The Structure of the Urokinase-Type Plasminogen Activator Receptor Gene

By Janet R. Casey, John G. Petranka, Jennifer Kottra, Donald E. Fleenor, and Wendell F. Rosse

The cellular receptor for urokinase-type plasminogen activator (uPAR) is a glycosylphosphatidylinositol (GPI)-anchored membrane protein that plays a central role in pericellular plasminogen activation. It contains 313 amino acid residues, including 28 cysteine residues in a pattern of three homologous repeats. The cysteine residue pattern suggests that uPAR belongs to a superfamily of proteins including CD59, murine Ly-6, and a variety of elapid snake venom toxins. A novel 1.7-kb uPAR cDNA was isolated that is missing exon 5 and that contains 380 bp not previously reported at the 5’ end. This cDNA was used to probe a human genomic library from which three clones were isolated and analyzed. The uPAR gene consists of 7 exons spread over 23 kb of genomic DNA. Exons 2, 4, and 6 code for homologous domains within the mature protein, as do exons 3, 5, and 7. CD59-like homologous pairs are encoded by exons 2-3, 4-5, and 6-7, respectively. The structure of the gene for uPAR further confirms the relationship of this molecule to the superfamily containing CD59, Ly-6, and the elapid snake venom toxins. © 1994 by The American Society of Hematology.

The UROKINASE-TYPE plasminogen activator (uPA) is one of two serine proteases capable of converting the proenzyme plasminogen into plasmin and thus plays an important role in cell migration and tissue remodeling. Specific receptor sites for uPA are located on the cell surface of neutrophils, monocytes, keratinocytes, and cultured endothelial cells as well as on several neoplastic cell lines. A 55- to 60-kD membrane receptor (designated uPAR) has been identified that binds both urokinase-type plasminogen activator (uPA) and pro-uPA with high affinity and localizes plasminogen activation near the cell surface. uPA binds to uPAR via an epidermal growth factor-like moiety at the amino terminal end of the uPA protein.

The uPAR gene has been assigned to chromosome 19 and Roldan et al. have isolated and sequenced a full-length cDNA that codes for the entire human uPAR receptor. The protein is 313 residues in length and has five potential sites for N-linked glycosylation. Compared with native uPAR protein (apparent molecular mass 55 to 60 kD), deglycosylated uPAR has an apparent molecular mass of only 35 kD, suggesting that the protein is in fact highly glycosylated.

The mature uPAR protein is attached to the cell membrane by a glycosylphosphatidylinositol (GPI) membrane anchor and is truncated at the COOH-terminus before its attachment. Like other GPI-linked proteins, uPAR is absent on affected peripheral blood monocytes and granulocytes from patients with paroxysmal nocturnal hemoglobinuria.

uPAR has a high content of cysteine residues that are arranged in a triplicate pattern that defines three homologous domains. This motif is shared with three other GPI-anchored membrane proteins (murine Ly-6, CD59, and squid glycoprotein Sgp-2) and with a variety of elapid snake venom toxins (the erabutoxin-bungarotoxin series), suggesting that these proteins comprise a superfamily of proteins. Further evidence of the relatedness of these proteins is provided by structural studies describing similar patterns of disulfide pairing in CD59 and the NH2-terminal domain of uPAR.

The genes of other members of the superfamily show similarities in organization: they may have an untranslated first exon; the characteristic protein unit is coded by two exons, the first usually being shorter than the second; and the last exon has an extensive 3’ untranslated region. We report the cloning and characterization of the human uPAR gene and show that its organization closely parallels that of the genes of the other members of the superfamily.

MATERIALS AND METHODS

Isolation of uPAR cDNA. A probe to uPAR was generated by gene-specific polymerase chain reaction (PCR) amplification of HeLa genomic DNA using oligonucleotide primers complementary to the uPAR cDNA sequence reported by Roldan et al. The sequence of the first primer was 5’GGCTGCTCCTCAGCCTGGCCCTGCC3’ (bases 961-985) and that of the second primer was 3’cagcacacagtccggctacaggg5’ (bases 1134-1159). A resulting 198-bp PCR product was cloned in plasmid pCR (Invitrogen, San Diego, CA) and its identity to bp 961-1159 of uPAR cDNA was confirmed by DNA sequencing.

A HeLa cDNA library in λ ZAP II (Stratagene, La Jolla, CA) was probed using the 198-bp PCR fragment that was labeled with α-dCTP using a random hexanucleotide DNA labeling kit (Pharmacia, Piscataway, NJ). A single 1.7-kb positive clone was isolated, amplified, and subcloned into pUC119. Dideoxynucleotide sequencing was performed using the Sequenase system (US Biochemical, Cleveland, OH).

Isolation of uPAR genomic clone. The full-length uPAR cDNA clone described above was radiolabeled with α-dCTP random priming and was used to screen a human placental genomic DNA library constructed in λ FIX II (Stratagene). Two positive clones were isolated and amplified. A third clone was identified when the same genomic DNA library was screened using an α-dCTP random primed labeled 131-bp PCR amplification product that was constructed using uPAR cDNA as the template and oligonucleotides 5’CACCAAGCCTTACGAGGT3’ (bp 301-318) and 3’GTAGTCTGTACTCGACACTG3’ (bp 411-432) of the uPAR cDNA sequence as primers.

Characterization of uPAR genomic structure. The gene structure of uPAR was analyzed by Southern analysis and by subsequent subcloning and sequencing of relevant restriction fragments. Restriction digests using BamHI, Hind III, Pst I, Pvu II, Xba I, EcoRI, Sac I, and Sma I were performed on the three genomic clones in phase λ. The restriction digests were electrophoresed on agarose gels and transferred to nitrocellulose filters. The filters were probed with γ-dATP (3,000 Ci/mmol; Amersham Corp, Arlington Heights, IL) end-labeled oligonucleotides. Prehybridization and hybridization

From the Division of Hematology/Oncology, Departments of Pediatrics and Medicine, Duke University Medical Center, Durham, NC. Supported by Grant No. 5 RO1 HL47831-02 from the National Heart, Lung, and Blood Institute of the National Institutes of Health. Nucleic acid sequences have been deposited in GenBank under accession numbers U07842 and U08839. Address reprint requests to Wendell F. Rosse, MD, Box 3934, Duke University Medical Center, Durham, NC 27710.

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. section 1734 solely to indicate this fact.
were performed at 42°C in a solution containing 4X SSC, 0.1% sodium dodeyl sulfate (SDS), 0.1% NaPPI, 10 mmol HEPES, pH 7.4, 5X Denhardts, 50 μg/ml heparin, and 50 μg/ml salmon testes DNA for 4 and 16 hours, respectively. The filters were washed sequentially in 2X SSC, 0.1% SDS; 1X SSC, 0.1% SDS; and 0.5% SSC, 0.1% SDS all for 20 minutes at 42°C. Autoradiographs were exposed at ~70°C for 4 to 24 hours using an intensifying screen. Positively hybridizing restriction fragments were subcloned into pUC19 for further analysis. Dideoxynucleotide sequencing was performed using oligonucleotide primers complementary to the cDNA.

Reverse transcription (RT)/PCR of HeLa RNA. RT/PCR was used to confirm the presence of the previously unreported 1.7-kb transcript in HeLa RNA. HeLa poly(A)+ RNA was isolated from total RNA using PolyATtract paramagnetic particles (Promega, Madison, WI). Reverse transcription was performed for 45 minutes at 42°C in a 20-μL reaction consisting of 400 ng of poly(A)+ RNA template (or 25 ng of uPAR cDNA for positive control), 50 pmol of oligonucleotide primer 5'CCACCTCTTTTCGACATG5' (bases 1-22 of cDNA) and 3'TATCCCTCCTGACAAAACT3' (bases 380-401 of cDNA) for further analysis. Dideoxynucleotide sequencing was performed using oligonucleotide primers complementary to the cDNA. Because intron 3 was not completely represented in the phage λ genomic clones, its size was estimated by analysis of overlapping restriction digests of genomic DNA.

Isolation of uPAR cDNA. A novel 1.7-kb uPAR cDNA was isolated from the HeLa cDNA library. Sequence analysis showed that the 5' end of the cDNA contained 380 bp not previously reported by Roldan et al.10 To confirm that the new 380 bp was not artifactual, RT/PCR was performed to electrophoresis on a 2% agarose gel containing ethidium bromide. Primers were chosen to amplify a fragment corresponding to the cDNA. The primer pair was complementary to the 5'-most sequence of the cDNA, 5'CAG-TATCCCTCTTTTCGACATG3' (bases 612-629 of the cDNA), 1X RT buffer (50 mmol/L Tris-HCl, pH 8.3; 75 mmol/L KCl; 3 mmol/L MgCl2; 30 mmol/L dithiothreitol [dTT]); 2 mmol/L each dNTP, 5 mmol/L DTT, 3 μg/ml bovine serum albumin (BSA), 0.5 μg/μL RNase inhibitor (US Biochemicals), and 10 U (or an equivalent volume of H2O for negative controls) of Moloney murine leukemia virus (M-MLV) reverse transcriptase (US Biochemicals). PCR was performed in a 100-μL volume consisting of 10 μL of the RT reaction, 50 pmol each of primers 5'CAG-TATCCCTCTTTTCGACATG5' and 3'TATCCCTCCTGACAAAACT3' (bases 504-521 of cDNA), 0.5 μmol/L each dNTP, 2.5 U AmpliTaq DNA polymerase (Perkin-Elmer, Norwalk, CT). Thirty thermal cycles were performed (1 minute at 94°C, 1 minute at 55°C, and 1 minute at 72°C), after which the reaction mix was subjected to electrophoresis on a 2% agarose gel containing ethidium bromide.

**RESULTS**

**Isolation of uPAR cDNA.** A novel 1.7-kb uPAR cDNA was isolated from the HeLa cDNA library. Sequence analysis showed that the 5' end of the cDNA contained 380 bp not previously reported by Roldan et al.10 To confirm that the new 380 bp was not artifactual, RT/PCR was performed using HeLa RNA as template. One member of the PCR primer pair was complementary to the 5'-most sequence of the cDNA. The unreported 380-bp segment that was found in the 1.7-kb cDNA was also found in the two genomic clones containing the 5' end of the gene, thus confirming that it was not an artifact of the cDNA. The first exon is 481 bp in length; its first 426 bp are untranslated and the remainder codes for a highly hydrophobic signal peptide. It is separated from exon 2 by an intron of 2.3 kb. Exon 2 is 111 bp in length and codes for the remainder of the signal sequence and the beginning of the mature protein; it is separated from exon 3 by an intron of 2.3 kb. The 144-bp exon 3 codes for mature protein and is separated from exon 4 by a 9-kb intron that was sized by genomic Southern blot analysis. Exon 4 is 162 bp in length, codes for mature protein, and is separated from exon 5 by an 800-bp intron. Exons 5, 6, and 7 code for the remainder of the mature protein and are 135, 147, and 563 bp in length, respectively; these exons are separated by 3.2-kb introns. Exon 7 contains a 312-bp untranslated segment at its 3' end. Each exon is flanked by appropriate consensus acceptor or donor splice sites (Fig 3).

The organization of exons in the uPAR gene reflects the pattern of triplicated homologous domains observed in the mature protein (Fig 4). Each protein domain is coded for by a pair of exons, with the intron break within each pair always occurring between the fifth and sixth cytoines.

The putative promoter region 5' to exon 1 was sequenced and examined for potential regulatory elements. No TATA or CAAT elements are evident, nor are the first 100 bases G+C rich (39%) as in the CD59 gene. However, there is a pair of octomer sequences at positions −50 and −34, as well as a GATA sequence at position −23 (Fig 3).

**DISCUSSION**

In this report we have identified and described the gene structure of the human uPAR. The structure was derived from analysis of three genomic clones from a human placental genomic library. The uPAR gene sequence is shown in Fig 3, and Fig 2 depicts the seven exons and six introns of varying sizes.

It was previously observed that the protein structure of uPAR belonged to a newly described Ly-6/lepalid venom toxin (EVT) superfamily.17-26 The fundamental structural unit typical of this family is a protein of about 70 amino acids containing 10 cysteines that form internal disulfide bridges producing a characteristic structure. This structure consists of a cysteine-rich hydrophobic core from which radiate three "loops" closed by the disulfide bonds. Behind this core is a small fourth loop. Most proteins in this superfamily consist of a single such unit, whereas uPAR consists of three such units in series; the first of these varies from the usual pattern in that one of the disulfide bridges holding a loop is missing.

Further evidence that these proteins belong to the same superfamily is seen in the fact that the structure of their genes is similar25-26 (Fig 5). In the uPAR gene, like those of the snake venom toxins, there is no untranslated first exon...
UPAR cDNA of Roldan et al.

Novel UPAR cDNA

Fig 1. Comparison of UPAR cDNA described by Roldan et al.11 with the novel cDNA described in this work. Exons are numbered, with dotted lines separating members of exon pairs that code for homologous repeats. ( ) Previously undescribed 380-bp 5' sequence. Sequence about the spliced-out exon 5 is shown. Exon 4 sequence is underlined, exon 6 sequence is double-underlined, and the amino acid translation is shown above. *The substitution of valine for isoleucine at the beginning of exon 6 in the alternatively spliced form. (B) 3' untranslated sequence.

as there is in the genes for CD59 and Ly-6C.1. Exon 1 of uPAR is similar to exon 2 of CD59 and Ly-6C.1 in that both exons have an untranslated 5' segment and code for the hydrophobic signal peptide.

The fundamental unit of the superfamily is coded for in each case by two exons, one (usually shorter) coding for the amino end of the unit and the other for the carboxyl end. In the uPAR gene, as in the genes of the other members of the superfamily, the exon break in each pair coding a subunit is always type 1 and occurs in the same general part of the molecule as in the other members of the superfamily. This suggests that the two-exon structure of the gene was present in the archetypal gene from which the members of the family are derived. Because uPAR contains three of the fundamental units characteristic of the EVT family, three exon pairs code for the complete structure. Each of these is like the exon pairs coding for a single unit, suggesting that the triplicate structure is the result of reduplication of the fundamental unit. However, there is not sufficient homology between the individual units either at the nucleic acid or at the protein level to ascertain at what point evolutionarily this reduplication occurred, or in what order.

The absence of exon 5 in our cDNA would lead to translation of a protein in which the carboxyl-terminal half of the second homologous repeat is missing. The mRNA remains in frame at the exon 4/exon 6 splice junction, so that the third homologous repeat as well as the GPI-anchoring signal would presumably still be intact (with the exception of the valine to isoleucine substitution at the splice junction). Such a molecule would have an uneven number of cysteines in the remaining half-domain of the second repeat, and the implications of this in terms of intramolecular bonding and protein folding are not known. If such a molecule were to be expressed, uPAR function might not be affected because

![Diagram of the uPAR gene structure](https://via.placeholder.com/150)

Fig 2. The structure of the uPAR-encoding gene. Exons are indicated by boxes; translated portions are shaded and untranslated portions are unshaded. The restriction sites used in assessment of the structure are indicated (P, Pst I; E, EcoRI; H, HindIII; B, BamHI; X, Xba I). The intron sizes are indicated below. The three genomic clones are shown below the gene structure and their lengths depict the portion of the gene contained within them.
Fig 3. Nucleotide sequence and salient features of the uPAR gene. Exons are denoted by boldface uppercase letters, with the previously unreported 380-bp cDNA sequence underlined. Protein sequence is indicated above translated portions of exons. Consensus splice donor and acceptor sites are indicated by double underlining.

Fig 4. Alignment and comparison of homologous exons of UPAR. Homologous exons are aligned based on conservation in the spacing of cysteine residues (shaded). Each repeat of the triplicated protein motif is coded for by a pair of exons (exons 2/3, 4/5, and 8/7) with the intron break within a pair always occurring between the fifth and sixth cysteine.
uPA-binding activity appears to reside in the amino terminal 87 residues of the first repeat. Kristenson et al. described an alternatively spliced mouse cDNA that results in premature termination of protein translation at a similar position in the second homologous repeat. That molecule could be translated in vitro to produce a protein of the appropriate size. In addition, in situ hybridization with antisense RNA probes showed that the truncated mRNA was both transcribed and differentially localized from full-length message in mouse gastrointestinal tissue. However, it was not shown that a functional uPA-binding protein could be translated in vivo.

In all the members of the superfamily that are membrane-bound, the last exon (exon 4 in CD59 and Ly-6C.1; exon 7 in uPAR) codes for the carboxyl terminus of the mature protein, as well as for the signal sequence for GPI anchor attachment, and ends with a long untranslated segment (Fig. 5). In the snake venom toxins, which are secreted, the last exon does not encode the characteristic hydrophobic sequence of amino acids necessary for GPI attachment. Pyke et al. have recently described an alternatively spliced uPAR mRNA in which exon 7 is replaced by an alternative final exon encoding a 29-residue hydrophilic sequence that presumably gives rise to a soluble secreted form of uPAR. This alternative exon 7 resides 578 bp downstream from the end of the originally described final exon and is homologous neither with the other exons of uPAR nor with the sequence of any known protein. Despite this, the cysteines at positions 240 and 250 of the alternative final exon are spaced at approximately the same interval as the first two cysteines of the typical exon 7; this provides the possibility, at least, that the intramolecular disulfide bonds that maintain loops one and two within the third homologous repeat may be preserved. Whether or not these disulfides are present, uPAR function might well not be affected because uPA-binding activity appears to reside in the first homologous repeat.

These studies show that the gene for uPAR closely resembles the gene structure of the other members of the EVT superfamily, but has been triplicated. The remarkable consistency with which this gene structure has been preserved suggests that there may be some evolutionary advantage to it. Further studies will be necessary to determine the biologic significance, if any, of the three uPAR mRNA variants that have thus far been described.

REFERENCES

7. Appella E, Robinson EA, Ulrich SJ, Stoppelli MP, Corti A,
The structure of the urokinase-type plasminogen activator receptor gene

JR Casey, JG Petranka, J Kottra, DE Fleenor and WF Rosse