Proc. Natl. Acad. Sci. USA
Vol. 95, pp. 6642–6646, June 1998
Biochemistry

Cloning and characterization of human protease-activated receptor 4

WEN-FENG XU*†, HENRIK ANDERSEN*, THEODORE E. WHITMORE†, SCOTT R. PRESNELL†, DAVID P. YEE†, ANDREW CHING†, TERESA GILBERT†, EARL W. DAVIE*‡, AND DONALD C. FOSTER†

*Department of Biochemistry, University of Washington, Box 357350, Seattle, WA 98195-7350; and †ZymoGenetics Inc., 1201 Eastlake Avenue East, Seattle, WA 98102

Contributed by Earl W. Davie, April 2, 1998

ABSTRACT **Protease-activated receptors 1–3 (PAR1, PAR2, and PAR3) are members of a unique G protein-coupled receptor family. They are characterized by a tethered peptide ligand at the extracellular amino terminus that is generated by minor proteolysis. A partial cDNA sequence of a fourth member of this family (PAR4) was identified in an expressed sequence tag database, and the full-length cDNA clone has been isolated from a lymphoma Daudi cell cDNA library. The ORF codes for a seven transmembrane domain protein of 385 amino acids with 33% amino acid sequence identity with PAR1, PAR2, and PAR3. A putative protease cleavage site (Arg-47/Gly-48) was identified within the extracellular amino terminus. COS cells transiently transfected with PAR4 resulted in the formation of intracellular inositol triphosphate when treated with either thrombin or trypsin. A PAR4 mutant in which the Arg-47 was replaced with Ala did not respond to thrombin or trypsin. A hexapeptide (GYPGQV) representing the newly exposed tethered ligand from the amino terminus of PAR4 after proteolysis by thrombin activated COS cells transfected with either wild-type or the mutant PAR4. Northern blot showed that PAR4 mRNA was expressed in a number of human tissues, with high levels being present in lung, pancreas, thyroid, testis, and small intestine. By fluorescence *in situ* hybridization, the human PAR4 gene was mapped to chromosome 19p12.**

Elucidation of the mechanisms by which proteases activate cells has been an intriguing question in cell biology (1). In recent years, a subfamily of G protein-coupled receptors capable of mediating cellular signaling in response to proteases has been identified (2–5). The first member of this family was the thrombin receptor presently designated protease-activated receptor 1 (PAR1). Thrombin cleaves an amino-terminal extracellular extension of PAR1 to create a new amino terminus that functions as a tethered ligand and intramolecularly activates the receptor (2). Knockout of the gene coding for PAR1 provided definitive evidence for a second thrombin receptor in mouse platelets and for tissue-specific roles for different thrombin receptors (6). PAR2 can mediate signaling after minor proteolysis by trypsin or tryptase but not thrombin (4). PAR3 was identified recently as a second thrombin receptor that can mediate phosphatidylinositol 4,5-bisphosphate hydrolysis; it was expressed in a variety of tissues (5). Because many other proteases such as factor VIIa (7), factor Xa (8), factor XIIa (9), protein C (10), neutrophil cathepsin G (11), mast cell tryptase (12), and plasmin (13) display cellular effects, it has been speculated that additional members of the PAR family exist (14, 15).

In this report, we describe the cloning of a human cDNA sequence coding for a fourth protease-activated receptor (PAR4). When transiently expressed in COS cells, PAR4 was activated by thrombin or trypsin. The extracellular amino-terminal extension of PAR4 contained a putative serine protease cleavage site, and its importance in receptor activation was demonstrated by mutagenesis experiments. The tissue distribution and chromosome localization are also described.

MATERIALS AND METHODS

Purified α-thrombin, trypsin, and other proteases were kindly provided by K. Fujikawa (University of Washington); human γ-thrombin was purchased from Enzyme Research Laboratories, South Bend, IN. The size-selected lymphoma Daudi cell cDNA library was a gift from S. Lok (ZymoGenetics). Peptides were synthesized and purified by HPLC before use. Nylon membranes and radioactive isotopes were purchased from Amersham. All cell culture media and supplements were from GIBCO/BRL.

A search of the public DNA sequence database dbEST and commercially available Incyte (Incyte Pharmaceuticals, Palo Alto, CA) expressed sequence tag (EST) LifeSeq database was carried out by using a modified version of the FASTA program (16). Human PAR1, PAR2, and PAR3 sequences were used as the query sequences.

Screening of the cDNA library was carried out by standard filter hybridization techniques with radioactive DNA probes labeled by random priming (Prime-it kit, Stratagene). cDNA inserts were sequenced on both strands by the dideoxynucleotide chain-termination method (17) using the Sequenase kit from United States Biochemicals. The cDNA used for the epitope-tagged PAR4 assay was prepared analogous to Flag-epitope-tagged PAR1 with an amino terminus sequence of MDSKGSSQKGSRLLLLLVVSNLLLCQGVVS ↓ DYKDDDDKLE-GG. This represents the bovine prolactin signal peptide, the putative signal peptidase site (↓), the Flag epitope DYKDDDDK, and a junction of LE providing a *Xho*I cloning site (5). This sequence was fused to Gly-18 in PAR4. Receptor cDNAs were subcloned into the mammalian expression vector pZP-7 provided by D. Prunkard (ZymoGenetics). Receptor expression on the COS cell surface was measured as specific binding of mAb M1 (IBI-Kodak) to the FLAG epitope at the receptor's amino terminus (18).

For the phosphoinositide hydrolysis assay, COS-7 cells were grown in DMEM with 10% fetal bovine serum (FBS). Cells were plated at 3.5×10^5/35-mm plate 1 day before transfection. Two micrograms of DNA were transfected with 12 μl of LipofectAMINE (GIBCO/BRL) for 5 hr. The cells were incubated overnight in DMEM with 10% FBS and then split

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviations: EST, expressed sequence tag; PAR1–4, protease-activated receptors 1–4, respectively.

Data deposition: The sequence reported in this paper has been deposited in the GenBank database (accession no. AF055917).

‡To whom reprint requests should be addressed.

into triplicate 35-mm wells. Forty-eight hours after transfection, the cells were loaded with *myo*-[^{3}H]inositol (2 μCi/ml; 1 Ci = 37 GBq) in serum-free DMEM and incubated overnight at 37°C. Cells were washed and treated with 20 mM LiCl in DMEM with or without protease or peptide activators added at various concentrations. Cells were then incubated for 2 hr at 37°C and extracted with 750 μl of 20 mM formic acid for 30 min on ice. The inositol mono-, bis-, and trisphosphates were purified through a 1-ml AG 1-X8 anion-exchange resin (Bio-Rad) (19) and radioactivity was quantitated by scintillation counting. In each hydrolysis assay, surface expression levels of receptors were determined in triplicate in parallel cultures.

For Northern blot analyses, three human multiple-tissue blots with 2 μg of mRNA per lane (CLONTECH) were hybridized with a ^{32}P-labeled 166-bp PAR product generated from human lymph node cDNA with PCR4-specific primers, 5′-TGGCACTGCCCCTGACACTGCA-3′ and 5′-CCCGTAGCACAGCAGCATGG-3′. Hybridization to human β-actin mRNA was used as a control for variation in abundance. The blots were hybridized overnight in ExpressHyb (CLONTECH) and washed at 50°C in 0.1× SSC/0.1% SDS followed by exposure to x-ray film.

The cDNA coding for PAR1 was isolated from a placental cDNA library by PCR. The DNA sequence was essentially identical to that reported (2) except for nucleotides 935-936 (CG → GC) and nucleotides 1315-1316 (CG → GC). These differences resulted in a change of Val-238 → Leu and a change of Ser-364 → Cys, respectively. These changes were confirmed by sequence analysis of the corresponding regions in the genomic DNA coding for PAR1.

The Human Genetic Mutant Cell Repository Human/Rodent Somatic Cell Hybrid Mapping Panel Number 2 (National Institute of General Medical Sciences, Coriell Institute of Medical Research) was used with PCR amplification to identify the somatic hybrid that contained the human PAR4 gene (20). PAR4-specific oligonucleotide primers (sense, 5′-GGTGCCCGCCCTCTATGG-3′, and antisense, 5′-TCGCGAGGTTCATCAGCA-3′) were used for the PCR amplification.

Subchromosomal mapping of the PAR4 gene was carried out by using the commercially available version of the Stanford G3 Radiation Hybrid Mapping Panel (Research Genetics, Huntsville, AL). The Stanford G3 RH Panel contains PCR-

```
1                                                                                 CTCCC ACGGGCTGGC TGGCAAGCGG CCCTGGTGGG TCTGCGGGGG CAGGGGCAGC
56   CTTCCTGGTT TATCTCCACC GGCGCGATCT GCTCGTCCGC CTCGGCTCCA GAAGCTGGGG CTCAGGGTCC GGCGAGGCAG GAAGCCTGAG GCCACAGCCC AGAGCAGCCT GAGTGCAGTC

1     M   W   G   R   L   L   L   W   P   L   V   L   G   F   S   L   S   G   G   T   Q   T   P   S   V   Y   D   E   S   G   S   T   G
176  ATG TGG GGG CGA CTG CTC CTG TGG CCC CTG GTG CTG GGG TTC AGC CTG TCT GGC GGC ACC CAG ACC CCC AGC GTC TAC GAC GAG AGC GGG AGC ACC GGA

34    G   G   D   D   S   T   P   S   I   L   P   A   P   R   G   Y   P   G   Q   V   C   A   N   D   S   D   T   L   E   L   P   D   S
275  GGT GGT GAT GAC AGC ACG CCC TCA ATC CTG CCT GCC CCC CGC GGC TAC CCA GGC CAA GTC TGT GCC AAT GAC AGT GAC ACC CTG GAG CTC CCG GAC AGC

67    S   R   A   L   L   L   G   W   V   P   T   R   L   V   P   A   L   Y   G   L   V   L   V   V   G   L   P   A   N   G   L   A   L
374  TCA CGG GCA CTG CTT CTG GGC TGG GTG CCC ACC AGG CTG GTG CCC GCC CTC TAT GGG CTG GTC CTG GTG GTG GGG CTG CCG GCC AAT GGG CTG GCG CTG

100   W   V   L   A   T   Q   A   P   R   L   P   S   T   M   L   L   M   N   L   A   T   A   D   L   L   L   A   L   A   L   P   P   R
473  TGG GTG CTG GCC ACG CAG GCA CCT CGG CTG CCC TCC ACC ATG CTG CTG ATG AAC CTC GCG ACT GCT GAC CTC CTG CTG GCC CTG GCG CTG CCC CCG CGG

133   I   A   Y   H   L   R   G   Q   R   W   P   F   G   E   A   A   C   R   L   A   T   A   A   L   Y   G   H   M   Y   G   S   V   L
572  ATC GCC TAC CAC CTG CGT GGC CAG CGC TGG CCC TTC GGG GAG GCC GCC TGC CGC CTG GCC ACG GCC GCA CTC TAT GGT CAC ATG TAT GGC TCA GTG CTG

166   L   L   A   A   V   S   L   D   R   Y   L   A   L   V   H   P   L   R   A   R   A   L   R   G   R   R   L   A   L   G   L   C   M
671  CTG CTG GCC GCC GTC AGC CTG GAT CGC TAC CTG GCC CTG GTG CAC CCG CTG CGG GCC CGC GCC CTG CGT GGC CGG CGC CTG GCC CTT GGA CTC TGC ATG

199   A   A   W   L   M   A   A   A   L   A   L   P   L   T   L   Q   R   Q   T   F   R   L   A   R   S   D   R   V   L   C   H   D   A
770  GCT GCT TGG CTC ATG GCG GCC GCC CTG GCA CTG CCC CTG ACA CTG CAG CGG CAG ACC TTC CGG CTG GCG CGC TCC GAT CGC GTG CTC TGC CAT GAC GCG

232   L   P   L   D   A   Q   A   S   H   W   Q   P   A   F   T   C   L   A   L   L   G   C   F   L   P   L   L   A   M   L   L   C   Y
869  CTG CCC CTG GAC GCA CAG GCC TCC CAC TGG CAA CCG GCC TTC ACC TGC CTG GCG CTG TTG GGC TGT TTC CTG CCC CTG CTG GCC ATG CTG CTG TGC TAC

265   G   A   T   L   H   T   L   A   A   S   G   R   R   Y   G   H   A   L   R   L   T   A   V   V   L   A   S   A   V   A   F   F   V
968  GGG GCC ACC CTG CAC ACG CTG GCG GCC AGC GGC CGG CGC TAC GGC CAC GCG CTG AGG CTG ACC GCA GTG GTG CTG GCC TCC GCC GTG GCC TTC TTC GTG

298   P   S   N   L   L   L   L   L   H   Y   S   D   P   S   P   S   A   W   G   N   L   Y   G   A   Y   V   P   S   L   A   L   S   T
1067 CCC AGC AAC CTG CTG CTG CTG CTG CAT TAC TCG GAC CCG AGC CCC AGC GCC TGG GGC AAC CTC TAT GGT GCC TAC GTG CCC AGC CTG GCG CTG AGC ACC

331   L   N   S   C   V   D   P   F   I   Y   Y   Y   V   S   A   E   F   R   D   K   V   R   A   G   L   F   Q   R   S   P   G   D   T
1166 CTC AAC AGC TGC GTG GAT CCC TTC ATC TAC TAC TAC GTG TCG GCC GAG TTC AGG GAC AAG GTG CGG GCA GGG CTC TTC CAA CGG TCG CCG GGG GAC ACC

364   V   A   S   K   A   S   A   E   G   G   S   R   G   M   G   T   H   S   S   L   L   Q   *
1265 GTG GCC TCC AAG GCC TCT GCG GAA GGG GGC AGC CGG GGC ATG GGC ACC CAC TCC TCT TTG CTC CAG TGA CACAAA GTGGGGAAGG CTGTACTGGG TCGAACAGGG

1370 TCCCTTCCCC CACTTCACGT CCTTCCTGGG ACCTCAGAAT GTGACCTTAT TTGGAAATAG GGTTGTTACA ACTGTCACTA GCGGAGGTCA CTTTGGAGAA GGGTGGGCCT TACATCCAGT
1490 GTGGGTGGTG TCCTCATAAG ATAAGGAGAG GCCAGGCCTG GTGGCTCACG CCTGTAATCC CAGCACTTTA AGAGGCCAAG GCGGATGGAT CACTTGAGCC CAGGAGTTCA ACACCAGCCT
1610 GAGCAACATG GTAAAACCCC ATCTCTACCA AAAATACAAA AATTAGCTGG GCTTGGTGGC TGGCGCCTGT AATCCCAGCT ACTCAGGAGA CTGAGGCAGA AGGATCGCTT GAACCTGGGA
1730 GGCAGAGGTT GCAGTGAGCC GAGATTGCGC CACTGGACTC CAGCCTGCGT GACAGAGAGC CTGTCTCTAA ATTAATTAAT TAATTAATTT AATTCAATTT TAAAAAGACG AAAAGTGACG
1850 GCCAGGTGCA GTGGCTCACG CCTATAATCT CAGCACTCTG GGAGGCCAAG ATGGAGGATT GCTTGAAGCC AGGAGTTTGG GACCAGCCTG GGCAACATAG GGGGATCCCA TCTCTACACA
1970 CAAAAAAATT TTTTAATGAA CCAGGCATTG TGGCATGCGC CTATAGTCCC AGCCACTCAA GAGGCACAGG CGGGAGGATC ACTTGAGCCT GGGAGGTTGT GGTTGCAGTG AGCTATGATT
2090 GTACCACTGC ACTCCAGCCT GGGCAACAGA GCAAGACCTT GTCTCAAAAA TAAACAAACT AAAATTAAAA AAAGAAGACG AGAGATAGTG GGTGTGGTGG CTCACACCTG CAATCCCAGC
2210 ACTTTGGAAG GCCGAGGTGG GCAGATCATC TGAGGCCAGG AGTTCAAGAC CAGCCTGGCT AACATGGTGA AATCCTATCT CTACCAAAAA TACAAAAATT AGCCAGGCGT GGTGGTGGGC
2330 ACCTGTACTG GGGAGGTGCC CACCCAGCTA CTGGGGAGGC TGAGTCAGGA GAATCGCTTG AACCTGGGAG GCGGAGGTTG CGGTCAGCTG AGATGGTGCC ACTGCACTCC AGCCTGGGCG
2450 AAAGAGCGAC TCTGTCTCCA AAAAAAAGAG AAGAGGAGAG GACACAGAGA CACACAGAGA AGAAAGCCAT GTGGCGGCAG AGGCAGAGAT GGGAGTGATG CGGACGGACA CAAACTAAGG
2570 GATGCCACGA TGCCAAGCAC AGCCAACAGC CACCAGCAGC CAGGAGACAG GCCTGGGACG GGCTCTCCCT CACAGCCTCC AGAGGGAACC AGCCCTGCCA CCACCTTGAC CCTGGACTTC
2690 TGGCCTGCAG AACTGTGAGA CAATAAACTC TCATTGTTTT AAGCTGCCTG GCATGTGGCA CTTTGTCAGG GCAGCCCAGG AATCTGAAAC AGGATCAAAC TCTGCTTCCT GGGCCCTGCC
2810 AGCATCTCTG GCTCGGCTTT CTGGGCTGGA TGCAGCCCAC GACGCACTGG TGTCTGAGAT GGGGCTGGAG CTGGGGCTGG GGCTGCATTC CCTGGAGACT CACTGCAAGT TCCTGCCCAG
2930 GAGGCTGAGG GCACCCCATC CTCAGTGCCC AATGCTGTGG CCCCACCAGG CCCAGAGCCT GGTTGGCCAT TCTCATGCCC ACCAGCTTCT GGCTTTGGGA TGTCTCTTGA GCAACCAGAA
3050 TAGCACCCCC AACTCTGCTC CCCAAAACCC ATCACTAGCA CGGCTCAGCC TCCTGCTATC CCCTGACTGC TGGGGACCCT CGCCTTCCCT CCTCTCACCT GCAGGCTGAT CCTTCTTTTC
3170 ACTTTCTGTC AATGTCACCA GGGATAAGGT GGGACAATGG GGGGTGGGGG TGGACAGTGT GTGCTGGGGG GTTCGGGTGC TGCAGACCTG GAACTCCCTT CTGCCAGGAT GTTGGCAGCC
3290 GGTTGTAAGC CTTGCACGGG ACAGACCACA CCCACCGCAA CCTCATCCCC TCAGCACTAA CCACATCCAC TCTCAACCCC GTCCCCTTCG CACTGACCAC ACCCACCCCG TTCGGCCCCG
3310 CCCCCCGCAC TGAACACTCC CGCCCTCAAC CCCGCACCCT CCGCACTCAC CTCCCCCTCG CCGCTCGACC CCGCCCTCAC CACACTGACC ACCCTCAACC CATTGCGCCC AGTCCCCACC
3430 ACAGTGACCA CACCCTCACT GGCTCGGCCC TGCCCCCAGT ATACTGACCA TTCCCCAGCC ACTTCCCTTC CGCACTTACC ACTCCCCCAG CCACGCCCCT CCCCGCTGAC CGCTCCTCCA
3550 GCCCCGCCTC CCCCGTACAG GCAGAGCGCC CGCCCACCTC TATGCTGCGT TCTCCTGACT TTACGTTGGC CCCTCCTCTG CCAAGCCCCC AGGGGAGCCC TCCCTGGCGT CCGAGGGTGG
3670 GAGTCGGGGT GTGGCAGGCC GCGGTGGGGG GCGGCAGTGG CTCCGCGCAC TCACCCGGGC CCCGGGCAGG GGCGCGCTCC ACTTCGTTGC ACGCGGGTCC GGCGCACAGT TCCCGGGCGA
3790 GTGGGCTGTG CGTGCTGACG TTGTAGAAGC GAGTGGCCTC GAAGGCTACG GGACGAGGGT GGCGGGTGAC CAAGTGCAGG CGCGACGGGT CAGGGACCGG GCCGGGCCGG GGGTGCGGGC
3910 GCGCGGGCCT ACCGGGTTCG TAGTAGTCGT ACACGGAGAC TGGCAGCGCC GACGTCCTGC CCACCACGCA CTCCCGGAGA GCACGGAACC GCACGCACGT CAGGCACCGG CTGGGGATCT
4130 GTGGGGCAGC GGCGGGCGCA GGCTCGACCC GGGCCAGGAG GCCCGGGGCG CTGAGCTCAG GCCCAGAACT GGCTGATTTC AGGGATACCC AGGACGCGTG AAACACAGAA GAAACGTGAT
4250 CCCATTTTCT TTTTTTCTTT TACTTTTCTT TTTTTTTTTT TTTCCTGAGA CAGAGTCTCG CGCTGTTGCC CAGGCTGGAG TGCAGTGGCG TGATCTCGGC TCACTGCAAG CTCGGCCTCC
4370 TGGGTTCAAA TGATTCTCCT GCCTCAGCCT CCCAAGTAGC TGGGATAACA GGCGCCCACC ACCGCACCCT GCTAATTTTT TGTATTTTTG ATCAAGACGG AGTTTCACCA TGTTGGCCAG
4490 GCTGGTCTCC AACTCCTGCC CTCAAGTGAT CCGCCTCGGT CCCATTTTTT ATTCTTTGGG TCCTTCCATC CCACTGGGAA AACGTCTCAG GTGGCCTCTG AAACACCACT CCTTTTTGTG
4610 TGTGTGCACG CATGGCTGAG CATGTGTGGG TGGGAGTCAG CACATTCACG ATACTGTGCA ATCATCACCT CTGTCTAGTT ACAGGACGGT TTCTTTCTCC CCCAAAGAAA CCCCATCGCC
4730 ATCAGCACTC ACTCCCCACT CCCCCAGCCC CTGGCAACCA CAAATCTTTC CAACTCTACG GATTTGCCTG TTCTGGGCAT TTCATGTCAA TGGAATCATG TACTCTGTGA AAAAAAAAAA
4850 AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAA
```

FIG. 1. Nucleotide and deduced amino acid sequences for human PAR4. The nucleotide sequence of the 4.9-kb PAR4 cDNA was determined. The amino acid sequence encoded by the longest ORF is shown. Consensus polyadenylation signals are underlined.

amplifiable DNAs from each of 83 radiation hybrid clones of the whole human genome, plus two control DNAs (the RM donor and the A3 recipient). A publicly available WWW server (http://shgc-www.stanford.edu) permitted chromosomal localization of markers. The PCR amplification with the same set of primers were set up in a 96-well microtiter plate and used in a RoboCycler gradient 96 thermal cycler (Stratagene). The PCR products were separated by electrophoresis on a 2% agarose gel.

RESULTS AND DISCUSSION

Cloning of PAR4 cDNA. An amino acid query sequence derived from the known receptors PAR1, PAR2, and PAR3 was used for searching various databases to identify ESTs with homology to these sequences. One Incyte EST sequence (INC373881) was identified that matched the three protease-activated receptor sequences starting in the fourth transmembrane domain (nucleotides 770-2140, Fig. 1). When this DNA sequence was translated into protein, the amino acid sequence shared 34% identity with PAR2 in the transmembrane region. A size-selected lymphoma Daudi cell line cDNA library with inserts greater than 2 kb was then screened with a 600-bp DNA probe from the EST sequence. A full-length cDNA clone (4.9 kb) was identified, sequenced on both strands, and designated as protease-activated receptor 4 (PAR4) (Fig. 1). The DNA sequence revealed a 5′ untranslated region (nucleotides 1–175), an ORF encoding a 385-amino acid protein (nucleotides 176-1333), and a long G+C-rich 3′ untranslated region containing several polyadenylation signals and a poly(A) tail (nucleotides 1334–4895).

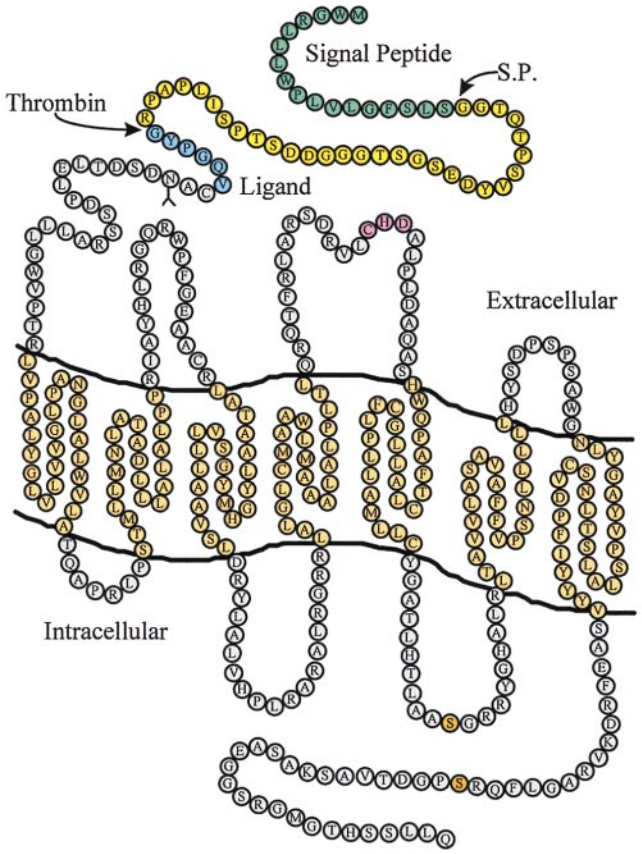

FIG. 2. Diagram illustrating the proposed seven transmembrane-domain organization for PAR4. The signal peptide is shown in green; the amino-terminal peptide cleaved by thrombin is in yellow; the tethered peptide ligand is in blue; the seven transmembrane-domain regions are in gold; the remaining extracellular and intracellular regions are shown in gray. The CHD sequence in the second transmembrane loop that is present in the four known PAR proteins is shown in pink. A potential serine phosphorylation site for CK II in the sequence SGR and a potential phosphorylation site for protein kinase II in the sequence SPGD are shown in orange (21). An attached Y refers to a potential N-linked glycosylation site, and S.P. refers to signal peptidase.

Table 1. Protease cleavage sites in PAR1, PAR2, PAR3, and PAR4

Peptide	Sequence
PAR1(37–61)	TLDPR ↓ SFLLRNPNDKYEPFWEDEEK
PAR2(32–56)	SSKGR ↓ SLIGKVDGTSHVTGKGVTVE
PAR3(34–57)	TLPIK ↓ TFRGAPPN SFEEFPFSALE
PAR4(28–52)	LPAPR ↓ GYPGQVCANDSDTLELPDSS
Hirudin	---DFEEI---

Regions important for fibrinogen anion exosite ginding in thrombin are underlined. ↓, Cleavage site.

A hydropathy plot of the amino acid sequence of PAR4 revealed that the receptor was a member of the seven transmembrane-domain receptor family as illustrated in Fig. 2. A hydrophobic signal sequence with a potential signal peptidase cleavage site was present at Ser-17/Gly-18. A putative cleavage site at Arg-47/Gly-48 for protease activation was also present within the extracellular amino terminus. Alignment of the PAR4 amino acid sequence with the three other known protease-activated receptors (22) also indicated that PAR4 was a member of the protease-activated receptor family with about 33% overall amino acid sequence identity with PAR1, PAR2, and PAR3. The extracellular amino terminus and intracellular carboxyl terminus of PAR4, however, have little or no amino acid sequence similarity to the corresponding regions in the other family members. Furthermore, the thrombin cleavage site in PAR4 differs substantially from that in PAR1 and PAR3 (Table 1). Also, in the second extracellular loop, PAR4 has only three amino acids (CHD) that matched the sequence of ITTCHDV that is conserved in PAR1, PAR2, and PAR3. The second extracellular loop was critical in determining the specificity of PAR1 from human and *Xenopus laevis* sources for their respective activating peptides (23).

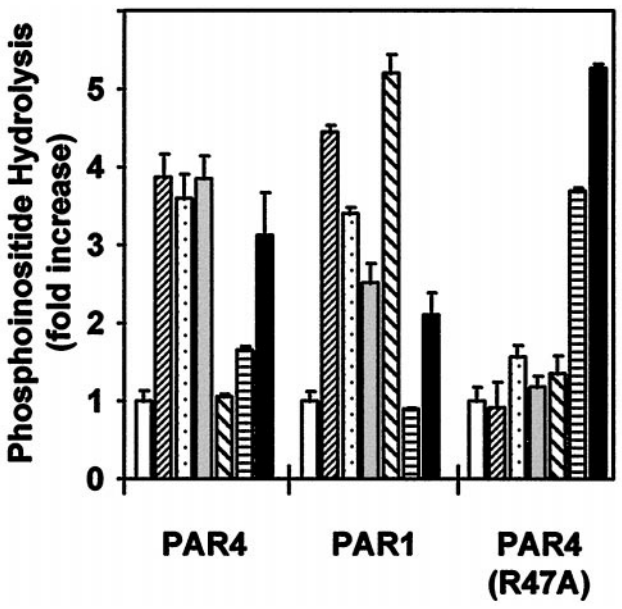

FIG. 3. Agonist activity of thrombin, trypsin, and activating peptides on COS cells expressing PAR1, PAR4, or PAR4 protease cleavage site mutant (R47A). Bars: open, control; upward hatched, thrombin (100 nM); stippled, γ-thrombin (100 nM); shaded, trypsin (100 nM); downward hatched, PAR1 activation peptide (100 μM); horizontally hatched, PAR4 activation peptide (100 μM); solid, PAR4 activation peptide (500 μM).

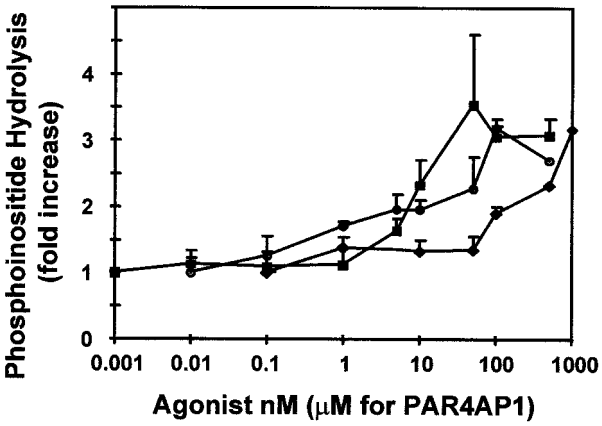

FIG. 4. Dose-dependent response of various agonists on PAR4. Concentration of proteases is presented as nanomolar, whereas peptide concentrations are presented as micromolar. EC_{50} values are estimated at 5 nM for thrombin and trypsin and about 100 μM for PAR4-activating peptide. ○, Thrombin; ■, trypsin; ◆, PAR4 activation peptide.

Activation of PAR4 by Thrombin and Trypsin. The similarity in sequence between PAR4 and the other protease-activated receptors suggested that PAR4 could be activated by an arginine-specific serine protease. COS cells transiently transfected with PAR4 cDNA responded to thrombin or trypsin addition (100 nM), resulting in phosphatidylinositol 4,5-bisphosphate hydrolysis. This was comparable to the thrombin-stimulated activation of PAR1 (Fig. 3). γ-Thrombin, which lacks a fibrinogen-binding exosite (24), was as effective as α-thrombin in the activation of PAR4 (Fig. 3). This is in contrast to the activation of PAR1 and PAR3 where γ-thrombin is much less potent than α-thrombin (1, 4). This is probably due to the presence of the thrombin binding site within the amino-terminal region of PAR1 and PAR3 (25–27). The thrombin-stimulated phosphoinositide hydrolysis with PAR4 was dose-dependent with a half-maximal concentration (EC_{50}) for thrombin and trypsin of 5 nM (Fig. 4). This was much higher than that for PAR1 and PAR3 (about 0.2 nM) (2, 5).

Other arginine/lysine-specific serine proteases including factors VIIa, IXa, and XIa; urokinase; or plasmin had little or no activity against PAR4. Small effects, however, were observed with factor Xa at high nonphysiological concentrations (100 nM). Chymotrypsin and elastase failed to activate PAR4 (data not shown).

Site-directed mutagenesis was then used to evaluate the importance of the putative cleavage site at Arg-47/Gly-48 in PAR4 activation. A cDNA coding for PAR4 with a single amino acid substitution, Arg-47 → Ala, was transiently expressed in COS cells. The putative cleavage site mutant (R47A) failed to respond to either thrombin or trypsin (Fig. 3). In contrast, a mutation of Arg-68 in the extracellular amino-terminal region (R68A) had no effect on the receptor activation by thrombin or trypsin in the phosphatidylinositol 4,5-bisphosphate hydrolysis assay (data not shown). These data further support the conclusion that the putative protease cleavage site of Arg-47/Gly-48 in PAR4 was critical for receptor activation.

Protease Receptor Activating Peptide. The protease-activated receptor family has been shown to be activated by a peptide derived from the amino terminus of the receptor protein. Accordingly, a hexapeptide (GYPGQV) corresponding to the unmasked amino terminus of PAR4 after the cleavage at Arg-47/Gly-48 was tested for its ability to stimulate COS cells expressing PAR4. The peptide readily activated both wild-type and mutant PAR4 (R47A) at 500 μM, whereas thrombin and trypsin only activate the wild-type PAR4 (Fig. 3). COS cells with no transfected DNA failed to respond to the activating peptide under the same condition (data not shown). The maximal response of cells expressing PAR4 to the activating peptide was comparable to the maximal response to thrombin or trypsin (Fig. 3). The activating peptide (SFLLRN) from PAR1 showed no activity toward PAR4 when tested at the same concentration. The EC_{50} of PAR4 activating peptide was about 100 μM, which is substantially higher than that of the activating peptide for PAR1 (2). The high EC_{50} for the activating peptide for PAR4 compared with thrombin or trypsin clearly reflects the difference between a built-in tethered ligand and a ligand in free solution.

Potential Intracellular Phosphorylation Sites. Because the termination of the signaling of PAR4 may occur by phosphor-

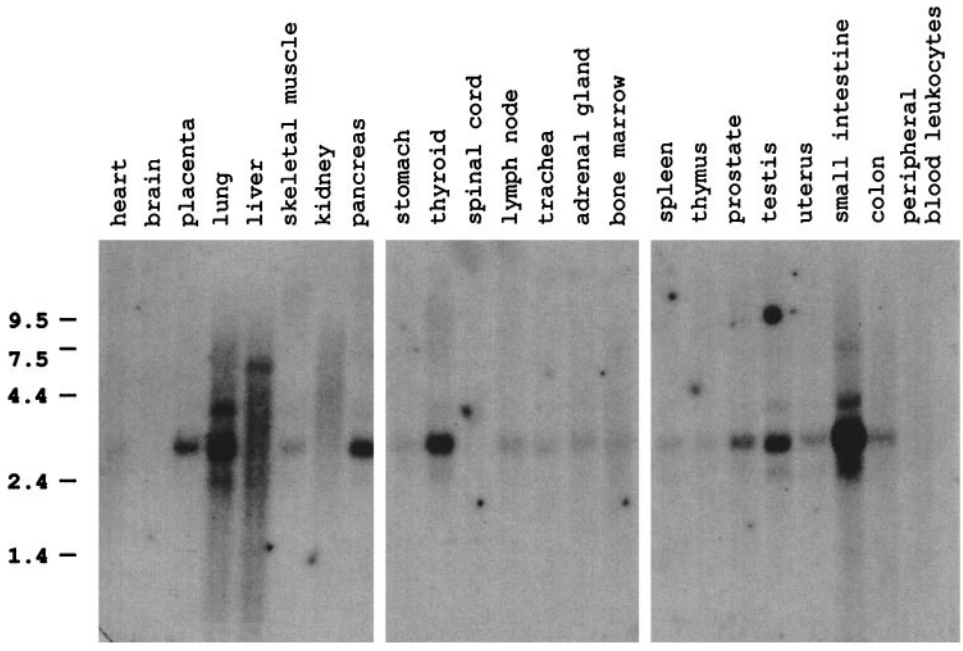

FIG. 5. Northern blot of human multiple tissue mRNA hybridized to a human PAR4 cDNA probe. Tissue sources are given at the top and the positions of sized markers (in kilobases) are indicated on the left. The predominate hybridizing species corresponds to an mRNA of about 2.7 kb.

ylation analogous to the β-adrenergic receptor (28), it was of interest to examine the intracellular regions of PAR4 for potential phosphorylation sites. A serine residue in the sequence SGR is present in the third intracellular loop of PAR4 that could be phosphorylated by protein kinase C, and another serine residue in the sequence SPGD is present in the carboxyl-terminal region that could be a substrate for casein kinase II (Fig. 2). Accordingly, the termination of PAR4 signaling by phosphorylation may be similar to other seven transmembrane-domain receptors.

Tissue Distribution of PAR4. Northern blot analysis of mRNA from 23 tissues showed that the PAR4 gene was expressed in most of the tissues tested with especially high levels in lung, pancreas, thyroid, testis, and small intestine (Fig. 5). The predominant band observed in all tissues was 2.7 kb, indicating that a polyadenylation signal located in the middle of the 3′ noncoding region was the preferred site for polyadenylation (Fig. 1). Moderate expression was also detected in placenta, skeletal muscle, lymph node, adrenal gland, prostate, uterus, and colon. No PAR4 expression was detected in brain, kidney, spinal cord, and peripheral blood leukocytes. From the tissue distribution of PAR4 mRNA, it is difficult to draw any conclusion about the physiological function of PAR4. The PAR4 mRNA was also detected in human platelets by reverse transcription-coupled PCR, although the expression of PAR4 was much less than that of PAR1 (data not shown).

Chromosome Localization of PAR4. The PAR4 gene was mapped to chromosomal location 19p12 by using a PCR and human/rodent somatic cell hybrid mapping method (20). This location was different from that of the PAR1 and PAR2 genes, which are located within approximately 100 kb of each other at chromosome 5q13. The location of the two latter genes suggested that they arose from a gene duplication event (29). At present the localization of PAR3 is unknown. Additional members of the PAR family probably exist that have evolved through a combination of retroposition and gene duplication (30).

The authors acknowledge the use of the Incyte LifeSeq database and are grateful to Dr. Si Lok for the Daudi cell cDNA library; Dr. Kazuo Fujikawa (University of Washington) for various proteases; Jeff Harris and Will Lint for oligonucleotide synthesis; Betty Haldeman, Mark Maurer, Dao Mai, and Karen Madden for technical assistance; and Dr. Dominic W. Chung for helpful discussion and careful review of the manuscript. This work was supported in part by Grant HL16919 from the National Institutes of Health and a fellowship to H.A. from the Danish Research Academy.

1. Scher, W. (1987) *Lab. Invest.* **57,** 607–633.
2. Vu, T. K. H., Hung, D. T., Wheaton, V. I. & Coughlin, S. R. (1991) *Cell* **64,** 1057–1068.
3. Rasmussen, U. B., Vouret-Craviari, V., Jallat, S., Schlesinger, Y., Pages, G., Pavirani, A., Lecocq, J. P., Pouyssegur, J. & Van Obberghen-Schilling, E. (1991) *FEBS Lett.* **288,** 123–128.
4. Nystedt, S., Emilsson, K., Wahlestedt, C. & Sundelin, J. (1994) *Proc. Natl. Acad. Sci. USA* **91,** 9208–9212.
5. Ishihara, H., Connolly, A. J., Zeng, D., Kahn, M. L., Zheng, Y. W., Timmons, C., Tram, T. & Coughlin, S. R. (1997) *Nature (London)* **356,** 502–506.
6. Connolly, A. J., Ishihara, H., Kahn, M. L., Farese, R. V., Jr. & Coughlin, S. R. (1996) *Nature (London)* **381,** 516–519.
7. Zioncheck, T. F., Roy, S. & Vehar, G. A. (1992) *J. Biol. Chem.* **267,** 3561–3564.
8. Altieri, D. C. & Edgington, T. S. (1990) *J. Immunol.* **145,** 246–253.
9. Wachtfogel, Y. T., Pixley, R. A., Kucich, U., Abrams, W., Weinbaum, G., Schapira, M. & Colman, R. W. (1986) *Blood* **67,** 1731–1737.
10. Hancock, W. W., Grey, S. T., Hau, L., Akalin, E., Orthner, C., Sayegh, M. H. & Salem, H. H. (1995) *Transplantation* **60,** 1525–1532.
11. Selak, M. (1994) *Biochem. J.* **297,** 269–275.
12. Hartmann, T., Ruoss, S. J., Raymond, W. W., Seuwen, K. & Caughey, G. H. (1992) *Am. J. Physiol.* **262,** L528–L534.
13. Chang, W. C., Shi, G. Y., Chow, Y. H., Chang, L. C., Hau, J. S., Lin, M. T., Jen, C. J., Wing, L. Y. C. & Wu, H. L. (1993) *Am. J. Physiol.* **264,** 16975–16979.
14. Coughlin, S. R. (1994) *Proc. Natl. Acad. Sci. USA* **91,** 9200–9202.
15. Molino, M., Woolkalis, M. J., Reavey-Cantwell, J., Pratico, D., Andrade-Gordon, P., Barnathan, E. S. & Brass, L. F. (1997) *J. Biol. Chem.* **272,** 11133–11141.
16. Pearson, W. R. (1990) *Methods Enzymol.* **183,** 63–98.
17. Sanger, F., Nicklen, S. & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. USA* **74,** 5463–5467.
18. Ishii, K., Hein, L., Kobilka, B. & Coughlin, S. R. (1993) *J. Biol. Chem.* **268,** 9780–9786.
19. Nanevicz, T., Wang, L., Chen, M., Ishii, M. & Coughlin, S. R. (1996) *J. Biol. Chem.* **271,** 702–706.
20. Kuestner, R. E., Elrod, R. D., Grant, F. J., Hagen, F. S., Kuijper, J. L., Matthewes, S. L., O'Hara, P. J., Sheppard, P. O., Stroop, S. D., Thompson, D. L., *et al.* (1994) *Mol. Pharm.* **46,** 246–255.
21. Bairoch, A., Bucher, P. & Hofmann, K. (1997) *Nucleic Acids Res.* **25,** 217–221.
22. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. (1990) *J. Mol. Biol.* **215,** 403–410.
23. Gerszten, R. E., Chen, J., Ishii, M., Ishii, K., Wang, L., Nanevicz, T., Turck, C. W., Vu, T. K. H. & Coughlin, S. R. (1994) *Nature (London)* **368,** 648–651.
24. Rydel, T. J., Yin, M., Padmanabhan, K. P., Blankenship, D. T., Cardin, A. D., Correa, P. E., Fenton, J. W. II & Tulinsky, A. (1994) *J. Biol. Chem.* **269,** 22000–22006.
25. Liu, L., Vu, T. K. H., Esmon, C. T. & Coughlin, S. R. (1991) *J. Biol. Chem.* **266,** 16977–16980.
26. Mathews, I. I., Padmanabhan, K. P., Ganesh, V., Tulinsky, A., Ishii, M., Chen, J., Turck, C. W., Coughlin, S. R. & Fenton, J. W., II (1994) *Biochemistry* **33,** 3266–3279.
27. Ishii, K., Gerszten, R., Zheng, Y. W., Welsh, J. B., Turck, C. W. & Coughlin, S. R. (1995) *J. Biol. Chem.* **270,** 16435–16440.
28. Ishii, K., Chen, J., Ishii, M., Koch, W. J., Freedman, N. J., Lefkowitz, R. J. & Coughlin, S. R. (1994) *J. Biol. Chem.* **269,** 1125–1130.
29. Kahn, M., Ishii, K., Kuo, W. L., Piper, M., Connolly, A., Shi, Y. P., Wu, R., Lin, C. C. & Coughlin, S. R. (1996) *Mol. Med.* **2,** 349–357.
30. Probst, W. C., Snyder, L. A., Schuster, D. I., Brosius, J. & Sealfon, S. C. (1992) *DNA Cell Biol.* **11,** 1–20.