Genomic Cloning and Characterization of the Human Thrombin Receptor Gene

STRUCTURAL SIMILARITY TO THE PROTEINASE ACTIVATED RECEPTOR-2 GENE*

(Received for publication, October 10, 1995, and in revised form, December 18, 1995)

Valentina A. Schmidt¶, Emilia Vitale§, and Wadie F. Bahou‡

From the ¶Department of Medicine, State University of New York at Stony Brook, Stony Brook, New York 11794-8151 and the §Departments of Psychiatry and Genetics and Development, Columbia University, New York, New York 10032

The seven-transmembrane segment thrombin receptor (TR) represents the prototype of a putative family of proteolytically cleaved receptors that may include the proteinase activated receptor-2. A panel of somatic cell hybrids retaining distinct portions of human chromosome 5 were used to establish that the human TR gene is present as a single-copy locus within the region 5q11.2 → q13.3, confirming our previous localization using fluorescent in situ hybridization analysis. To further characterize the TR gene, overlapping clones from a human genomic library were isolated. Genomic analysis confirmed that the TR gene is of limited complexity, spanning ~27 kilobases and containing two exons separated by a large ~22-kilobase intron. Primer extension analysis using two 30-mer oligonucleotide primers known to be contained within the first exon identified the predominant transcription initiation site 351 base pairs upstream from the initiator methionine in both human umbilical vein endothelial and human erythroleukemia cells. Sequence analysis of the 5-flanking region revealed that the TR promoter is of limited complexity and includes a GATA motif, octamer enhancer sequences, AP-2-like sites, and Sp1 sites. These data provide evidence for remarkable similarity at the gene level between both proteolytically cleaved receptors described to date.

The serine protease α-thrombin plays a critical role in hemostasis and thrombosis via interactions with specific coagulation proteins and cells diversely involved in regulatory functions of the vessel wall. α-Thrombin is one of the most potent of the physiological stimuli for platelet aggregation (1), modulates the endothelial cell hemostatic response (2–4), and is mitogenic for vascular smooth muscle cells (5) and fibroblasts (6). G-protein-coupled thrombin receptor (TR)2 structurally similar to other members of the seven-transmembrane segment receptor family (7) has been isolated from a megakaryocytic (Dami) cell line (8). The cDNAs for similar receptors have been identified and cloned from human endothelial cells (9), CCL39 hamster lung fibroblasts (10), and rat vascular smooth muscle cells (11). Activation of the receptor by α-thrombin and/or synthetic ligands representing the new N terminus after thrombin cleavage (12–14) results in dual coupling to phospholipase C and adenylyl cyclase (15). Molecular mechanisms of thrombin receptor activation have been studied by this and other laboratories, and these results suggest that critical structural determinants regulating receptor activation exist within the long extracellular domain and the second extracellular loop (16, 17).

Despite the extensive and rapid accumulation of data directed toward elucidation of cellular activation mechanisms mediated by this receptor, little is known about the molecular genetics of the thrombin receptor. The concept of an extended gene family has recently been underscored with the isolation and cloning of a second proteinase-activated receptor (PAR-2) (18). Like the thrombin receptor, PAR-2 is activated by proteolytic cleavage and by synthetic peptides corresponding to the new N terminus after cleavage. Whereas trypsin unequivocally activates this receptor, the presence of additional physiological enzyme agonist(s) remains unproven although probable (18). We have now completed the molecular characterization of the human thrombin receptor gene and provide further evidence for remarkable similarity at the gene level between the human thrombin receptor and PAR-2 genes. These data provide conceptual support for the presence of a more extended gene family of proteolytically cleaved receptors that may have evolved from a common primordial gene.

MATERIALS AND METHODS

Supplies, Reagents, and Cell Lines—Restriction enzymes were purchased from Stratagene (La Jolla, CA), and avian myeloblastosis virus reverse transcriptase was from Seikagaku America, Inc. (Rockville, MD). Nylon membranes were purchased from Schleicher & Schuell, and T7 DNA polymerase (Sequenase) was purchased from U. S. Biochemical Corp. Oligonucleotide primers were synthesized on an Applied Biosystems Model 381A single-channel synthesizer (G. D’Angelo, Molecular Biology Core, SUNY/Stony Brook). Chinese hamster ovary cells containing chromosomal human hamster somatic cell hybrids were kindly supplied by Dr. T. C. Gilliam (Columbia University, New York, NY) and propagated in Dulbecco’s modified Eagle’s medium (Life Technologies, Inc.) supplemented with 10% fetal calf serum and 0.2 mM proline. HHW105 contains a single human chromosome 5 as its only human component, HHW213 contains a single human chromosome 5 lacking ~95% of the long arm of chromosome 5 with an intact 5p, and HHW1064 contains a single human chromosome 5 with a deletion.

*This work has been supported by grants from the American Heart Association, New York State Affiliate, and National Institutes of Health Grant HL02431. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked “advertisement” in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

¶Present address: Dept. of Microbiology and Molecular Genetics, UMDNJ Medical School, Newark, NJ 07103.

§To whom correspondence should be addressed: Division of Hematology, Health Sciences Center T15-040, State University of New York, Stony Brook, NY 11794-8151. Tel.: 516-444-2059; Fax: 516-444-7530; E-mail: WBahou@epo.som.sunysb.edu.

1 The abbreviations used are: TR, thrombin receptor; PAR-2, proteinase-activated receptor; kb, kilobase(s); PCR, polymerase chain reaction; HEL, human erythroleukemia; ON, oligonucleotide.
within the region 5q11.2 → 5q13.3 (19). Human umbilical vein endothelial cells were isolated from pooled primary cultures of human umbilical veins and propagated as described previously (20). HEL cells were propagated in RPMI medium supplemented with 10% fetal calf serum, penicillin (100 units/ml), and streptomycin (100 μg/ml).

Genomic Analysis and Library Screening—DNA from normal human volunteers or rhesus monkeys was extracted from peripheral blood leukocytes as described previously (20) and quantitated by absorption spectrophotometry at 260 nm. Approximately 5–10 μg of DNA were digested with various restriction enzymes for Southern blot analysis using cDNA probes directly labeled in low melting agarose with [α-32P]dCTP. Distinct thrombin receptor cDNA clones included a ~1.3-kb insert encompassing the open reading frame cloned into the BamHI site of pGEM (5'-open reading frame, nucleotides 222-1502) and the ~1.8-kb insert spanning the 3'-untranslated region (nucleotides 1503-3111) (9). Blots were washed to high (0.1 M NaCl, 0.1% SDS) or low (0.1 M NaCl, 0.5% SDS, 1 M EDTA, pH 8.0, 10 M NaCl sodium phosphate at 68 °C for 1 h) or low (0.1 M NaCl, 0.5% SDS, 1 M EDTA, pH 8.0, 10 M NaCl sodium phosphate at 55 °C for 30 min) stringency and analyzed by autoradiography with Kodak XAR-5 film with an intensifying screen at ~80 °C for 3–10 days. For some experiments, filters were stripped according to the manufacturer's recommendations and adequacy confirmed by overnight autoradiography.

A human genomic library cloned into the bacteriophage EMBL3 was kindly supplied by Dr. W. Schubach (SUNY at Stony Brook). Library screening was completed with the 32P-radiolabeled TR cDNA insert essentially as described previously (20), utilizing Escherichia coli host strain NM539. Positive phage clones were plaque-purified, and the DNA was purified from minilysates by standard methods (21). Alternatively, P1 genomic clones were obtained by PCR using oligonucleotide primers spanning the second exon (Genome Systems, Inc., St. Louis, MO). Genomic fragments were extensively characterized by end-ordered partial digestion and Southern blot analysis, and individual fragments were subcloned into pBluescript (Strategene, La Jolla, CA) or M13mp18 (Sigma) for sequence analysis using dideoxy chain termination (22). Exon-intron boundaries were defined by comparison of genomic DNA sequence with that of the published cDNA (8, 9). Sequence analysis was performed using the Wisconsin Genetics Computer Group Package (23).

RNA Preparation and PCR Analysis—Confluent human umbilical vein endothelial cells in the second to fifth passages were directly harvested with a rubber policeman, and total cellular RNA was isolated by immediate solubilization in guandine hydrochloride and serial ethanol precipitation (20). Reverse transcription was completed by incubating 10 μg of endothelial cell RNA with 1 μg of a 14-mer oligodeoxynucleotide primer at 41 °C for 1 h using 10 units of avian myeloblastosis virus reverse transcriptase (Seikagaku America, Inc.) in a solution containing 50 mM Tris/HCl, pH 8.3, 50 mM KCl, 8 mM MgCl2, 0.2 mM dNTPs, and 2.7 μg/ml amantadine D. Samples were then incubated with 500 ng of DNase-free RNase (Boehringer Mannheim) for 45 min at 37 °C in the presence of 16 μg/ml EDTA prior to phenol-chloroform extraction and precipitation using 2.5 M sodium acetate, pH 5.2, and ethanol. The primer extension product was then analyzed by acrylamide gel electrophoresis in parallel with sequencing reactions using the identical oligodeoxynucleotide primer and a ~3-kb genomic fragment cloned into the HindIII site of M13mp18, known to contain the first exon and 5'-untranslated region.

RESULTS AND DISCUSSION

Gene Localization Using Somatic Cell Hybrid—Previous work in this laboratory using fluorescent in situ hybridization of metaphase chromosomes localized the human thrombin receptor gene to the region 5q13 (24). To confirm this localization using molecular techniques and to exclude the possibility that this represented cross-hybridization to a pseudogene, genomic analysis was completed using a panel of somatic cell hybrids retaining distinct portions of chromosome 5 as their sole human components. As demonstrated in Fig. 1, Southern blot analysis using the thrombin receptor cDNA as probe confirmed the presence of a single cross-hybridizing fragment in total human genomic DNA and HHW105, which contains a single copy of an intact chromosome 5. No cross-hybridizing fragments were evident using DNA from Chinese hamster ovary cell lines HHW213, which lacks the majority of 5q, or HHW1064, which contains an intact chromosome 5 specifically deleted within the region q11.2 → 13.3 (19). These data are in full agreement with our initial gene localization studies and further demonstrate that this region of the cDNA is present as a single copy in the human genome, consistent with previous observations in the rat (11).

Thrombin Receptor Genomic Characterization—To further characterize the thrombin receptor gene, more extensive analysis was completed using human genomic DNA and two cDNA

![Fig. 1. Genomic analysis using human:hamster somatic cell hybrids.](http://www.jbc.org/

Downloaded from [jscholar.ucsf.edu](http://www.jscholar.ucsf.edu) by guest on June 24, 2017
probesspanningnucleotides222-3111. Initialevaluationusing
the 1280-base pair 5
9-open reading frame as probe demon-
strated the presence of few (1–2) cross-hybridizing fragments
with all restriction enzymes tested (Fig. 2
A). Five of the en-
zymes revealed single hybridizing fragments, suggesting a
gene of limited size and complexity. Furthermore, genomic
analysis using EcoRI-digested rhesus monkey DNA demon-
strated an identical pattern under high stringency wash con-
ditions, suggesting that this portion of the TR gene is highly
conserved and similarly organized in nonhuman primates. Hy-
bridization of the duplicate blot with the
1.6-kb 3
9-untrans-
lated region cDNA again demonstrated the presence of 1–2
hybridizing fragments with all restriction enzymes (Fig. 2
B).

Taken collectively, these data suggested that the thrombin
receptor gene is of limited size and complexity, a pattern that
has been previously described for a number of other seven-
transmembrane receptor genes. To date, the coding regions
of the three b-adrenergic receptors, the two a-adrenergic recep-
tors, the five muscarinic cholinergic receptors and the 5HT-1A
serotonin receptors have all been shown to be intronless, sug-
gesting the evolution from a common ancestral gene (7).

To determine if structurally related genes are present in
humans, individual filters were stripped, and Southern blot
analysis was repeated under low stringency conditions. No
novel cross-hybridizing fragments were demonstrable, inconsis-
tent with the presence in the human genome of a structur-
ally related pseudogene. Thus, although a second proteolyti-
cally cleaved receptor has been recently identified (18), and the
presence of other thrombin receptors has been postulated,
these data confirm that they are not highly homologous to the
TR. Indeed, the murine putative proteinase-activated receptor
(PAR-2) displays only 28% identity to the murine and 30%
identity to the human thrombin receptor at the protein level,
although certain regions within the transmembrane and extra-
cellular loops appear more highly conserved (18).

The initial characterization of the gene was then confirmed
by isolating genomic clones encompassing the TR. Approxi-
mately 1 × 10
6 plaques were screened from a human genomic
bacteriophage λ library using the TR cDNA as probe with the
isolation of a single ~18-kb genomic clone (λ11A-1). Southern
analysis confirmed that this fragment contained the majority of

FIG. 3. Schematic diagram displaying the structural organiza-
tion of the thrombin receptor gene. Exons are indicated by solid
rectangles. Relevant restriction endonucleases utilized for genomic
mapping are indicated: H, HindIII; S, SalI; E, EcoRV; P, PvuII. Lambda
phage and P1 clones with approximate ends are depicted.

To more precisely characterize the TR genomic organization,
we initially employed a comparative PCR strategy using total
genomic DNA or reverse-transcribed endothelial cell RNA as
templates. Distinct oligonucleotide primer pairs spanning the
full-length cDNA effectively amplified the identically sized
fragments from base pair 490 to the 3′-end of the cDNA (data
not shown). We were unable, however, to amplify the remain-
der of the 5′-sequence using total genomic DNA, suggesting the
presence of a large intron upstream of this region.

The initial characterization of the gene was then confirmed
by isolating genomic clones encompassing the TR. Approxima-
tely 1 × 10
6 plaques were screened from a human genomic
bacteriophage λ library using the TR cDNA as probe with the
isolation of a single ~18-kb genomic clone (λ11A-1). Southern
analysis confirmed that this fragment contained the majority of
the coding sequence, although it lacked the 5'-exons separated by a large intronic splice junction encompassing the TR coding sequence. Although it lacked the 5'-exons separated by a large intronic splice junction encompassing the TR coding sequence, the coding sequence proved to be identical to the published cDNA, except for the presence of a CG inversion at nucleotides 935–36 (CG \rightarrow GC, Leu^{237}, unchanged; Val^{238} to Leu^{239}), as described previously in the endothelial cell TR cDNA homologue (9).

Determination of the Transcription Initiation Site—Primer extension analysis was used to determine the thrombin receptor transcription initiation site using two 30-mer oligonucleotide primers (ON 1715 and ON 1716) known to be contained within the first exon, as elucidated by studies outlined above. Primer extension analysis was completed with cellular RNA from both human umbilical vein endothelial cells and HEL cells, cell lines known to express a functional thrombin receptor. Both these lines contained a single ~3.5-kb hybridizing TR transcript as demonstrated by Northern blot analysis (not shown). As shown in Fig. 4, the identical primer extension product was demonstrable using total cellular RNA or poly(A)-mRNA from either cell line. No primer extension product was seen using transfer RNA as a control (not shown), confirming the presence in both these cell types of a single predominant RNA transcript. Analysis of the TR primer extension product in parallel with a sequencing reaction using the identical primer and the ~3-kb HindII genomic fragment (Fig. 3) identified the predominant transcription start site to a guanine nucleotide 351 base pairs upstream of the initiator methionine. This site conforms to a canonical splice acceptor site (26), allowing us to empirically define the 5'-border of the first exon. These results were confirmed using a second oligonucleotide (ON 1716), again demonstrating the presence of a primer extension product at the same site (not shown).

Sequence Analysis of the TR 5'-flanking Region—A ~3-kb HindII fragment known to contain the first exon and a portion of the 5'-flanking region (see Fig. 3) was sequenced on both strands to further characterize potential regulatory sequences involved in TR expression (Fig. 5). A region typical of an Alu J-subfamily of short interspersed repetitive sequences was identified and is displayed in Fig. 5A. Alu sequences represent approximately 5–6% of the total human genome (27) and appear approximately every 5–8 kb in human genomic DNA. Interestingly, similar Alu sequences have been identified within the promoter regions for other genes including transforming growth factor-\(\alpha\) (28) and the integrin \(\beta_3\) subunit (29), although a potential role in transcriptional gene regulation remains undetermined. Neither a TATA box nor CCAAT sequences (30) are evident adjacent to the transcription initiation site. Although a canonical TATA recognition sequence (TATA-AAA) is evident at nucleotide positions \(-431 \rightarrow -425\), its relevance in mediating gene transcription is currently unresolved, although functional TATA boxes at this distance from the transcription site are unusual in eukaryotic genes. Further sequence inspection of the 5'-regulatory region revealed potential cis-acting DNA elements, including Sp1 binding sites (GGGCGG) (31) present at nucleotides \(-1008\) (inverted), \(-579,
The potential role of either of these two latter sequences (GATA motifs, octamer sequences) in regulating cell-specific expression remain uncharacterized, however, because thrombin receptor expression is evident in a wide number of cell types in addition to megakaryocytes (and platelets).

The seven-transmembrane segment thrombin receptor represents the prototype of a novel class of proteolytically cleaved receptors that mediate signaling events by functional coupling to G-proteins. The identification of PAR-2 reinforced the concept that circulating proteases (in addition to α-thrombin) may affect cellular events through such proteolytically cleaved receptors. Although a physiological enzyme substrate for PAR-2 has not been identified, preliminary observations from other laboratories suggest that both receptors display similar activation mechanisms. The data presented in this manuscript demonstrate that these functional properties also extend to the structural organization of the genes. Both genes contain two exons separated by a large intron, both genes encode identical numbers of amino acids within the first exon, and the cleavage sites for both gene products are similarly contained within the larger second exon. Thus, although the proposed gene family is currently limited to two family members, we would speculate that other similarly organized genes are present in humans, presumably evolving from a common ancestral gene.

Addendum—Since the submission of this manuscript, the human homologue of the PAR-2 gene has been isolated and characterized (36). Like the TR and murine PAR-2, the human PAR-2 gene has essentially the identical genomic organization. Interestingly, human PAR-2 co-localizes with the human TR gene at 5q13.

Acknowledgments—We thank Cheri Potter and Andrea Wong for assistance with some of these experiments and Shirley Murray for help with the preparation of this manuscript.

Fig. 5. Analysis of the thrombin receptor gene 5′-flanking sequence. A, sequence analysis of the 5′-regulatory region is displayed with relevant restriction sites indicated. The Alu-like repeats are underlined, the transcription initiation site is delineated by the arrow, the start of thrombin receptor cDNA done A+1 (9) is depicted by the black diamond, and the start of the original published cDNA sequence (8) is represented by the star. The location of primer ON 1715 used for primer extension analysis in Fig. 4 is depicted by the thick black line above the sequence. B, schema summarizing the putative transcriptional regulatory sequences identified within the 5′-regulatory region. This sequence has been deposited into the GenBank data base and assigned the accession number U35634.

and 1 and AP-2-like elements (GSSGWGSC and YCSCC-MNSSS)2 (32) at nucleotides −965, −822, −552, −254, and −242. Potential recognition sequences for the Ets family of transcription factors (SMGGAWGY) (33) are evident at −1478, −564, −536, −152, and −67 (PU.1 element), and TEF-1 elements are evident at −214 and −179 (34). The thrombin receptor 5′-flanking region also contains sequences known to represent binding sites for the erythroid nuclear factor protein NF-E1 (WGATAMS) at nucleotides −1309 and −1241 (invert.

2 The abbreviations used in nucleic acid motifs are as follows: M represents A/C, R is A/G, W is A/T, S is C/G, Y is C/T, and N refers to A/C/G/T.


Genomic Cloning and Characterization of the Human Thrombin Receptor Gene: STRUCTURAL SIMILARITY TO THE PROTEINASE ACTIVATED RECEPTOR-2 GENE
Valentina A. Schmidt, Emilia Vitale and Wadie F. Bahou

doi: 10.1074/jbc.271.16.9307

Access the most updated version of this article at http://www.jbc.org/content/271/16/9307

Alerts:
- When this article is cited
- When a correction for this article is posted

Click here to choose from all of JBC's e-mail alerts

This article cites 35 references, 20 of which can be accessed free at http://www.jbc.org/content/271/16/9307.full.html#ref-list-1