segment of exon 10 of hTSHR gene contains three tandem repeats of the poly(A) sites, from which are expressed three full-length TSHR mRNAs with distinct 3'UTR length. The longest one appears to be a predominant transcript. From these data, together with (i) the previously reported organization of hTSHR genome and (ii) use of the Draft Human Genome to localize the unidentified sequence in the alternatively spliced form of truncated hTSHR, we propose the complete structure of hTSHR gene. Rather than 10 exons, our analysis suggests that hTSHR gene seems to contain 13 exons and 12 introns. At least three full-length TSHR mRNAs with distinct poly(A) sites and five alternatively spliced forms of TSHR mRNAs are expressed from the single hTSHR gene.

Key words: TSH receptor, Alternate polyadenylation, Alternative splicing

(Endocrine Journal 49: 175–180, 2002)

THE cDNA for the human TSH receptor (hTSHR) has been cloned in 1989–1990 [1–4]. The full-length hTSHR cDNA is approximately 4 kilobases (kb) in length, and contains a single open reading frame of 2292 base pairs (bp), which encodes a 764 amino acid polypeptide, and ~100 bp of the 5' untranslated region (UTR) and ~1.6 kb of the 3' UTR [1–5]. However, Northern blot analysis of hTSHR messenger ribonucleic acid (mRNA) expression in the thyroid gland reveals multiple mRNA transcripts with major, longer forms of 3.9 to 4.6 kb in length and several minor, smaller ones ranging from ~1.2 to 1.8 kb [3, 6, 7]. Since hTSHR is encoded by a single

gene located on chromosome 14q31 [8, 9], the multiple mRNA transcripts on Northern blot analysis suggest either multiple transcription initiation sites, alternative splicing or alternate polyadenylation (poly(A)) sites. Indeed studies on the transcription initiation site of hTSHR gene performed by Gross *et al.* [10] have identified multiple transcription initiation sites that produce 52 to 157 bp of the 5' UTRs. Moreover, alternatively spliced forms of truncated hTSHR lacking the entire transmembrane/cytoplasmic regions have previously been cloned [6, 7, 11]. These forms, ranging from 250 bp to 1.7 kb in length, are likely to correspond to some of the minor, smaller transcripts detected by Northern blot as mentioned above.

In addition, when we cloned hTSHR cDNA with two synthetic oligonucleotides based on the reported amino acid sequences of a putative TSHR and related receptors [1], we found three distinct alternate

Received: October 3, 2001

Accepted: December 28, 2001

Correspondence to: Yuji NAGAYAMA, M.D., Department of Pharmacology 1, Nagasaki University School of Medicine, 1-12-4 Sakamoto, Nagasaki 852-8523, Japan

poly(A) signals in the full-length hTSHR cDNA clones (Nagayama & Rapoport, unpublished data), further suggesting the use of the alternate poly(A) signals at the 3' UTR of hTSHR gene. The present study was therefore designed to characterize these TSHR mRNA variants with the different 3' UTR. Our data demonstrated that the full-length hTSHR mRNA transcripts utilize three alternate poly(A) signals, and that hTSHR gene appears to be expressed predominantly from the third poly(A) signal. Furthermore, comparison of the 3'UTR sequence obtained here, as well as the intronic sequence we have previously obtained [10], with the Draft Human Genome in NCBI database verifies the location of these sequences in hTSHR gene.

Materials and Methods

Nucleotide sequencing

Three full-length hTSHR cDNA clones with different sequences at 3' ends (#4, #15 and #22) used in this study have previously been cloned (refs. 1 and 5 and Nagayama & Rapoport, unpublished data). The nucleotide sequences of the 3' UTRs of the clones #15 and #22 were determined on both strands by the dideoxynucleotide chain termination method using an ABI 377 automated DNA sequencer (Perkin-Elmer, Foster, CA). The clone #4 has already been sequenced completely [5]. Nucleotide homology searches were performed against the Draft Human Genome in NCBI database (Homo sapiens genomic contig sequences) [12].

Northern blot analysis

Total RNA from a human thyroid tissue obtained at the time of subtotal thyroidectomy of a Graves' patient was extracted with ISOGEN (Nippon Gene, Tokyo, Japan). Written informed consent was obtained from the patient. Twenty-five µg of total RNA was subjected to Northern blot analysis as previously described [5]. cDNA probes specific for TSHR mRNA were synthesized by polymerase chain reaction with TSHR cDNA #22 as a template. The oligonucleotide primers used were 5'-TTA ATG GGT CCC CAC AGA TGG TCC-3' (sense) and 5'-CAT GTT TGA GTG CAA ATT TTA GT-3' (antisense) for P1, 5'-AGA TTC GTG TTT ATA TAA

AAG AGT-3' (sense) and 5'-GAC TTT TCT AGC AGG ACA TTA-3' (antisense) for P2 and 5'-ACC TAT TAA TCA TCT CTT C-3' (sense) and 5'-TAT ATT TGA GGG AAA GGG-3' (antisense) for P3. Cyclophilin cDNA was also used for rehybridization to quantify the amounts of RNA loaded. The probes were labeled with [α - 32 P] dCTP by BcaBEST Labeling Kit (Takara Biochemicals, Osaka, Japan) to a specific activity of approximately 1×10^9 cpm/µg DNA.

Results and Discussion

The complete nucleotide sequences of the 3' UTR of hTSHR cDNA clones #4, #15 and #22 are shown in Fig. 1. The shortest (#15), intermediate (#4) and longest (#22) clones were 950, 1572 and 1941 bp in length, respectively. The sequences of the shortest (#15) and intermediate (#4) clones were completely identical to the 5' end of the longest clone (#22). The comparison of these sequences to the Draft Human Genome in NCBI database (Homo sapiens chromosome 14 working draft sequence segment (NT_010140.3 | Hs14_10297) in Homo sapiens genomic contig sequences) revealed no intronic sequence in this region. These data indicate that the 3' UTR of hTSHR gene is intronless and that the nucleotide sequence of the longest clone (#22) corresponds to the 3' end of exon 10. Thus, exon 10 of hTSHR gene comprises 1412 bp of the translated region encoding the transmembrane/cytoplasmic regions and 1942 bp of the 3' UTR (Fig. 2).

There are three poly(A) signals of the hexanucleotide AATAAA (shown in bold in Fig. 1) as well as some variants in the 3' UTR. Because the poly(A) signal is typically located 10 to 30 nucleotides upstream of the 3' end of mature mRNA [13, 14], the intermediate (#4) and longest (#22) clones seem to use the AATAAA sequence as a poly(A) signal which are localized ~15 bp upstream of their 3' ends. However, the AATAAA sequence in the 3' end of the clone #15 is unlikely to be utilized as poly(A) signal because this is localized ~130 bp upstream from the 3' end. Instead the AATATA sequence, which is 18 bp upstream from the 3' end (also shown in bold in Fig. 1), might be an alternate poly(A) signal for clone #15. Previous mutagenesis studies suggest that the efficacy of polyadenylation in this variant (AATA-TA) is ~10% of the wild type sequence [15].

Fig. 1 Nucleotide sequence of the 3' UTR of hTSHR gene. Numbering follows from the 3' end of the coding region. Arrows depict the 3' ends of 3 variants with alternate poly(A) signals. The poly(A) signals (AATAAA) and one of its variant (AATATA) are shown in bold. Underlined are the sequence data downstream of the 3' UTR obtained from NCBI database. The putative GT-rich region downstream of the 3' end of TSHR gene is shown in italics.

To determine which of three transcripts is predominantly expressed in the thyroid gland, we performed Northern blot analysis using three cDNA probes (P1-3) shown in Fig. 2. P3 is a common probe for all three transcripts, P2 is specific for the intermediate and longest ones, and P1 is specific for

the longest one. As shown in Fig. 3, all probes, which were essentially the same in size and had very similar specific activities, hybridized to a single transcript of ~4 kb long with the similar intensities. This suggests that the third AATAAA sequence appears to be the predominant poly(A) signal in the full-

Fig. 2 Structure of the hTSHR gene and exon usage in alternatively spliced forms of truncated TSHR mRNAs and the full-length TSHR transcripts with different poly(A) signals. Exons and introns are shown by boxes and thin lines, respectively. Black indicates the coding region of the full-length TSHR, and gray depicts the coding regions unique to the truncated TSHR. P1–3 are cDNA probes used for Northern blot analysis. pA, poly(A) signal; AAAn, poly(A) tail; *, stop codon.

Fig. 3 Northern blot analysis of hTSHR gene expression in a human thyroid gland. Twenty-five μ g of total RNA was analyzed for hTSHR mRNA expression by Northern blot analysis. The size was estimated by running 1 kb DNA ladder (GIBCO) in an adjacent lane. P1, P2 and P3 (see Fig. 2) were used as cDNA probes in lanes 1, 2 and 3, respectively.

length hTSHR mRNAs.

In addition to the poly(A) signal mentioned above (the hexanucleotide AATAAA), the upstream and downstream elements are also known to define accuracy and efficacy of a 3'-processing site [13, 14]. The upstream element, a T-rich region, is not present in TSHR gene. The downstream element is a GT- or T-rich region, usually within 50 nucleotide 3' of the poly(A) site. The nucleotide sequence downstream

of the 3' end of TSHR exon 10 (obtained from NCBI database; underlined in Fig. 1) is T-rich, whereas the corresponding regions of the first and second poly(A) sites are not. This finding, together with the variant poly(A) signal used for the shortest clone (#15), may explain the efficient cleavage and addition of poly(A) at the third poly(A) site.

In addition, we used the Draft Human Genome in NCBI database to localize the unique sequence encoding the C-terminus of the alternatively spliced, truncated TSHR we had previously cloned (ST4, hTSHR-v1.3 and TSHR-I in refs. 6, 7 and 11, respectively). As shown in Figs. 2 and 4, the 3' end sequence of TSHR-I (ST4) localized to intron 8, ~160 bp downstream of the 3' end of exon 8. Because this sequence encodes a short peptide, this must be called "exon". We therefore propose to name this "exon 8A". Exon/intron boundary is conserved to a canonical splice consensus sequence [16] (gt—ag, underlined in Fig. 4). Indeed exon 8A has the typical poly(A) signal near its 3' end (shown in bold in Fig. 4).

Hunt *et al.* [6] have also cloned other truncated TSHR variants (ST1–3 and 5 in Fig. 2). Because the "intronic" nucleotide sequences in these variants are not available, their precise localization cannot be determined. We speculate that the 3' end sequences of ST1 and ST5 may occur in intron 1, and intron 8, downstream of the 3' end of exon 8A, respectively.

Exon 8 CAAGCTTGCT									
<u>GTGAGTAAGA</u>	CATACAAAAG	AATATTTTAG	TCTTTT	TTTTT	TCTTTT	TTTTT	TTTTT	GGAAT	60
AAATGGAGAC	ATTAAGGGGA	GAATCTTATG	TTCAAGGGTA	GAAAATGTTG	CTGTCTCGGG				120
TAAAGGCTTC	TGCAAGTCCC	TCTTTCCTG	GCTGTATAGG	CTACCTCTTG	GAAGAAAGTC				180
CTTGTCCTTT	GAGACTCAGA	AGGCCCCACG	CTCCAGTATG	CCATCATGAT	GCCTGCTAAG				240
GCAGCCACCT	TGGTGACAT	GCTCACAGAG	GCTCTGTTCA	TGGAGCAGCT	GCTGTTTGA				300
AAATTTTGAA	ATGCAAGATC	CACAAC TAGA	TGGAAGGCAC	TCTAGTCTTT	GCAGAAAAA				360
ATGTACCTGA	ATGTACATTG	CACAATGCCT	GGCACAAAGA	AGGAAGAATA	TAAATGATAG				420
TTCGACTCGT	CTGTGGAAGA	ACTTACAATC	ATGGGGAAAG	ATGGAATAAA	AACATTTT				480
AAACAGCATG	AAAACATACC	TATTGTGCA	TTTCCTTTGA	TTTACACCC	CACAT....				

Fig. 4 Nucleotide sequence of the 3' end of exon 8 (10 bp) and the 5' end of intron 8 (530 bp) of hTSHR gene obtained from GENE BANK. The 3' end of exons 8 and 8A are boxed. The putative poly(A) signal AATAAA is shown in bold. The consensus sequence of exon/intron boundary (gt-ag) is underlined. Numbering starts at the 5' end of intron 8.

With the same reason as the definition of exon 8A, we call these sequences "exons 1A and 8B". Furthermore, ST2 and ST3 are unspliced at their exon 3/intron 3 boundary and contain the 5' end of intron 3, suggesting that exon 3 is longer than originally reported. We here call it "exon 3'". Again all the truncated variants have the typical poly(A) signals near their 3' ends (6). It is reported that the expression levels of ST4 and ST5 mRNAs are 25 to 50% of that of the full length mRNA, and those of ST1-3 mRNAs are too low to be detected on Northern blot analysis [6, 11].

What is the physiological significance of alternative splicing and alternate poly(A) usage in the 3' UTR of the TSHR gene? Alternative splicing is an important molecular mechanism for protein diversity derived from a single gene [17]. Thus different lengths of truncated TSHRs lacking the transmembrane/cytoplasmic region can be produced by alternative splicing, which may be secreted as soluble TSH binding proteins into peripheral circulation [11]. In contrast to use of differential poly(A) sites following alternative splicing, the tandem poly(A) sites in the 3'UTR produce multiple forms of mRNA that differ in their 3'UTR and not in the coding region. It is possible that alternate poly(A) usage impacts on the final amount of a protein product, be-

cause the different forms of mRNAs may have different stabilities and/or translatabilities. Indeed tissue-specific or cell cycle-/developmentally-regulated differential poly(A) usage has been described in some genes such as eukaryotic initiation factor 2 α [13]. However, hTSHR mRNA levels in the thyroid gland are remarkably stable and resistant to regulation by chronic TSH stimulation [18]. Further studies will be required to clarify these issues.

Finally, from our results together with previously published data on hTSHR gene structure [10], we propose the complete organization of hTSHR gene, as shown in Fig. 2. Thus hTSHR gene consists of 13 exons and 12 introns. Exon 10 has three tandem repeats of the poly(A) sites. Five alternatively spliced forms and three full-length variants with the alternate poly(A) signals are so far identified to be expressed from hTSHR gene; the longest clone appears to be the predominant mRNA transcript.

Acknowledgment

We thank Dr. Basil Rapoport (Cedars-Sinai Research Institute, Los Angeles, LA, USA) for revising this manuscript.

References

1. Nagayama Y, Kaufman KD, Seto P, Rapoport B (1989) Molecular cloning, sequence and functional expression of the cDNA for the human thyrotropin receptor. *Biochem Biophys Res Commun* 165: 1184-

- 1190.
2. Misrahi M, Loosfelt H, Atger M, Sar S, Guiochon-Mantel A, Milgrom E (1990) Cloning, sequencing and expression of human TSH receptor. *Biochem Biophys Res Commun* 166: 394–403.
 3. Libert F, Lefort A, Gerard C, Parmentier M, Perret J, Ludgate M, Dumont JE, Vassart G (1989) Cloning, sequencing, and expression of the human thyrotropin (TSH) receptor: evidence for binding of autoantibodies. *Biochem Biophys Res Commun* 165: 1250–1255.
 4. Frazier AL, Robbins LS, Stork PJ, Sprengel R, Segaloff DL, Cone RD (1990) Isolation of TSH and LH/CG receptor cDNAs from human thyroid: regulation by tissue specific splicing. *Mol Endocrinol* 4: 1264–1276.
 5. Kakinuma A, Chazenbalk GD, Filetti S, McLachlan SM, Rapoport B (1996) Both the 5' and 3' non-coding regions of the thyrotropin receptor messenger RNA influence the level of receptor protein expression in transfected mammalian cells. *Endocrinology* 137: 2664–2669.
 6. Hunt N, Willey KP, Abend N, Balvers M, Jahner D, Northemann W, Ivell R (1995) Novel splicing variants of the human thyrotropin receptor encode truncated polypeptides without a membrane-spanning domain. *Endocrine* 3: 233–240.
 7. Graves PN, Tomer Y, Davies TF (1992) Cloning and sequencing of a 1.3 kb variant of human thyrotropin receptor mRNA lacking the transmembrane domain. *Biochem Biophys Res Commun* 187: 1135–1143.
 8. Rousseau-Merck MF, Misrahi M, Loosfelt H, Atger M, Milgrom E, Berger R (1990) Assignment of the human thyroid stimulating hormone receptor (TSHR) gene to chromosome 14q31. *Genomics* 8: 233–236.
 9. Libert F, Passage E, Lefort A, Vassart G, Mattei MG (1990) Localization of human thyrotropin receptor gene to chromosome 14q31 by in situ hybridization. *Cytogenet Cell Genet* 54: 82–83.
 10. Gross B, Misrahi M, Sar S, Milgrom E (1991) Composite structure of the human thyrotropin receptor gene. *Biochem Biophys Res Commun* 177: 679–687.
 11. Takeshita A, Nagayama Y, Fujiyama K, Yokoyama N, Namba H, Yamashita S, Izumi M, Nagataki S (1992) Molecular cloning and sequencing of an alternatively spliced form of the human thyrotropin receptor transcript. *Biochem Biophys Res Commun* 188: 1214–1219.
 12. Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acid Res* 25: 3389–3402.
 13. Edwards-Gilbert G, Veraldi KL, Milcarek C (1997) Alternative poly(A) site selection in complex transcription units: means to an end? *Nucleic Acid Res* 25: 2547–2561.
 14. Wahle E (1995) 3'-end cleavage and polyadenylation of mRNA precursors. *Biochim Biophys Acta* 1261: 183–194.
 15. Sheets MD, Ogg SC, Wickens MP (1990) Point mutations in AAUAAA and the poly (A) addition site: effects on the accuracy and efficacy of cleavage and polyadenylation in vitro. *Nucleic Acid Res* 18: 5799–5805.
 16. Mount S (1982) A catalogue of splice junction sequences. *Nucleic Acid Res* 10: 459–472, 1982.
 17. Lou H, Gagel RF (2001) Alternative ribonucleic acid processing in endocrine systems. *Endocr Rev* 22: 205–225.
 18. Maenhaut C, Brabant G, Vassart G, Dumont JE (1992) In vitro and in vivo regulation of thyrotropin receptor mRNA levels in dog and human thyroid cells. *J Biol Chem* 267: 3000–3007.