A Conserved TATA-Less Proximal Promoter Drives Basal Transcription From the Urokinase-Type Plasminogen Activator Receptor Gene

By Emilia Soravia, Alexandra Grebe, Pasquale De Luca, Kristian Helin, Theodore T. Suh, Jay L. Degen, and Francesco Blasi

The urokinase-type plasminogen activator receptor (uPAR) focuses at the cell surface the activation of pro-uPA and, hence, the formation of plasmin, thus enhancing directional extracellular proteolysis. To characterize the transcriptional regulatory mechanisms that control receptor expression, we have cloned an uPAR DNA segment containing upstream regulatory sequences from both the human and murine genomes. We report that a proximal promoter, contained within 180 bp from the major transcription start sites of the human uPAR gene, drives basal transcription. This region lacks TATA and CAAT boxes and contains relatively GC-rich proximal sequences. A subregion of this sequence, highly conserved between human and murine genes, contains most of the promoter activity and is specifically bound by HeLa nuclear proteins, one of which belongs to the SP1 class.

© 1995 by The American Society of Hematology.

THE MIGRATION OF normal and neoplastic cells across tissue barriers is regulated by plasmin-mediated degradation of extracellular matrix proteins, as well as activation of factors such as hepatocyte growth factor and transforming growth factor (TGF)-β. Activation of plasminogen and other factors is mediated by urokinase-type plasminogen activator (uPA), which is held on the cell surface by a specific membrane receptor (uPAR). Three inhibitors, PAI-1, PAI-2, and PN-1, can bind the uPA moiety of this complex to regulate its activity. In addition to recruiting migration-regulatory factors at specific areas of the pericellular space, the receptor also exerts the action to unload stable uPA- inhibitor complexes from the cell surface by a mechanism of internalization/degradation. This process also involves a second receptor protein identified as the α2-macroglobulin receptor. The uPAR mRNA encodes a protein containing 313 amino acids plus the signal peptide, which is processed during biosynthesis by glycosylation, carboxy-terminal processing, and addition of a glycosyl-phosphatidylinositol (GPI) anchor. The polypeptide sequence exhibits three regions of internal homology with a distinctive distribution of cysteines and disulfide bonds common to other GPI-anchored proteins of the LY6/elapid venom toxin superfamily.

The three domains contain approximately 90 amino acids each, and the amino terminal domain contains ligand-binding activity. The uPAR is expressed in a variety of normal and neoplastic cells. Its expression is influenced by a number of hormones and growth factors that may superimpose a cell type-specific modulation over basal receptor expression. The expression of uPA, PAI-1, and PAI-2 also exhibit similar characteristics, thus suggesting some kind of connection between regulatory mechanisms of the components of the plasminogen activation system. Like uPA, the uPAR is found at very high levels in human tumors where it is localized at the invading edge. To investigate the mechanisms that regulate transcription of the uPAR gene, we have isolated upstream DNA sequences, identified the transcription start sites, and defined a minimal promoter region required for basal transcription of the human uPAR gene. We have found that 181 bp upstream from the transcription start site contain substantial basal promoter activity and are bound by multiple nuclear factors, including SP1, and enclose a region highly conserved between human and murine uPAR genes.

MATERIALS AND METHODS

Cloning and nucleotide sequence determination of human uPAR 5' flanking region. A human leukocyte genomic library in XEMBL3 phage (obtained from Clontech, San Diego, CA) was screened using human uPAR cDNA as hybridization probe. Three identical recombinant phage clones containing uPAR sequences were identified from 107 plaques. Five DNA fragments, representing the whole insert, were separated from the phage vector following digestion with Sal I and BamHI and subcloned into pEMBL8 plasmid vector. The DNA insert was mapped by digestion with different restriction enzymes and hybridized to cDNA fragments and to a 40-bp long synthetic oligonucleotide covering the 5' end of the published cDNA data not shown). The nucleotide sequence of the human uPAR genomic clone was determined on both strands of the DNA by the dideoxynucleotide chain termination method. Oligonucleotide primers were synthesized with an Applied Biosystems DNA synthesizer (Applied Biosystems, Foster City, CA). Oligonucleotides for DNA sequencing were chosen from the sequence of the uPAR cDNA.

RNA isolation and primer extension. U937 cells were maintained in RPMI supplemented with 10% fetal calf serum (FCS). Cells were stimulated 48 hours before harvesting with 75 mmol/L phorbol myristate acetate (PMA). Total RNA was isolated using the guanidinium thiocyanate method. Poly A+ RNA was purified by Oligo d-T cellulose chromatography (Pharmacia, Uppsala, Sweden).
sequences up to -1275 (Fig 1A) was subcloned in pSP73 (Promega) and mung bean nuclease and religated. Deletion endpoints were Kpn I and Sal I (in the polylinker), treated with exonuclease IIP5 to yield pPXb73; this intermediate vector was then digested with EX0l-l Position in cDNA8

<table>
<thead>
<tr>
<th>Exon</th>
<th>Position in cDNA</th>
<th>Size of Exon (bp)</th>
<th>Amino Acid</th>
<th>Interrupted</th>
<th>5'-Donor Site</th>
<th>Intron (kb)</th>
<th>3'-Donor Site</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1-101</td>
<td>101</td>
<td>Ala-19</td>
<td></td>
<td>TCCAG gtaggg</td>
<td>2.6</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>102-212</td>
<td>111</td>
<td>Glu-56 (34)*</td>
<td></td>
<td>GGGAAgtgagc</td>
<td>2.4</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>213-356</td>
<td>154</td>
<td>Gly-104 (82)*</td>
<td></td>
<td>ACTCTGTgtagag</td>
<td>&gt;5</td>
<td></td>
</tr>
</tbody>
</table>

The sizes of introns I and II were determined by agarose gel electrophoresis of isolated DNA fragments that were not entirely sequenced. Exon sequences are shown in upper case letters, and intron sequences are in lower case letters.

* The numbers in parentheses indicate amino acid positions in the mature polypeptide.

Transcription initiation sites were searched with the primer extension method using primer P1 of 19 nucleotides (complementary to +122 to +141, with sequence 5'CCGTGGCTTACATTCTCTGC3') and P2 of 31 nucleotides (+41 to +72 of the published cDNA, with sequence 5'AGCCGCAGACGGCGGGTGACCATGTGCGGC3'). The primers were end-labeled with γ32P ATP using T4 polynucleotide kinase and purified by denaturing 10% polyacrylamide gel. For primer extension reactions, the labeled oligonucleotide was mixed with 2 μg of poly A+ RNA or 25 μg of total RNA for 10 minutes, heated to 70°C, and allowed to anneal by cooling at 42°C. The annealed primer was extended using 100 U of murine leukemia virus reverse transcriptase in the presence of 2 mMol/L deoxynucleoside 5' triphosphates, in 50 mMol/L Tris-HCl, 75 mMol/L KCl, 3 mMol/L MgCl2, and 10 mMol/L DTT, for 1 hour at 37°C in a 50 μL volume. The extended products were analyzed by polyacrylamide gel electrophoresis (6% denaturing polyacrylamide gel)

RNase protection analysis. A DNA fragment between position -401 and +107 of the human uPAR gene was generated by polymerase chain reaction (PCR) amplification and insert into Xba I site in the polylinker of pSP73 vector (Promega, Madison, WI). This construct, linearized with Xma III (position -148), was transcribed in vitro with T7 RNA polymerase in the presence of α-32P-UTP to generate the probe for RNase protection. A total of 50 μg of total RNA isolated from PMA treated U937 cells and 25 μg of RNA from PMA-treated HeLa cells (PMA concentration was 100 ng/mL) for 12 and 6 hours, respectively) were hybridized using 2 X 104 cpm of probe in 80 PL for 16 hours at 42°C in 50% formamide, 40 mmol/L TRIS-HCl (pH 7.5), 300 mmol/L NaCl, 7.5 mmol/L EDTA. After hybridization, RNase buffer, the mixture was digested with 10 μg/mL RNase A and 1,000 U/mL RNase T (Boehringer Mannheim, Germany) at 32°C. Products were resolved on 8% denaturing polyacrylamide gels.

Construction of huPAR-CAT expression vectors. A 2.3-kb ShpI/Nrfl genomic fragment, containing the 5' flank region extending up to +46 into exon 1, was then cloned into the Shp I Usma I sites of pSP73 vector (Promega) to produce ps'5'R73 plasmid. ps'5'R73 was digested with Shp I and Cfu I sites and the 2.3-kb insert upstream of the AUG was then ligated into Shp I/Acc I sites of pCAT-Basic vector (Promega) to yield construct C1 (see Fig 3). A 450-bp Xba I fragment from C1 was inserted in either orientation in the Xba I site of pCAT Basic vector, to generate C2 and C3. Constructs C4 and C5 were obtained by insertion of PCR-amplified fragments containing sequences -41 to -181 and +41 to -71, respectively, in the Xba I site of pCAT Basic vector. All constructs were verified by sequencing DNA sequencing. 3' deletions (C6, C7, and C8) were obtained by exonuclease III digestion; a Pst I genomic fragment containing the first exon, part of the first intron and upstream sequences up to -1273 (Fig 1A) was subcloned in pSP73 (Promega) to yield pPX173; this intermediate vector was then digested with Kpn I and Sal I (in the polylinker), treated with exonuclease III and mung bean nuclease and religated. Deletion endpoints were determined by sequencing. A select number of 3' deletions were then digested with Xba I and subcloned into the Xba I site of pCAT-Basic vector (C6, C7, and C8). Construct 9 was derived by digestion of pPXU73 with Pst I and Xba I and subcloning of the isolated fragment in Pst I and Xba I in pCAT Basic. Construct 10 was obtained by inserting in C9 digested with Xba I, a fragment excised from C7 with Xba I. In constructs C1-C3, C6-C8 and C10, 31 and 28 nucleotides, respectively, of the pSP73 multiple cloning site separate the uPAR 5' flanking sequence from the Xba I site of pCAT Basic, which is located 41 nucleotides upstream the CAT translation initiation codon. In constructs C4 and C5, the uPAR 5' flanking sequences are directly fused to the Xba I site.

Cell culture, transfections, CAT assay. The HeLa cell line was routinely maintained in Dulbecco modified Eagle medium supplemented with 10% fetal bovine serum. Twenty-four hours before transfection, 7 X 104 cells were transferred into 100-mm plates. A total of 10 μg of plasmid DNA construct, 2 μg of pXGHS DNA internal control plasmid (Allegra HGH; Nichols Institute Diagnostic, S. Juan Capiistrano, CA), and 10 μg of carrier DNA were coexpressed with calcium phosphate. The medium was replaced after approximately 16 hours. Forty-eight hours posttransfection, the secreted growth hormone was measured and cells were harvested. CAT assays were performed as described.30 After heating to 65°C for 10 minutes, 5-50 μL of cell extract was incubated with 300 μmol/L n-butyryl coenzyme A and 0.2 μCl of 14C-chloramphenicol (Du Pont, Wilmington, DE) in 100 μL of 250 mMol/L Tris-HCl pH 8. After 2 hours at 37°C, the reaction products were isolated by extraction with 200 μL of xylene, back extracted twice with 200 μL of 10 mMol/L Tris-HCl (pH 7.5)-1 mMol/L EDTA, mixed with 5 mL of scintillation fluid, and radioactivity was counted in a Beckman scintillation counter (Beckman Instruments, Irvine, CA). Protein concentrations were determined with the Bradford protein assay (Bio-Rad, Richmond, CA) and the results of CAT assays were normalized to 50 μg protein and to the secreted HGH as determined with a radioimmunoassay kit (Allegra HGH; Nichols Institute Diagnostic). CAT activities of the samples were in the linear range of the assay. Each experiment was performed at least three times and in duplicate. The Drasophila melanogaster Schneider line 2-3S was cultured at 27°C in Schneider's Drosophila Medium (GIBCO, Grand Island, NY) supplemented with 10% fetal bovine serum and penicillin/streptomycin (GIBCO). A total of 1 X 105 cells were seeded in 100-mm dishes and transfected 24 hours later with the calcium phosphate precipitation method using 1 μg of plasmid (Pharmacia), 5 μg of promoter-CAT plasmid and P6_Spi or P6_O.30 Harvesting was 48 hours posttransfection. Transfection efficiency was normalized to β-galactosidase activity. Protein concentration of the extracts was measured, and 50 μg were analyzed for CAT enzyme activity using 14C Chloramphenicol and thin layer chromatography. Conversion values were obtained by a Phosphoimage System (Molecular Dynamics, Kemsing, UK).

Electrophoretic mobility shift assays. Nuclear extracts from HeLa cells were prepared according to the method of Dignam.30 A 256-bp fragment (-260 to position -4) used as probe was synthe-
The positive clone λuPAR-1, carrying a 14-kb insert, was analyzed by restriction mapping and Southern blot hybridization using as probes a set of fragments generated from the previously cloned cDNA (see Materials and Methods) (Fig 1A). Adjacent and overlapping subfragments of the DNA insert were subcloned in plasmid vectors and sequenced, showing that AuPAR-1 contained the 5' end of the uPAR gene. We, therefore, determined the intron-exon junctions in this region and characterized by restriction mapping and sequencing the 5' flanking region (Fig 1A). Comparison of the sequence obtained with that of the cDNA showed the presence of sequences flanking the 5' end of the uPAR cDNA, coding sequences including the cDNA up to the codon for residue gly104 and intervening sequences 3' of this codon (Fig 1A). The coding sequences covered the signal peptide and the whole first domain of uPAR and were arranged in three exons: exon 1 included the transcription start sites (see below), the 5'-untranslated end of the cDNA and most of the signal peptide extending through the codon for the 19th residue (ala) in the signal peptide; exon 2 (111 bp) encoded from residue 20 in the signal peptide through the 19th residue of the mature protein; exon 3 (144 bp) contained sequences throughout the codon for residue gly82 of the mature protein (Fig 1B). The size of the intervening sequences was 2.6 kb for intron 1 and 2.4 kb for intron 2. The 3' portion of the phase insert ended with about 5 kb not containing coding sequences and, hence, probably belonging to intron 3, separating the third and fourth exons. The sequences at the intron-exon junctions are reported in Table 1.

Identification of the transcription initiation site. The cDNA for uPAR contains a 45 nucleotides-long untranslated region upstream of the ATG initiation codon.5 To identify the 5'-flanking region of the uPAR gene, it was necessary to define its transcription initiation site. For this we used RNA of PMA-treated U937 monocytic cells and HeLa cells

RESULTS

The genomic structure of the first domain of uPAR. We have isolated sequences encoding the human urokinase receptor (uPAR) gene from a human genomic leukocyte library in phage λEMBL-3 (see Materials and Methods section).
that both express high levels of uPAR.\textsuperscript{18,21} Oligonucleotide primers corresponding to two distinct segments of the uPAR mRNA were separately hybridized to either total or poly A\textsuperscript{+} RNA from U937 cells and incubated with reverse transcriptase. Primer extension products were analyzed on the same gel along with the corresponding uPAR genomic DNA or with an unrelated DNA ladder (Fig 2). The extension products could be identified from their size. Primer P1 corresponds to the sequence +122 to +141, primer P2 to +41 to +72 of the cDNA sequence.\textsuperscript{8} Figure 2B shows that primer P1 was extended to products of roughly 145 nucleotides in length, while primer P2 was extended to products of about 75 nucleotides (Fig 2C). This result indicates that the extension of both primers ended at comparable positions on the uPAR genomic sequence and that the major transcription initiation site resided only three to six nucleotides upstream of the uPAR promoter sequence (Fig 1).

![Fig 1. (Cont'd) (B) Sequence of the 5' flanking region of the uPAR gene. Nucleotides are numbered from the transcription initiation site (see text), which is indicated with the arrow (+1). Downstream initiation sites are in boldface type. A screening of the nucleotide database for matching sequences indicated the presence of two blocks of Alu-like repetitive elements (enclosed in boxes) in the distal portion. The sequences of a reverse CAAT box, a TATA box in the upstream Alu element, and of potential regulatory elements are underlined.](image-url)
Fig 2. Identification of the transcription start site of the human uPAR gene by primer extension and RNase protection of RNA from PMA stimulated U937 monocytes. (A) The oligonucleotides P1 and P2 used for primer extension are indicated and matched with the uPAR cDNA sequence. Radiolabeled primers are shown as thick lines and extended primers as thin dashed lines. The size of the extended products is indicated. (B) Primer extension using P1 primer (complementary to +122 to +141) and 25 μg total RNA. Extended products were separated on a 6% sequencing gel. A, G, C, and T lanes represent dideoxy-sequencing reactions used as size marker together with a labeled 82-bp fragment (lane 1). P1 extended products are shown in lane 2. (C) Primer extension with P2 primer (+41 to +72) and 2 μg of Poly A+ RNA (lane 1). Lanes A, G, C, and T represent dideoxy-sequencing reactions using P2 primer on double-stranded uPAR genomic template. The length of the extended products (indicated with the arrows) was estimated by comparison with the sequencing ladder. (D) Transcription start sites (arrowheads); a partial match to the consensus initiator sequence of the DHFR gene is indicated with dashed lines. The A at −25 (boldface type) indicates a minor extension products. The underlined sequences represent the positions of the two primers.
of the 5' terminus of the cDNA clone. Moreover, the products obtained with primer P2 were resolved as a group of six bands mapped in a sequence of 16 nucleotides (see Fig 2C). Figure 2D shows the DNA sequence of the flanking region of the human uPAR gene with the identification of the various transcription start sites. The most upstream of the start sites, an A following a C, was arbitrarily given the position +1. This site is located six nucleotides upstream of the 5' terminus of the cDNA and 52 nucleotides upstream of the translation initiation codon. A minor extension product of 170 nt was also noticed with primer P1 on total RNA (Fig 2A indicated with a small arrow) that could represent a minor start site at position −25. Although the data of Fig 2 do not rule out the possibility of additional transcription initiation events at more upstream sites, use of primers at positions −42 to −71 and −160 to −194 of the genomic sequence did not show any additional start site (not shown). To confirm the location of the transcription start site, we performed RNase protection analysis with RNA from PMA-stimulated U937 and HeLa cells. Total RNA was hybridized to a 255 nucleotides probe containing 200 nucleotides upstream and 55 nucleotides downstream of the translation initiation codon. In both cases 110 nucleotides were protected, placing the transcription start sites at the same position identified by primer extension analysis. The DNA sequence in proximity of this start site (−12 CCTTCCCTGA-GGCCA +1) shows homology with a consensus Initiator (Inr) element sequence of the DHFR gene (ATTTCA[130]GCCA)41 (Fig 1B). The presence of Inr-like elements and the occurrence of multiple transcription initiation sites are
Fig 3. Functional analysis of the uPAR promoter deletions. (A) Schematic representation of the uPAR promoter showing restriction sites used for deletions. Numbering is relative to the major transcription start site. (B) Structure of deletion mutants and CAT activity measured after transfection into HeLa cells. CAT activity is given relative to the parental promoter-less vector and is corrected for transfection efficiency. Results are from three experiments and are given as mean ± standard error.

The promoter of the human uPAR gene. The XuPAR-I insert contained about 4.5 kb upstream of the first uPAR exon (Fig 1A). Within 1.5 kb upstream (Fig 1B), we identified two Alu-type repetitive elements in the distal portion (enclosed in boxes, Fig 1B). To identify the promoter region, we fused to the bacterial CAT gene the sequences from -2.3 kb to +46 and the sequences from -0.4 kb to +46 and measured CAT activity after transfection in HeLa cells. In four distinct sets of experiments, we observed that transfection with constructs carrying 2.3 kb (construct C1) or 0.4 kb (construct C2) of upstream sequences elicited a 24-fold and 35-fold higher activity, respectively, than the promoterless vector (Fig 3). Because these two constructs drove comparable CAT gene expression, the promoter must be contained within 0.4 kb from the transcription start site. Furthermore, the activity of this region is orientation-dependent (construct C3). We also noticed that construct C2 was 1.5-fold more active than construct C1, and this might indicate the presence, upstream of position -401, of a weak negative regulatory element that remains to be characterized. A better resolution of the active region was obtained using construct C4 ending at position -181, which showed 83% of the activity of construct C2. The role of proximal sequences was confirmed by 3' deletions constructs (C6-C8) which showed eightfold reduction in activity. Construct C9 and C10, lacking sequences from -401 to +46, did not show any activity. We conclude from these data that 181 nucleotides upstream of the transcription start site account for most of the transcription-promoting activity and that sequences -401 to -182 contribute weakly to the overall activity. By contrast, more upstream sequences do not contribute any activity, at least under basal conditions (see also constructs C9 and C10). Within the proximal 0.4 kb, we observed a sequence of 240 nucleotides containing 63% GC, proximal to the transcription start site. This region lacks TATA and CAAT boxes and perfect matches to typical SP1 consensus elements, although contains five SP1-like sequences, each with a single mismatch: three copies of the sequence GGGAGG are observed at position -44 to -39 and between nucleotides -102 and -88, and the sequences GGGCCG and GGGCTG are observed at positions -175 to -170 and -73 to -68, respectively. A PEA3 consensus sequence is present more upstream (-241 to -248). The most proximal 5' flanking 181 bp contain one mismatched AP2-like sequence (-168 to -161), one AP1 consensus sequence (-182 to -176), a combined Glucocorticoid Responsive/AP1-like Element (-69 to -63) (Fig 1B). Alignment of the human promoter with the corresponding murine sequence is shown in Fig 4.
Fig 4. Comparison of proximal promoter sequences of the human and murine uPAR genes. The BESTFIT program identified 68% identity between the human (upper row) and the murine (lower row) sequences. Boldface letters correspond to the transcription start sites. Dots above indicate match to the DHFR Inr consensus. GA-rich stretches are underlined; putative binding sites for known transcription factors are also indicated by boldface letters. The sequences corresponding to oligo A and oligo C are indicated.

shows 68% identity in a region of about 135 nucleotides, and that one of the human transcription start sites coincides with the major transcription start site determined with murine cells mRNA. Both the human and mouse promoters lack TATA boxes and present a moderately high GC content in the proximal 134 nucleotides (66%). The mouse sequence contains multiple canonical SP1 binding sites and a perfect match with an AP1 site (TGACTCA). Beyond this area of homology, only another AP1 site is conserved between human and mouse (position -182 to -176 in the human sequence).

In vitro protein DNA interactions at the uPAR gene promoter. To identify nuclear factors that interact with the human uPAR promoter, we used nuclear extracts from HeLa cells. Binding assays were done with a 256-bp fragment (-4 to -260) encompassing the most active promoter region. Large amounts of specific DNA fragment were used in the binding reactions to resolve individual complexes formed with this long probe, presumably containing multiple protein binding sites. Under these conditions, one major and six minor complexes were resolved in native polyacrylamide gel using HeLa nuclear extracts (Fig 5, lane 1). These complexes were observed reproducibly and appear to be specific, as they were all competed out with excess unlabeled fragment (lanes 2 and 3), but not with Escherichia coli DNA fragments (lanes 10 and 11), or poly d(IC) (not shown). Complex 3 (indicated with an asterisk in Fig 5) was only partially self-competed, but was not at all competed by nonspecific competitors (Fig 5, lanes 10 and 11) or oligonucleotides bearing AP1 and AP2 consensus sequences did not compete any of the seven complexes (Fig 5, lanes 4 to 7). Using a double-stranded oligonucleotide bearing sequences proximal to the transcriptional start site from -44 to -6, we observed that complex 5 was specifically competed (data not shown). Furthermore, an oligonucleotide containing sequences from -44 to -6 (oligo A) produced a specific complex in electrophoretic mobility shift assays (Fig 6).

<table>
<thead>
<tr>
<th>probe</th>
<th>Competitor</th>
<th>Fragment-260/-4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Free probe</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>10</td>
<td>50</td>
<td>100</td>
</tr>
<tr>
<td>E. coli DNA</td>
<td>*</td>
<td></td>
</tr>
</tbody>
</table>

Fig 5. Electrophoretic mobility shift assay using a 32P-labeled fragment from -260 to -4 as probe. Self-competitions and competitions with other DNAs are as indicated. The asterisk corresponds to the major complex that is only partially self-competed. Arrows indicate complexes specifically competed by SP1.
Identification of the SP1 transcription factor binding sites in the uPAR minimal promoter. The huPAR promoter resembles many other mammalian promoters lacking a proximal TATA element that have been shown to require SP1 for transcriptional activity. DpnI protection assay with recombinant SP1 showed three distinct footprints between nucleotides −125 and −24 (Fig 7), comprised within the most active region and containing two noncanonical SP1 sites (GGAG motif). We note that this motif has been shown to be bound by SP1 in the early promoter of the glia-cell specific virus JC. When HeLa nuclear extracts were used for footprinting analysis, we observed protections at positions −93 to −73 (site 1), −70 to −45 (site 2), −44 to −37 (site 3), and −29 to −19 (site 4). Of these, only sites 1 and 3 were similar, although not exactly identical, to the protections observed with recombinant SP1. Site 4 is comprised within the sequence of oligo A, which has been shown above, to be bound by a HeLa nuclear factor. A comparison of the footprints observed with pure recombinant SP1 and with HeLa nuclear extracts is in Fig 7C.

SP1 activation of the uPAR promoter in Schneider cells. As a preliminary step to establish a functional role of SP1 in the activity of the basal promoter, we cotransfected Drosophila Schneider cells that lack endogenous SP1 with huPAR-CAT constructs C2 (−401 to +46) or C7 (−401 to −227) and with a SP1 expression plasmid (Fig 8). In this experiment, both constructs exhibited activity depending on SP1 expression (no activity was detected in control transfections with expression plasmid alone), thus suggesting that SP1 can drive transcription from the uPAR promoter. Construct C7 only showed 2.5-fold lower activity than construct C2, probably owing to the presence of residual SP1 sites between positions −227 and −401.

DISCUSSION

The uPAR gene belongs to the Ly-6/elapid venom toxins multigene family, including genes for lymphoid cell antigens, and snake venom toxins. Members of this family are attached to the plasma membrane via a glycosyl phosphatidylinositol anchor. The sequence of the uPAR protein suggests the presence of three repeated structural domains with the amino-terminal domain having ligand-binding activity. Unlike uPAR, members of the Ly6-elapid venom toxins superfamily are single domain polypeptides. The structural homology among the members of this superfamily is defined on the basis of the conserved cysteine spacing and disulphide bond distribution, which has been determined for the amino terminal ligand-binding domain of uPAR, for CD59 and for the venom toxin, whose crystal structure has also been determined. The sequence of the uPAR genomic DNA shows that the signal peptide is encoded by a separate exon and that there is no untranslated exon, unlike members of the Ly-6 gene family. The uPAR ligand-binding domain is encoded by two exons (see Fig 1A). This basic two-exon structure, also described for the single domain proteins of the superfamily, is most probably the ancestral domain from which different members of the superfamily have evolved. Transcription from the uPAR gene initiates at several sites in proximity of the 5′ end of a previously characterized uPAR cDNA clone. Consistent with multiple transcription initiation sites, the 5′ flanking sequences of the uPAR gene lacks a TATA box that usually determines transcription initiation from a single major site. Primer extension and RNase protection experiments gave no evidence of other upstream initiation sites. A proximal basal promoter was contained in a region of 401 bp from the transcription start site and drove expression of uPAR-CAT fusions after transfection in HeLa cells. Deletion of the region encompassing the transcription start sites led to a strong reduction of CAT activity. These deletions still had some residual activity, which may depend on more upstream sequences (C6-C8); however, construct C9, missing altogether the −401 to +46 region, is totally devoid of activity. Transfection experiments overall showed that most of the activity is contained within 181 bp. There is a high sequence conservation of human and murine genes in the 135 nucleotides, which include the transcription start sites, thus suggesting the functional importance of this promoter domain. In the absence of a TATA box, transcriptional activation may function through SP1 bound to GC rich sequences. It has been proposed that SP1 might recruit the TFIID complex and RNA polymerase II via an intermediate tethering factor. Alternatively, the
TATA binding component of the TFIIID complex might interact with sequences around position -30, even in the absence of TATA motif, although with a low-affinity, and SP1 might help to stabilize this interaction. The region surrounding the transcription start site may be bound by nuclear proteins that recognize an Initiator (Inr) element; (for reviews, see Weiss and Reinberg). Although there is no universal consensus sequence motif for Inr, some sequence similarities exist, and they have allowed grouping of distinct Inr sequences. The human uPAR transcriptional initiator site exhibits a partial match to the DHFR Initiator element that binds a factor termed HIP1 (house keeping initiator protein). The simultaneous binding of HIP1 and of SP1 has been shown to be sufficient for accurate transcription initiation from the DHFR gene. We have investigated the binding of HeLa nuclear proteins to uPAR promoter.

Our data indicate that SP1 binds to this region as confirmed by DNase footprinting with purified SP1 and HeLa nuclear extracts. The role of SP1 in promoter activity was also confirmed by cotransfection of Drosophila Schneider cells (which lack SP1), showing that the effect was dependent on the number of SP1-binding sites in uPAR-reporter gene. At least another factor, distinct from SP1, binds in proximity of the transcription initiation site (sequences -44 to -6) of both human and murine promoters. uPAR is transcriptionally induced by PMA and TGFβ1 treatment in human U937 and A549 cells respectively. Moreover, uPAR levels appear to depend on the level of expression of uPA. Therefore, it is interesting to note the presence in the human and mouse proximal promoter of consensus binding sites for transcription factors AP1, PEA3, and NFkB. It is possible that promoter specificity, inducibility, and strength are generated
by appropriate interactions between multiple DNA-binding proteins within a short proximal regulatory sequence. Some of these sites may confer cell type-restricted expression, whereas others may be active ubiquitously or confer inducibility. Such complexity could account for different modulatory effects of hormones, growth factors, and cytokines in different cell types during pathologic or physiologic conditions requiring turnover of the extracellular matrix.

ACKNOWLEDGMENT

The authors are grateful to Ann Roldan for the initial isolation of plages carrying human uPAR sequence and to Karin Holm for help with the sequencing.

REFERENCES


4. Conese M, Olson D, Blasi F: Protease ninxin-1-urokinase complexes are internalized and degraded through a mechanism that requires both urokinase receptor and α2-Macroglobulin receptor. J Biol Chem 269:17886, 1994


42. Weis L, Reinberg D: Transcription by RNA polymerase II: Initiator-directed formation of transcription-competent complexes. FASEB J 6:3300, 1992
53. Means AL, Farnham PJ: Transcription initiation from the dihydrofolate reductase promoter is positioned by HIP 1 binding at the initiation site. Mol Cell Biol 10:653, 1990
A conserved TATA-less proximal promoter drives basal transcription from the urokinase-type plasminogen activator receptor gene

E Soravia, A Grebe, P De Luca, K Helin, TT Suh, JL Degen and F Blasi