The Murine Urokinase-Type Plasminogen Activator Gene†

Sandra J. Friezner Degen,‡ Janice L. Heckel,‡ E. Reich,§ and Jay L. Degen*,‡

Division of Basic Science Research, Children's Hospital Research Foundation, Cincinnati, Ohio 45229, and Department of Pharmacological Sciences, State University of New York, Stony Brook, New York 11794

Received May 29, 1987; Revised Manuscript Received August 14, 1987

ABSTRACT: The murine urokinase-type plasminogen activator (uPA) gene has been isolated from a BALB/c liver DNA cosmid library and its nucleotide sequence established. The gene is organized into 11 exons comprising 34.7% of the 6710 base pair (bp) region spanning the interval between the presumed transcription initiation and polyadenylation sites. The transcription initiation site is flanked by common RNA polymerase II promoter elements, including a TATA box and a potential transcription factor Sp1 binding site. A large polypurine tract of the structure $(AG)_{22}(AGGG)_{16}(AG)_{28}$ is located 79 bp upstream of the 5′-terminus. It was highly sensitive to the single-strand-specific nuclease S_1, suggesting a non-B-DNA conformation of unknown significance. Consistent with the well-documented influence of adenosine cyclic 3′,5′-phosphate (cAMP) on uPA gene expression, there is a dodecanucleotide homologous to proposed regulatory sequences identified in other cAMP-modulated genes. Comparison of the murine uPA gene to the previously described porcine and human uPA genes revealed an unusually high degree of evolutionary (interspecies) sequence conservation that was not limited to exons but included introns and flanking sequences as well.

Urokinase-type plasminogen activator (uPA)[1] is one of two known mammalian serine proteases that convert plasminogen to the active protease plasmin. There is considerable evidence that plasminogen activators participate in a variety of processes that involve extracellular proteolysis. Most of these can be viewed as examples of tissue remodeling and cell migration, and they include mammary gland involution (Ossowski et al., 1979), ovulation (Beers et al., 1975), trophoblast implantation (Strickland et al., 1976), angiogenesis (Gross et al., 1983), hormone processing (Virji et al., 1980), inflammation (Vassalli et al., 1977), spermatogenesis (Fritz et al., 1982), and tumor cell metastasis (Ossowski & Reich, 1983; Mignatti et al., 1986). The regulation of uPA production is of unusual interest both because of the diversity of biological context in which it occurs and because uPA production is influenced by many agents, including modulators of intracellular cAMP (Degen et al., 1985; Nagamine & Reich, 1985; Nagamine et al., 1983; Dayer et al., 1981), tumor promoters (Degen et al., 1985; Belin et al., 1984), growth factors (Grimaldi et al., 1986), glucocorticoids (Medcalf et al., 1986), and oncogene products (Unkeless et al., 1974; Miskin et al., 1978). In several systems, it has been shown that changes in uPA gene transcription can account for alterations in cellular uPA synthesis (Degen et al., 1985; Grimaldi et al., 1986), although the underlying mechanism(s) remain(s) to be defined. Because of the ease of experimental manipulation and the large body of available genetic information, murine systems have been favored in many investigations either of the biological roles of uPA or of tissue- and/or cell-specific patterns of uPA expression [for a review, see Danø et al. (1985)]. To provide a basis for detailed studies at the molecular level, in this paper we report the isolation and complete nucleotide sequence of the murine uPA gene.

Materials and Methods

Screening of the Mouse Cosmid Library. The construction, storage, and initial screening of the mouse BALB/c liver DNA cosmid library have been described previously (Steinmetz et al., 1985). Clones containing the uPA gene were identified by in situ hybridization using a probe derived from a previously described porcine uPA cDNA clone (Nagamine et al., 1984). A 1085 bp internal *Pst*I fragment which codes for amino acids 104 through the final amino acid 422 and also contains 128 bp of 3′-noncoding sequence was isolated, labeled by nick-translation using deoxynucleoside [α-^{32}P]triphosphates, and used in hybridization mixtures (Steinmetz et al., 1985) at a concentration of 0.5×10^6 cpm/mL (specific activity $\sim 10^8$ cpm/μg). Positive bacterial isolates were retested by in situ hybridization at 60 °C in hybridization mixtures containing 6 × SSC [1 × SSC = 15 mM sodium citrate (pH 7.0)/0.15 M NaCl], 0.04% ficoll 400, 0.04% bovine serum albumin, 0.04% poly(vinylpyrrolidone), 1 mM EDTA, 0.5% SDS, and heat-denatured ^{32}P-labeled cDNA probe. The filters were washed at 60 °C in 6 × SSC containing 0.5% SDS and exposed to Kodak X-Omat AR5 film overnight at −70 °C with intensifying screens.

Cosmid DNA Sequence Analysis. The methods used to prepare and isolate plasmid subclones containing cosmid-derived fragments, to map restriction endonuclease cleavage sites, and to sequence DNA fragments by the Maxam and Gilbert chemical cleavage technique have all been described in detail (Degen et al., 1986). DNA sequences were analyzed by using the programs of Queen and Korn (1984) on an IBM-AT computer.

Identification of Transcription Initiation Site. Oligonucleotide primer extension analysis was performed to determine the 5′-terminus of murine uPA mRNA. Primer, complementary to 30 nucleotides of the murine uPA mRNA,

† This study was begun at the Friedrich Miescher Institut, Basel, Switzerland, and completed at the Children's Hospital Research Foundation, Cincinnati, OH. The research was supported in part by National Institutes of Health Grant CA44611.

* Correspondence should be addressed to this author.

‡ Children's Hospital Research Foundation.

§ State University of New York.

[1] Abbreviations: uPA, urokinase-type plasminogen activator; PEPCK, phosphoenolpyruvate carboxykinase; cAMP, adenosine cyclic 3′,5′-phosphate; bp, base pair(s); kb, kilobase(s); SDS, sodium dodecyl sulfate; EDTA, ethylenediaminetetraacetic acid; Tris-HCl, tris(hydroxymethyl)aminomethane hydrochloride; tPA, tissue plasminogen activator.

0006-2960/87/0426-8270$01.50/0

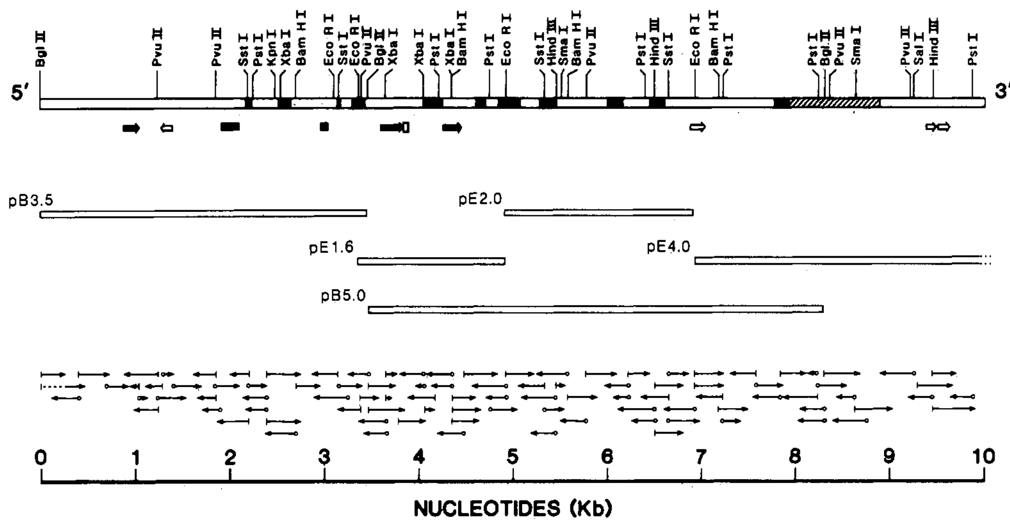

FIGURE 1: Organization of the murine uPA gene. A partial map of restriction endonuclease cleavage sites is presented. *Filled areas* and *hatched areas* indicate the protein-coding and -noncoding portions of exons, respectively; *unfilled areas* represent introns and flanking DNA. (A) Placement of repetitive DNA. B1 and B2 family repeats are represented by *open* and *solid arrows*, respectively; the arrows indicate the direction of transcription from the putative polymerase III promoters of these repeat units. Polypurine repeats and alternating purine/pyrimidine sequences are illustrated by *solid and open boxes*, respectively. (B) Fragments subcloned to facilitate analysis of the gene. A series of overlapping *Bgl*II and *Eco*RI subclones were generated (the *Eco*RI site at the 3′ flank of pE4.0 was cosmid derived) and named according to their approximate size in kilobases. (C) Sequenced fragments of the gene; *bar-tailed arrows* indicate regions sequenced on the mRNA-like strand, and *circle-tailed arrows* indicate regions sequenced on the complementary strand. The sequence was not read in the dashed region.

was prepared by using an Applied Biosystems synthesizer, 5′ end labeled with [γ-^{32}P]ATP (3000 Ci/mmol, New England Nuclear) and T_4 polynucleotide kinase, and purified by electrophoresis on a 12% denaturing polyacrylamide gel (Maxam & Gilbert, 1980). Gel autoradiography was performed to localize the 30-mer, and a gel fragment containing the primer was excised; it was washed briefly in water and broken into small fragments, and the oligomer was eluted overnight at 37 °C in 0.5 M ammonium acetate containing 1 mM EDTA. The eluted oligonucleotide was finally purified by chromatography on a NENSORB 20 cartridge (New England Nuclear). Total cellular RNA was isolated from mouse MSV-3T3 fibroblast cultures treated with 160 nM phorbol myristate acetate for 15 h (phorbol ester treatment raised cellular uPA–mRNA levels from ~50 molecules/cell to ~500 molecules/cell; our unpublished data) by sedimentation through CsCl (Degen et al., 1985). Poly(A+) RNA was prepared by oligo(dT)–cellulose chromatography (Aviv & Leder, 1972). Hybridization mixtures (30 μL) were prepared containing 10 μg of RNA, 300 000 cpm of ^{32}P-labeled oligonucleotide (specific activity 1500–3000 Ci/mmol), 20 mM Tris-HCl (pH 7.5), 0.6 M NaCl, 4 mM EDTA, 0.1% SDS, and 10% formamide and then incubated for 15 h at 68 °C. The nucleic acids were precipitated by the addition of 2.5 volumes of absolute ethanol, dried briefly under vacuum, and then redissolved in 20 μL of reverse transcriptase mix [100 mM Tris-HCl, pH 8.3, containing 100 mM NaCl, 10 mM dithiothreitol, 10 mM $MgCl_2$, 1 mM dATP, 1 mM dCTP, 1 mM dTTP, 1 mM dGTP, 600 units/mL human placental RNase inhibitor (Bethesda Research Laboratories), and 750 units/mL AMV reverse transcriptase (Life Sciences Inc., St. Petersburg, FL)]. Reaction mixtures were incubated for 1 h at 37 °C, and the nucleic acids were then ethanol precipitated as above. Precipitates were dissolved in 2 μL of DNA sequencing gel sample buffer and analyzed by electrophoresis on a 6% sequencing gel and autoradiography (Maxam & Gilbert, 1980).

S_1 Nuclease Sensitivity of Polypurine Sequences. S_1 nuclease hypersensitive sites were mapped by the indirect end-labeling method of Wu et al. (1980). DNA samples (2 μg) were incubated at 37 °C for 30 min in reaction mixtures (50 μL) containing 30 mM NaOAc (pH 4.2), 0.3 M NaCl, 3 mM zinc acetate, and 0–20 units of S_1 nuclease (Sigma). The nucleic acids were ethanol precipitated, digested with *Bgl*II, and analyzed by agarose gel electrophoresis. The DNA fragments were transferred from the agarose gel to nitrocellulose (Southern, 1975) and hybridized (Degen et al., 1986) with probes (labeled with deoxynucleoside [^{32}P]triphosphates by nick-translation) complementary to regions adjacent to the *Bgl*II cleavage sites (for further details, see Figure 5).

RESULTS

Isolation of the Murine uPA Gene. A BALB/c liver DNA cosmid library [generously provided by Dr. M. Steinmetz; see Steinmetz et al. (1985)] was screened for the uPA gene using a porcine uPA cDNA hybridization probe (see Materials and Methods). Three positive colonies were identified from among the (6–9) × 10^5 cosmid-containing clones analyzed, and one cloned isolate was selected for further evaluation. The portion of the cosmid (cosMuPA-1) which hybridized to the porcine cDNA probe was characterized by digestion with selected restriction endonucleases and Southern blot hybridization. This isolate proved to harbor an insert that included the entire uPA gene. A detailed map of restriction endonuclease cleavage sites is illustrated in Figure 1.

Nucleotide Sequence of the Murine uPA Gene. To simplfy the characterization of the uPA gene, a series of overlapping *Bgl*II and *Eco*RI fragments that encompassed its sequence was subcloned into the pNNL vector (Grosveld et al., 1982). The sequencing strategy is given in Figure 1, and the nucleotide sequence is presented in Figure 2. Of the 9950 nucleotides sequenced, 56% were determined for both strands, and 86% were established at least twice. A comparison of the data in Figure 2 with those previously reported for a murine uPA cDNA (Belin et al., 1985) confirmed the presence of the uPA gene in cosMuPA-1 and established the gene organization (see Figures 1 and 2); it also identified four differences between the gene and cDNA sequences: three of these were in the coding region, but they did not modify any amino acid, while the fourth affected the length of a poly(T) tract in the 3′ noncoding region (see legend, Figure 2). It is unclear whether

```
TATGGAATCT CTGATTCAGT GGTCCTGTGG GTTGCTAGCA GCTCGGCTGC GCAAGGAAAA CGTATTCTGA AGAACGATGG TCACCTGCCT ACCCAAAGCT  -2082
GGGTATTCAA TGTGTACTTT CCTATCCAGA GGTTGGCATT CTGGGCCACT GGCTGGGGTA AAGTCAAGCA GCCTCCTCCT TCCTACCTCC TGGCATTCTT  -1982
TTCCAAGGTC CAGGTTGACA GTAAAATGTT GTAGGCTCAG GCTTATGGAA ATAAAGCAAA GCTGGGGCAC TAATGAATGG TGCGGGCTCA GGCACCTCTT  -1882
CTGTGGAAGG AGCATCAAAA TGACCACTGG TCCTTGGCTC TGAGGCACAT GTTGCAGTGA CTGACCCTTG GAATCTTTTC TGAGCACTCC CTACTTTGCC  -1782
TGAGGGGTGT GTGCATGTGT GCATGTGTGC AAGTGTGTGC ATGAGGGAGG GTGCTTTGTC TTCAAGATGT TCAGGGCTCA TTTACACATG ACCTGTCTTT  -1682
ATATAAGCCT ATTTTATTAG ATAAATTATT AGATAAATAA TTGTTTTCGT CTTAACATGA TGATGTAAAA TGCTGAACAC AAGTGCATTT ACAGTGTGGA  -1582
ATTGGCAACA GAAGAGATGG CTAGGCTCAC ACCCAAAAAT AAAATGTGTA ATTATTTTTT TAGAATGACT GATACGAGGA CGTGTTTAAA GTTTTTAAAA  -1482
CAGTGTGAAA TAGTATTCAG TGAAAAATAA ACATTCCTTC ATCTCTTCCT TCAGACCCAC AGCCTCTCCA CTACAAGGGA AGAGCAGACA CCCCCACTGT  -1382
TGTTGCATGT TGATAGCCCA CACTAACGAG GTTTGAGTAT TCCTTGCTTT TTTTTCCTAT CCTGTTACAA TGTACTGTTC ATGTTACTGT TCACATCACT  -1282
GACTGCTCTT CCAGAGGTCC TGAGTTCAAT TCCCAGCGCC ACATGGTGGC TCATGACCAT CTACAATGGG ATCTGGTACT CTTTTCTGGT GTGTCTGAAG  -1182
AGAGGACTGT GCACTCGCAT ACATAAAATA AATAAATCTT TAAAAAAAAG ACACCCCCCT TGATTGCTTA TTTGTGTCTG TGTGCAACCA CAATACCAGT  -1082
GAGGTATGCA CTCAGTCCTG ATCTATGATC CTCTGTATTG GTTGTATTCT TTTCCAGACA AATAAATAGA TTGTTCATTA TAGGCATTTT TCAAAAGACC  - 982
CATTGACACT TCCATAGCAG CTGCTGTGCC AAGACTTTGC TCCCACAATC CTAACATAGT AACCATTACT GCTTTTTTAA AAATCATAAT TATTTTTGAG  - 882
GATTTCACTA TGTAAATTAG TCTGGGCTTG AACTCGCAGA GATCCACCTG CCTCTTCCTC CCTAGTGTTT AGATAAAAGA CGTGCACCAT TACGTCTGGC  - 782
CCTTGCCATA AATTTTAAAA TTGTCTTTCA AAGATTGACC TTAAACCAAA CAGCAAATCT GAATAAAATC TGACCTTGCC TTTCACCTTC TGATGACAGT  - 682
ATTACACTCC CTGACGACAA ACTTCACTCT TGTCTTCTGA TTCACTGCTT GCATTAGTGG CATTTGGAGA ACTCAGCATT TGACATGTGG GAGCCTTTGT  - 582
TAGTAGGTAT TTTTTATTGT AAAGGAACTG CGACTTATAC CCCTCTATCA GACATCTGAA TCAGTGTCGT AGGCAGGTAG GGGGACAGGT TGGAGAAGAA  - 482
CTGATTAAAC CACTAAGGAA GAGGCTTGAG ATCAGCGAGC CAATGGCTGG AGCGCCTTCA CCACGCTACA CTGGGGCCGC ACTAGGTGAA TGAAAGAAAG  - 382
GAAGAATGTT CAAGCCGCCG GATCATCGCT CGATCCAGAC AGACTGCGTT AAGTTTGCTC AGCTGAAATT CCGTGACTTC GTCAAAGTTG GGAAGCAAGC  - 282
GCGGTCCAGT TGCGGTGGGA TGCAGGAAAA GGAAAAGGAG AGAGAGAGAG AGAGAGAGAG AGAGAGAGAG AGAGAGAGAG AGAGGGAGGG AGGGAGGGAG  - 182
GGAGGGAGGG AGGGAGGGAG GGAGGGAGGG AGGGAGGGAG GGAGGGAGAG AGAGAGAGAG AGAGAGAGAG AGAGAGAGAG AGAGAGAGAG AGAGAGAGAG  -  82
AGCCGCCCTC CAGGGAACCT GGGCGGGGCC AGGGCTCTGG CGGGCCCTAA TAAAGGGCGA GCAGCGCCGA GCAGAGCCTG TAGCCCCAGA GCTCTGTCTG     19
TCATCCAACC AGTCCTTGCG TGTCTGCCAG CGCCCTTCCG CTGCAGTCAC CGGTGAGTGC TGTTGGTCTG AAGCAAGCCT GGCGGGATGA GGCAGCCAGG    119
GCTCCCGCAT GCCTCCCTTC CCCCTACCTT GGCTGGCGGA ACTGTGGGCA AGGTCACCAC TCCAGCCCTT CGCGCCCCTC TACAGAGAGG TTCCATGGTG    219
TTGTGCGGAT TCAGAGCCCG CAGAGGGGAG AGACTGCCCG GCTTGGGGAG GTTGGTCACT GATGGCTTGC CCCGCAGGGT ACCTGGAGTG GCTTCCTTCC    319
CTTGGTCTGT AGTAACTCTG CCACCTTCGA GCTGCTCCGC TTCTTGTCCT GACTTCTCCT TCCTTTGCAG AACTGCTGTC TAGAGCCCAG CGGCACTACC    419

MetLysValT rpLeuAlaSe rLeuPheLeu CysAlaLeuV alValLysAs nSerGlu
ATGAAAGTCT GGCTGGCGAG CCTGTTCCTC TGCGCCTTGG TGGTGAAAAA CTCTGAAGTG AGTGGTCTCG CTGCTTTAGC ACCATCAGGA AGGGGCTTGC    519
AGGATCCCTT AAGCAGCATC AGGGGAAAAA TGGGGGCTGC ACGGGGAACT TAGGCATCAA AGGCAGGTCC AGGCTTTCCC AGGAAATAGG ACAATGTATC    619
AGTGGAGGGC TTGTGCACCC AAAGAGGTTT GCACTATCTG GCAAGGGAGG AAGAAGCCAC GGGGAGTACC TTAGCCCAAG GGCACCTGGT TTGTGTGAAG    719
TTTGCTTAAG TCAGTCCATG TCTGGGTGCT GGCTAGGAAT AAACAGAAAG GGGAGAGACA GACAGGGGTG GGGTGGGAGA AAGAGAGAGA GAGAGAGAGA    819
GAGAGAGAGA GAGAGAGAGA GAGAGAGAGA GAGAGAGAGA ATATTATGAG TGAATGAATA TCACTGGAAG GGATTTTGAG GTGGGGACCT GTTTATCCTG    919

                                             GlyGlySer VaILeuGlyA laProAspGl uS
AACATGAATT CCCTAGAGCA TGTCACCTTC ATCTCTTGCA GGGTGGCAGT GTACTTGGAG CTCCTGATGA ATGTGAGTAT CTGCTTCCTT GCACAATAGT   1019
TGGCTGCACA GAGACCCTTG AAAAACCTTA GGAGACATAC CCTCCCTGTC CCTGCTGAAA GGCTGGCTCC CCACTTGATC CTTGCTTACC CCTCCTTTGC   1119

         e rAsnCysGly CysGlnAsnG lyGlyValCy sValSerTyr LysTyrPheS erArgIleAr gArgCysSer CysProArgL ysPheGlnGl
ATTTGCTAGC AAACTGTGGC TGTCAGAACG GAGGTGTATG CGTGTCCTAC AAGTACTTCT CCAGAATTCG CCGATGCAGC TGCCCAAGGA AATTCCAGGG   1219

yGluHisCys GluIleA
GGAGCACTGT GAGATAGGTA TGGGGATTTG GACTTGGAAT GTGGGAGTGG GGGAGGACCA GAGATCTTAG AACAGGGACA GATGGGTGGG ATGCAGAAGC   1319
AGGCAGAAGC TGGCCTTGGA GGTGTGGGTC TGTGAGCCCA GCACTTAGGA GGAGACTGAA GCCTGGCCTC CATAGTAAGT CCTTGTCTCA AAAGGCGGGC   1419
GAGGCGCAGC TAGAGAGACA GGTGAGTGAT TAAGAACACT GGCTGCTCTT CTAGACATCC TATAGTTTGA GTTTCAGCAC CCACAGAGGT GGGTTACAGC   1519
CATCTGTACC CCCAGTCCCA GGGAATCTGA TGCCCTCTTC TGACCTCTTC CAGCCCAAGT GACATACATG GTGTACAAGT ATACATAAAG GAAAAACACT   1619
CATACACATA AAATAAGCAA ACAAACAAAC AAAACAGGCA GCAGAAGTTG GGAGTCACAC ACACACACAC ACACACACAC ACACACACAC ACACACACTC   1719
ATGAAGCAGT GGCTATACAA GTGTGAAAAA AAGGGTGAAT CTCCCTCATA TCACCTGACA GGTCTGAAAC CGTGTCACCT CTGAAATGCC TGTCCAAACC   1819

                                    spAlaSe rLysThrCys TyrHisGlyA snGlyAspSe rTyrArgGly LysAlaAsnT
TCATCCCTTT TCTAATACTC TGCACTCCTC AAATCATTTC TAGATGCATC AAAAACCTGC TATCATGGAA ATGGTGACTC TTACCGAGGA AAGGCCAACA   1919

hrAspThrLy sGlyArgPro CysLeuAlaT rpAsnAlaPr oAlaValLeu GlnLysProT yrAsnAlaHi sArgProAsp AlaIleSerL euGlyLeuGl
CTGATACCAA AGGTCGGCCC TGCCTGGCCT GGAATGCGCC TGCTGTCCTT CAGAAACCCT ACAATGCCCA CAGACCTGAT GCTATTAGCC TAGGCCTGGG   2019

yLysHisAsn TyrCysAr
GAAACACAAT TACTGCAGGT AGGTGGTGAC TGAGTACCAA GAATCCTTCC CAAGGGGGAT AGGGAGGTGG CTCAGCAGTT AAGAGCACAG ACTGCCTTTC   2119
CAGAGCACCT AGTTTGATTC CCAGGGCAGC TCGTGACAGT CTTTAACACC TGTTCTAGAG GATCCGATGC CCTCTTCTGG CCTCACTGGG CAGGCATGCA   2219
CTGTCATGAA TATAGGAAAA CACTTATACA CATTAAAAAC AACATCCCTT CCCCCATCGT GGCCTCTTAG AAACCTTTGT TATCACCATG GTATACCTGG   2319
GATGGGAATC CTGGCACAAG AATCCAGGTC TCTGGTTGAG CCTTTGTTGG AAGGGAGGAT ACAGAGAAGA CATTCGGGCT TGGCATGACA TTCCCTATCT   2419

               gAsnPr oAspAsnGln LysArgProT rpCysTyrVa lGlnIleGly LeuArgGlnP heValGlnGl uCysMetVal HisAspCysS
CTTTGTGTTA CCAGGAACCC TGACAACCAG AAGCGACCCT GGTGCTATGT GCAGATTGGC CTAAGGCAGT TTGTCCAAGA ATGCATGGTG CATGACTGCT   2519

erLeuS
CTCTTAGTGA GTGTCGCTGA CTGCTTATGA CAACGGGGTG GGAAGAGACA AACTCTATTG TCACTGCAGG AGGGATGAGA AGTGAGGTTG GCCTCAGAGA   2619

                                                     e rLysLysPro SerSerSerV alAspGlnGl nGlyPheGln CysGlyGlnL
CTCTTCATCA TTGCTGTCTC CCCCAAACAT GTGTCTCTTT CTTTTCTAGG CAAAAAGCCT TCTTCGTCTG TAGACCAACA AGGCTTCCAG TGTGGCCAGA   2719

ysAlaLeuAr gProArgPhe LysIleValG lyGlyGluPh eThrGluVal GluAsnGlnP roTrpPheAl aAlaIleTyr GlnLysAsnL ysGlyGlySe
AGGCTCTAAG GCCCCGCTTT AAGATTGTTG GGGGAGAATT CACTGAGGTG GAGAACCAGC CCTGGTTCGC AGCCATCTAC CAGAAGAACA AGGGAGGAAG   2819

rProProSer PheLysCysG lyGlySerLe uIleSerPro CysTrpValA laSerAlaAl aHisCysPhe Il
TCCTCCCTCC TTTAAATGTG GTGGGAGTCT CATCAGTCCT TGCTGGGTGG CCAGTGCCGC ACACTGCTTC ATGTACGTCC ATCCCTTTGT CCCTTCTCTC   2919
TGACTCTTCC ACCCAACCCC AAGACTGTCC TTCCTCCTTC CCTATGGACG GTTACAATGT CATTCTCCTG CTAACCCTCT AACCATGCAG CTTGTGGTCT   3019

                                                                                                     eGlnLeuP
TGGGTACAAG TAATACTTTG AGGCCTCTGG GGTGGAGTGG AGAGAGTGAC CGGACTTTGT GAGACCAGGC TGACATGTTT CATTTCTCAT AGTCAACTCC   3119

roLysLysGl uAsnTyrVal ValTyrLeuG lyGlnSerLy sGluSerSer TyrAsnProG lyGluMetLy sPheGluVal GluGlnLeuI leLeuHisGl
CAAAGAAGGA AAACTACGTT GTCTACCTGG GTCAGTCGAA GGAGAGCTCC TATAATCCTG GAGAGATGAA GTTTGAGGTG GAGCAGCTCA TCTTGCACGA   3219

uTyrTyrArg GluAspSerL euAlaTyrHi sAsnAspIle A
ATACTACAGG GAAGACAGCC TGGCCTACCA TAATGATATT GGTGAGCAGA AAGCTTAGTT ATCAGAAAGG CTAAAGTAGT GGTGGGAAAT GTTGGGGGAC   3319
```

```
TTGAAGCCCG GGATTTATAT AACGAGACGG ATGAGGAAGA GTGCAGAATG AGATACATGA GAAGCTGAGG GGTGTGGGGA TCCTCTGTGG AGACCTTGAA   3419
TTTCCCAAAC AGATAGATTC TTCTAAGTAG AAACAATCTT ACAGGCATAC GGCTTAGGCT GAGAATGCCC TGTTTGTACA AAGTAGGATG GATGCTTCTT   3519
CTCTGTATAC CAGAATATAG AAGGTATAAA GCAAAGCCTT GGCTGGATTT CAGCTCAGCT CCCTCAGCAG GAAACAACCT GTTCAGCTGT ATATGGTAGA   3619
TTTTGTTGCC CGAACATCTG TCATCTGATG AAATAAAGCA TTTGGAGAAT GTGGCAGGGG AGGCTTCAGG GTAACAAGAT ACCAGCAGAC CTTTTGGATC   3719
TCTGTGACTC CCATGCCACG AGTATAGATC AATGCTCAGC ATTGGTAGGG GAGAGATGAT GACCATCTGA CACAGTGATA ACCTTTCCCC TTTGACCTTT   3819

                laLeu LeuLysIleA rgThrSerTh rGlyGlnCys AlaGlnProS erArgSerIl eGlnThrIle CysLeuProP roArgPheTh
CCCTTCCCCA CCCAGCCTTG CTGAAGATAC GTACCAGCAC AGGCCAATGT GCACAGCCAT CCAGGTCCAT ACAGACCATC TGCCTGCCCC CAAGGTTTAC   3919

rAspAlaPro PheGlySerA spCysGluIl eThrGlyPhe GlyLysGluS erGluS
TGATGCTCCG TTTGGTTCAG ACTGTGAGAT CACTGGCTTT GGAAAAGAGT CTGAAAGTAG TGACAGATGA AGCTCACTGA GAGAGTCTGG GGGAGTGTTA   4019
TGGTCCAGAG CAAAGAGCAG ACTATCAAAG GAAGACTGTG GAAACAGGAC TGGAAACATT ATGGAGGGCC AGGGATAGAG TAGGGGGAGA TGGGCAAGCA   4119
AGTCAAACAG GGTGTGAACA ATTGTGAGTG AAGTAAAAGA CTCAGATTGG AGAAACAAGA ACAAGAGCTT TTCATAGCTG GGATATGTTT TTTATCTTCA   4219

                                                                    erAspTyr LeuTyrProL ysAsnLeuLy sMetSerVal
CCCCTGCAGA GAGTCTCATT TATAGACACA TCTTAATGCA AACATCTGTT TGTTCCATCT AGGTGACTAT CTCTATCCAA AGAACCTGAA AATGTCTGTT   4319

ValLysLeuV alSerHisGl uGlnCysMet GlnProHisT yrTyrGlySe rGluIleAsn TyrLysMetL euCysAlaAl aAspProGlu TrpLysThrA
GTAAAGCTTG TTTCTCATGA ACAGTGTATG CAGCCCCACT ACTATGGCTC TGAAATTAAT TATAAAATGC TGTGTGCTGC GGACCCAGAG TGGAAAACAG   4419

spSerCysLy s
ATTCCTGCAA GGTAAGACTC TCAAGCACCC CTCTTTATCA CCCCAACTCC CCAGAGCTCT TGGATTTGAT CTAACAACCC TGGGGAGTCT CTTTCCAGCC   4519
AACAATCTAA GAATCAAGGA CTTAGGTCTT TGGGAGCTTG TCCCAATACT TATAGGTTCA AACGTTGGGC ATGAGTCCCT GTGCTATATG CGTTTTAGAC   4619
TAAAAGGGAC CAAGACTGCT AAAAAAAAAA TAACCCAGAC ATGGTAGAGC ATACCTATAA CCTAGCACTC TTAGCATTTT GAATGCTGAG ACAGGAGGAT   4719
CATGAGTTTA AGGCCAGCCC AAACTACAGA GTGAGAATTC AAGGCCTGTG CCACATAGCA AGATCCTTTC TCAAATAAAA CAAGGAAAGC ACAAACCATA   4819
AAAACCAAGA CAATAGCAAC AAAGGGATGG GTGCCTAGCT CAGTGGTTTA GTGCTTGCTT ACTATGCTCA AAGTCCTGAG TTCAAGTCTC AACTCAGGGA   4919
CTGGAGATCA TAAGAAAAAC TAAGAGTCTG GGGACGTGGC TTAGTTGGTA GAGTACTTAC CCAGCATGAA AGAAGCCCTG GATCCAGTCC TCGGCACTGT   5019
ATGTAATGAC CCAGGCCTGC AGTATCAGCC CTTGAGGGGG TCAGAATCAG TTTATAAGTT CTTCAGTTAC AGAGTGAGTT CAAAGCCAGC CTGAAAGACA   5119
TGGGCTTATG AGACTCTGTC CCAATCTGAA AGAACACAAC CAACCAACCA ACCACCACCA CCAACAGCAA AATATAGATA CTATTCAAAT CACTTCTGGG   5219
CCTTTGGCAA GACAAGTGAA ATCAACATAA TTCTATTGTT CAGGATCGCA GTGAATTACC AAAGATCAGG TAGGAAAGGA AGGAGAAGTC TTAAAGAGAC   5319
TATGAACTGG TAAATAAAGA GACGGAAGGA AAAAGGAAGC ATGTGTCAGT TGGAAAAAAC AAAACTAAGA CTGAGCATGC TGTGTGCCAA GGCGAGGAAT   5419
AGCAGGGTCT GGGAAAGCAC TGGAGAGTGG GAGAGGAAAG CTAAGACTTT TTACTCTTGA TTCGGTAGAA AATGGGGAGT TGCGAATGTC TCTGACTCTG   5519

                                                                                                 G lyAspSerGl
GGAACCTCTC CCCGTTCTCT CCCGTGGCTG GGTAGTGGCC CTTCCCTCAG TTCTTCCAGG GCTTCACCTC TTTATCTTTG GCTTCCCAGG GCGATTCTGG   5619

yGlyProLeu IleCysAsnI leGluGlyAr gProThrLeu SerGlyIleV alSerTrpGl yArgGlyCys AlaGluLysA snLysProGl yValTyrThr
AGGACCGCTT ATCTGTAACA TCGAAGGCCG CCCAACTCTG AGTGGGATTG TGAGCTGGGG CCGAGGATGT GCAGAGAAAA ACAAGCCCGG TGTCTACACG   5719

ArgValSerH isPheLeuAs pTrpIleGln SerHisIleG lyGluGluLy sGlyLeuAla Phe***
AGGGTCTCAC ACTTCCTGGA CTGGATTCAA TCCCACATTG GAGAAGAGAA AGGTCTGGCC TTCTGATGGC CCTCAGGTAG CTGAGGGAAG AAACAGATGG   5819
GTCACTTGTT CCCATGCTGA CCGTCCTCTC TGCAACAGAG TCGTCAAATG GAGGGAAGAA GCTGAAAAGA CAGGTTTTGC ATTGATCCTC TGCTGTGCTG   5919
CCCACCAGGG TGAGCGCCAA TAGCATTACC CTCAGACACA GGCCTGGGTG CTGGCCATCC AGACCCTCCC GACCAGGATG GAAAGTTGGT CCTGACTCAG   6019
GATGCTATAG ACCAGGAGTT GCCTTTTTAT GGACTAAAGC CATCTGCAGT TTAGAAAACA TCTCCTGGGC AAGTGTAGGA GGAGAGCTGT TTCCCTTAAT   6119
GGGTCATTCA TGAGATCTGC TGTTGGGAAA TAAATGATTT CCCAATTAGG AAGTGCAACA GCTGAGGTAT TGTGAGGGTG CTTGTCCAAT ATGAGAACGG   6219
TAGCTTGAGG AGTAGAGACA CTAACGGCTT GAGGGAACAG CTCTAGCATC CCATGAATGG ATCAGGAAAT GTTATATTTG TGTGTATGTT TGTTCACTCT   6319
GCACAGGCTG TGAGTATAAG CCTGAGCAAA AGCTGGTGTA TTTCTGTATC TAACTGCAAG TCTAGGTATT TCCCTAACTC CAGACTGTGA TGCGGGGCCA   6419
TTTGGTCTTC CATGTGATGC TCCACGTGAA TGTATCATTC CCGGGCGTGA CCCGTGACTA GCACTAAATG TCGGTTTCAC TTTTTATATA GATGTCCACT   6519
TCTTGGCCAG TTATCTTTTT TTTTTTTTTT TTTTTTTTTT ACTAATTAGC CTAGTTCATC CAATCCTCAC TGGGTGGGGT AAGGACCACT TCTACATACT   6619
TAATATTTAA TAATTATGTT CTGCTATTTT TATTTATATC TATTTTTATA ATTCTGAGTA AAGGTGATCA ATAAATGTGA TTTTTCTGAA GATTCTGGTT   6719
TCTCCATGAT TCTTGTGTGA CAGGGAAGAG GGGGACATTA AAAGGAAGAA AATAATGAGG GCTACGTGCA TCTTAGTTTC ATTTGGGGTT TGCTTGGACT   6819
TTTTTTGGAT GAGAATGCAT GGATGAGGCT GCTGATCCAA GCCAGGCACG GTCCTAGTCC ACCTGAAGGC TAAATGAAGA TTGGTGCAAA TTCAAGGTCA   6919
GCCTGACGAT GTGGTTATTT CAAGGCCAGC TAGGCTACAT AGCAAGACAT TGTCTTTAAA AAAAAATGCG CAAGAAAGAA AAGAAAAAAA TCTGATTCAA   7019
ACAAAGCAGC TGAGTCGGTG CTGTCGACGG GGTCAGGTAA TGAAGATACT TGTGTTTGCA GCTCTTGGTC CCCCGCTGAA ACTACTTGTA ACGCTTCTGG   7119
CCTCTGTAGG CACCAACACC CATGCACACA CACAGATGAT TACAAATAAG TCTTACAGAA GAAAACATGA AAAAAATCAG TGTCTCACAC CTGTCATCCC   7219
AGCAAGTGAG AGGCTGAGGC AAGAAGACTC CTGTGAGTTT GAAGCCAATT TTGCTACAAA GCTTTAGTCT TAAAACAGAC AAAATAAAAC AAAAAGTGGG   7319
GTGGTAGTGG TATGCCTTTA ATCTCAGCAG AGGCAGAGGT TCGAGGCCTG ACTTGTCTAC AGAGTGAGTT CAGGACAGCC AGGAGCTACA CATAGAAACC   7419
TTGTCTCAAA ATAACAATAA AATAATAATA ACAACAAAAC CAATAAAACT AAACCATTGT GAATCTGGGA TTCCAGAAAG CAAACATACT TTTCCATCAT   7519
CTGTGTGTAG GCTGATGCTA AATTTCCGCT GTGCTAATGG AGCTTATCTG CACTTAATGT GGCCTTGGGA AGGTACAGAA GGAGAGTTCC AGGGTTGGCC   7619
TTCATAGCAC CTAAGTTACA AAACAGGCCA CAGGCTGCGG CTTGGTAAGC GGTGTTCGGG TTGAGCTGCA GCTCACAGGT GCTTCCTCAG CCTGGTGCTA   7719
TTGGGCAGAG TACCTCGTTT ATTATTAATT AATTAATTAA TTAATTAATT                                                         7769
```

FIGURE 2: Nucleotide sequence of the murine uPA gene and its 5′- and 3′-flanking regions. The site of initiation of gene transcription has been defined as nucleotide +1. The site of polyadenylation is nucleotide 6710. The 2181 nucleotides of 5′-flanking sequence presented are numbered −1 to −2181. The numbers to the right of each line correspond to the terminal nucleotide of that line. The deduced amino acid sequence is given above the DNA sequence. Exon 1 and the portions of exons 2 and 11 which are not protein coding are underlined. The termination codon is shown with three asterisks. A^{3860}, T^{4316}, T^{4319}, and $T_{25}^{6535-6559}$ in the gene sequence were reported as G, C, C, and T_{39} in a cDNA (Belin et al., 1985).

any of these changes might be due to polymorphisms or cloning artifacts.

Gene Organization. The size of the murine uPA gene, from the site of transcription initiation (see below) to the polyadenylation site, was 6710 bp. Like the porcine and human uPA genes (Nagamine et al., 1984; Riccio et al., 1985; see Discussion), the murine gene was organized into 11 exons separated by 10 intervening sequences. The 5′ noncoding sequence was interrupted by the first intervening sequence, whereas the 3′ noncoding region was entirely contained in exon 11 (see Figures 1 and 2 and Table I). The exons ranged in size from 31 to 1102 bp and together totaled 2382 bp (34.7%) of the gene sequence; the intervening sequences ranged from 137 to 1178 bp in length and totaled 4382 bp (65.3%) of the gene (see Table I). All exon–intron splice junctions followed the GT-AG rule (Breathnach et al., 1978) and agreed with the consensus sequences proposed by Mount (1982).

Site of Initiation of Transcription. The site of initiation of transcription was determined by primer extension analysis using a synthetic oligonucleotide primer and poly(A+) RNA isolated from phorbol ester induced MSV-3T3 mouse fibroblasts (see Materials and Methods). The primer was a 30-nucleotide DNA fragment complementary to the region numbered +41 to +70 in Figure 2; it was 5′ end labeled with [γ-^{32}P]ATP and T_4 polynucleotide kinase, hybridized to poly(A+) RNA, and extended with reverse transcriptase (see

Table I: Comparison of the Murine, Human, and Porcine uPA Genes

	size (bp)			sequence identity (%)[e]					
				mouse/human		mouse/porcine		human/porcine	
exon/intron[a]	mouse	human[c]	porcine[d]	DNA	protein	DNA	protein	DNA	protein
1	71	88	85	55.7	*f*	48.2	*f*	55.7	*f*
2	87	88	88	71.6	52.6	73.9	52.6	85.2	78.9
3	31	28	34	32.3	20.0	24.3	16.7	67.6	54.5
4	108	108	108	79.6	69.4	88.0	86.1	88.9	77.8
5	175	175	175	82.3	72.9	77.1	71.2	85.1	83.1
6	92	92	119	84.8	73.3	81.5[g]	76.7[g]	84.8[g]	83.3[g]
7	223	220	220	74.4	70.7	73.5	68.0	87.3	85.1
8	149	149	149	71.8	63.3	72.5	61.2	81.2	77.6
9	141	141	141	77.3	68.1	76.6	76.6	80.1	68.1
10	149	149	149	80.5	74.0	78.5	78.0	93.3	96.0
11	1102	1106	1119	69.1	75.9	68.9	70.7	77.3	79.3
total exon	2328	2344	2387						
A	318 (–)[b]	306 (–)	329 (–)	55.9		61.9		64.2	
B	484 (0)	417 (0)	452 (0)	56.4[h]		59.0[h]		68.6	
C	137 (I)	146 (I)	157 (I)	66.2		62.3		62.4	
D	626 (I)	603 (I)	329 (I)	57.8[i,j]		60.8[i]		65.3[j]	
E	396 (II)	193 (II)	187 (II)	66.3[k]		59.5[k]		71.7	
F	143 (I)	157 (I)	161 (I)	63.1		48.5		49.4	
G	220 (II)	221 (II)	225 (II)	63.3		63.4		71.1	
H	574 (I)	666 (I)	644 (I)	58.8		55.8		63.4	
I	306 (I)	346 (I)	326 (I)	59.4		58.2		65.0	
J	1178 (0)	989 (0)	655 (0)	57.8[l,m]		55.3[l]		71.2[m]	
total intron	4382	4044	3465						
total[n]	6710	6388	5852						

[a] Exons are numbered, and introns are lettered. [b] Roman numerals in parentheses indicate intron placement; introns occurring within noncoding sequence, between codons, between the first and second nucleotides of a codon, and between the second and third nucleotides of a codon are labeled with a dash, 0, I, and II, respectively. [c] Data from Riccio et al. (1985). [d] Data from Nagamine et al. (1984). [e] Sequence alignment and percent sequence identity were established by using the Microgenie programs of Queen and Korn (1984). [f] Exon is entirely noncoding. [g] Calculated omitting the 27-nucleotide/9 amino acid segment unique to porcine exon 6 (see text). [h] Calculated omitting nucleotides 801–858 of the mouse gene that contain the repeat $(AG)_{29}$. [i] Calculated omitting nucleotides 1428–1745 of the mouse gene that contain a B2 family repeat and the repeat $(AC)_{21}$. [j] Calculated omitting nucleotides 1429–1736 of the human gene that contain an Alu family repeat. [k] Calculated omitting nucleotides 2075–2281 of the mouse gene that contain a B2 family repeat. [l] Calculated omitting nucleotides 4686–5223 of the mouse gene that contain a B1 family repeat and the NE sequence (see text). [m] Calculated omitting nucleotides 4853–5159 of the human gene that contain an Alu family repeat. [n] Total size of gene from transcription initiation site to polyadenylation site.

Materials and Methods). The products were analyzed on a DNA sequencing gel together with a 346 bp 5′ end-labeled *Hpa*II–*Ava*II fragment (–275 to +71, Figure 2) which had been subjected to the chemical cleavage sequencing reactions of Maxam and Gilbert (1980). The extension products were 70–73 nucleotides in length (see Figure 3), indicating that the 5′ end of the gene is at or near the nucleotide labeled +1 in Figure 2. These results were confirmed by S_1 nuclease protection analysis using the 5′-labeled *Ava*II–*Hpa*II fragment as a hybridization probe (data not shown).

Repetitive DNA. Several classes of repetitive DNA were identified in the murine uPA gene. One class of repetitive sequence consisted of the B1 and B2 family repeats (Kramerov et al., 1979; Ryskov et al., 1983; Kalb et al., 1983; King et al., 1986): B2 elements were identified in the 5′-flanking region of the gene and in the fourth and fifth intervening sequences, whereas B1 elements were found in both 5′- and 3′-flanking sequences as well as in the tenth intervening sequence (see Figure 1). A comparison of these B elements to reported (Kalb et al., 1983; Krayev et al., 1982) consensus sequences is shown in Figure 4. These data indicate that (1) the B elements associated with the murine uPA gene frequently consist of partial B repeats, often lacking the putative RNA polymerase III split-promoter found in the complete B elements, and (2) within shared sequences the B elements of the murine uPA gene are approximately 70% identical with each other and to previously described B family repeats.

A second class of repeat found in the murine uPA gene was an alternating purine–pyrimidine homocopolymer of the structure $(AC)_{21}$ (see fourth intron, Figures 1 and 2). Sequences of this type $[(AC)_{20-60}]$, although a common feature of eukaryotic genomes ($\sim 10^5$ copies in the human genome), are notable because they have been shown both to adopt left-handed conformation (Z DNA; Hamada et al., 1984a) and to enhance gene transcription in vivo (Hamada et al., 1984b). However, since no direct study has yet been undertaken, any influence of the alternating purine–pyrimidine repeat on murine uPA gene expression remains to be established.

A third class of repeat consisted of two polypurine elements. One of these, of the form $(AG)_{29}$, was found in the second intervening sequence, and a second, having the structure $(AG)_{22}(AGGG)_{16}(AG)_{28}$, was located 79 bp upstream of the proposed transcription initiation site (see Figures 1 and 2). The proximity of this latter sequence both to the transcription initiation site and to presumed regulatory sequences (see below) might prove to be of some functional or regulatory significance (see Discussion). In this context, the results of S_1 nuclease hypersensitivity studies have suggested that polypurine sequences may adopt non-B-DNA structure (Htun et al., 1984; Pullyblank et al., 1985), and we therefore tested the S_1 susceptibility of our two polypurine elements. Supercoiled plasmid DNA containing the 3.5-kb *Bgl*II fragment of the murine uPA gene (which includes over 2 kb of 5′-flanking sequences, both polypurine sequences, and exons 1–4; see Figures 1, 2, and 5) was incubated with various concentrations of S_1 nuclease. The products were subsequently digested with *Bgl*II and then analyzed by agarose gel electrophoresis. As shown in Figure 5A, murine uPA gene sequences were specifically sensitive to S_1 nuclease digestion, and five heterogeneous size fragments were

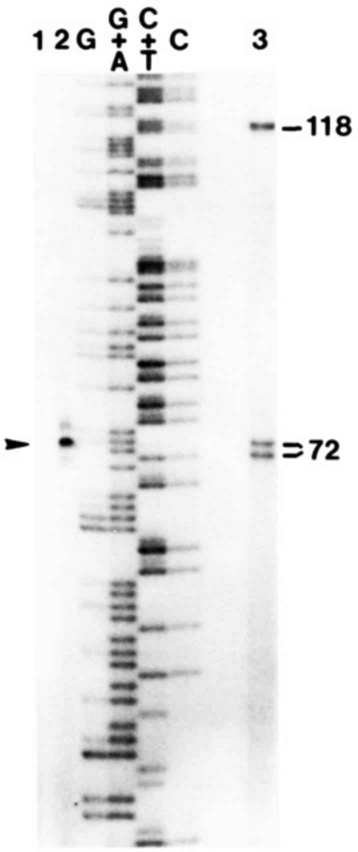

FIGURE 3: Determination of the site of transcription initiation by primer extension analysis. Oligonucleotide primer complementary to the region numbered +41 to +70 in Figure 2 was 5′ end labeled with [γ-^{32}P]ATP and T_4 polynucleotide kinase. Hybridization mixtures were prepared containing labeled oligonucleotide and either 10 μg of *Escherichia coli* tRNA (lane 1) or 10 μg of poly(A+) RNA isolated from MSV-3T3 cells treated with phorbol myristate acetate for 15 h (lane 2). The primer extension products generated by incubation with the four deoxynucleoside triphosphates and reverse transcriptase were displayed on a DNA sequencing gel together with 5′ end-labeled *Hae*III fragments of ϕX174-RF (lane 3) and the Maxam and Gilbert chemical cleavage products of the 346 bp 5′ end-labeled *Ava*II–*Hpa*II fragment (−275 to +71, Figure 2). The sizes (in nucleotides) of the ϕX174 standards are given at the right; the two 72-nucleotide fragments seen in lane 3 are resolved complementary strands.

produced; no S_1 nuclease cleavage of the plasmid vector was detected at any enzyme concentration. The location of the S_1 nuclease cleavage sites was defined by the indirect end-labeling method of Wu (1980). Briefly, the fragments generated by S_1 nuclease cleavage were transferred to nitrocellulose and hybridized to probes derived from either the 5′ or the 3′ ends of the *Bgl*II insert. As seen in Figure 5B,C, the S_1 nuclease hypersensitive regions mapped within or close to the two polypurine regions (Figure 5D). An identical pattern of S_1 sensitivity was found in experiments with linear DNA substrate (consisting of a gel-purified 3.5-kb "insert"), indicating that DNA superhelicity or torsional stress is not required for S_1 hypersensitivity. We noted further that S_1 hypersensitivity increased progressively as the NaCl concentration in the reaction was decreased from 300 to 3 mM, a result consistent with that previously reported for purine homocopolymer sequences in a sea urchin histone locus (Hentschel, 1982).

Promoter Elements. The 79 bp region intervening between the 5′-flanking polypurine sequence and the proposed site of transcription initiation contained two common RNA polymerase II promoter elements: a typical TATA sequence ($^{-34}$TAATAAA^{-28}; see Figure 2) and a potential transcription factor Sp1 binding site ($^{-62}$TGGGCGGGGC^{-53}; see Figure 2). The putative Sp1 binding site in the murine uPA gene matches precisely the extended consensus sequence reported by Briggs et al. (1986). In addition to these common promoter elements, a sequence which resembles a proposed cAMP–control element was also identified (for details, see Discussion).

Discussion

The tissue-specific and hormone-dependent expression of uPA has been extensively documented in the mouse (see the introduction). To provide a foundation for studies to define the features of the uPA gene critical to the regulation of uPA production, we have established the complete nucleotide sequence of the murine uPA gene. In addition, together with the available porcine and human uPA gene sequences (Nagamine et al., 1984; Riccio et al., 1985), these data provide a unique opportunity to compare orthologous serine protease genes from three species and to identify conserved structural and regulatory features.

Mammalian uPA Genes. Although some unique features have been noted (see below), the murine, porcine, and human uPA genes are remarkably similar. First, all three genes are relatively small (see Table I), ranging in size from 5852 bp (porcine) to 6710 bp (murine). Indeed, the uPA genes can be described as compact in comparison to many other serine protease genes: exon sequences comprise 34.7%, 36.7%, and 40.8% of the murine, human, and porcine uPA genes, respectively, whereas exon sequences account for less than 10% of the tPA (Degen et al., 1986), prothrombin (Degen & Davie, 1987), plasminogen (Malinowski et al., 1984), and factor IX (Yoshitake et al., 1985; Anson et al., 1984) genes. Second, exon size, their nucleotide sequence, and the corresponding amino acid sequence are highly conserved between species (Table I). Even the entirely noncoding exon 1 and the largely noncoding exon 11 have 50–75% sequence identity between species (Table I).

Intron Location. Excluding introns C and F of the porcine gene, the locations of the intervening sequences in the murine, human, and porcine uPA genes are precisely the same with respect to both nucleotide and amino acid sequence. Introns C and F of the porcine gene are exceptional in that the splice junction "donor" sites have apparently changed to encode two additional amino acids in exon 3 and nine additional amino acids in exon 6. Nagamine et al. (1985) have reported that the 5′ splice site of the porcine intron F may be variable and subject to alternative splicing. The introns of the murine uPA gene are not all the same with respect to the site of codon interruption: categorized as proposed by Sharp (1981) (i.e., type 0 introns fall between codons, type I introns fall between the first and second nucleotides of a codon, and type II introns fall between the second and third nucleotides of a codon), the murine uPA gene contains type 0, type I, and type II introns. However, the orthologous introns of the murine, porcine, and human uPA genes are invariably of the same type (see Table I).

Intron Sequence. If one excludes recognizable repetitive DNA sequences [i.e., homocopolymers, Alu (human), and B (murine) sequences], then, with the exception of intron J (see below), the sequences of corresponding introns in the murine, porcine, and human uPA gene are clearly similar; computer alignment of the introns from each species revealed 50–70% sequence identity (Table I). Interestingly, this maintenance of intron sequence identity has not been generally observed between orthologous genes in higher eukaryotes. For example, despite almost 90% sequence identity in the exons of the mouse and chicken α-actin genes, essentially no homology is observed

```
A.

                                                                            CG
B1conc:   CCGGGCGTGG TGGCGCACGC CTTTAATCCC AGCACTCGGG AGGCAGAGGC AGGCGGATTT CTGAGTTCGA
B1a:      --A-A---AA ---T-----T ----T---TA -A----A--- ---A------ ---T----C- --------A-
B1b:      xxxxxxxxxx xxxxxxxxxx xxxx-C--TT ----T-TT-A -T--T---A- ---A----CA x------TA-
                                                                            TG
B1c:      xxxxxxxxx- --T-T---A- --G-C----- ----AGT-A- ----T----- -A-AA--C-C -------T--
                     A                                                        G
B1d:      xxx--G---- --TG-T-T-- --------T- ----xGA--C --xxxxxxxx xxxxxxxxxx xxx-------

B1conc:   GGCCAGCCTG GTCTACAGAG TGAGTTCCAG GACAGCCAGG GCTACACAGA GAAACCCTGT CTCGAAAAAC
B1a:      -C----A--A A-T----T-- ---AA---xx xxxxxxxxxx xxxxxxxxxx xxxxxxxxxx xxxxxxxxxx
                                 AA
B1b:      --------CA AA-------- -------A-- xxxx---T-T --C---T--C A-G-T---T- ---A--T--A
B1c:      A----ATT-T -x-----A-- xxxxxxxxxx xxxxxxxxxx xxxxxxxxxx xxxxxxxxxx xxxxxxxxxx
                                                    A
B1d:      ----T-A--T ---------- -------x-- ---------- --------T- ------T--- ---A---T-A

B.

B2conc:   GGGCTGGTGA GATGGCTCAG CGGGTAAGAG CACCCGACTG CTCTTCCGAA GGTCCGGAGT TCAAATCCCA
B2a:      xxxxxxxxxx xxxxxxxxxx -T-T-C-C-T ---Tx----- -------AG- -----T---- ----T-----
                                                                     T
B2b:      xx---A-A-- --CA-G-G-- T-AT-----A ---TG-x--- ------TAG- CA---TA--- -TG-G-TT--
B2c:      ---A-A-G-- -G-------- -A-T------ ---Ax----- -CT----AG- -CA--Tx--- -TG-T-----

B2conc:   GCAACCACAT GGTGGCTCAC AACCATCCGT AACGAGATCT GATGCCCTCT TCTGGAGTGT CTGAAGACAG
B2a:      --Gx------ ---------T G------TAC --T-G----- -G-A-T--T- -----T---- -------G--
                   A                  CCCCCAGTCCC                      TC  CCC
B2b:      ---C-----G -----G-T-- -G-----T-- --G-GA---- ---------- ----ACC-C- -A----TG-C
            C                                T   G
B2c:      -GG-G-T-G- -ACA-x--TT T-A--C-T-- TCA------C ---------- -----CC-CA ---GGC-GGC

B2conc    CTACAGTGTA CTTACATATA ATAAATAAAT AATCTTTAC
B2a       Gx--T---C- --CG----C- TA-------- ---------A
             TG
B2b       A--------- -AAGT---C- TA--GG---A -C-CTCA---
                             G             T
B2c       A-G--C---x CA-GA----- GA---C-CT- --CA-A---A
```

FIGURE 4: Comparison of the B1 and B2 repeats of the murine uPA gene. (A) A B1 consensus sequence reported by Kalb et al. (1983) is shown on the top line with the underlined sequences indicating the putative split–promoter for RNA polymerase III. The B1 sequences of the mouse uPA gene are listed below; a dash indicates nucleotide identity with the consensus sequence, an x indicates the nucleotide is missing relative to the consensus sequence, and additional nucleotides that are not found in the consensus sequence are listed above each line. The B1 repeats listed are located as follows in Figure 2: B1a = −882 to −781; B1b = 4686–4798; B1c = 7199–7280; B1d = 7317–7433. (B) A B2 consensus sequence reported by Krayev et al. (1982) is shown on the top line with the underlined sequences indicating the putative RNA polymerase III split–promoter. The B2 sequences listed are located as follows in Figure 2: B2a = −1295 to −1139; B2b = 1428–1624; B2c = 2075–2255.

in the introns (Hu et al., 1986). However, the maintenance of intron sequence homology is not exclusively observed in the uPA genes; the introns of the goat ϵ^{I}-globin gene are ~65% identical with those in the human ϵ-globin gene[2] (Shapiro et al., 1983). With the limited number of mammalian genes which have been fully sequenced in more than one species, it is difficult to accurately estimate how frequently intron sequence homology is maintained; however, the uPA gene seems to be one of the exceptional cases in which the level of intron sequence identity approaches the level of exon sequence identity between species. The basis for conservation in the noncoding regions of the uPA gene is unclear.

Novel Insertion Element in Intron J. On the basis of computer alignments, intron J of the murine uPA gene, in addition to a B1 element, contains a ~380-nucleotide element for which there is no counterpart in the porcine or human genes. This element, which we call NE, includes nucleotides 4846–5223 in Figure 2. The position of NE immediately 3′ to a B1 element is consistent with a model in which the B1 and NE sequences were inserted into the gene simultaneously, most likely as a single, combined B1–NE unit. A comparison of the NE sequence to the GenBank data base did not reveal any related sequences in the mouse or any other species. We are currently investigating whether NE-hybridizing sequences occur elsewhere in the murine genome.

[2] Calculated by omitting a 226 bp insertion element in the goat ϵ^{I} gene (Shapiro et al., 1983).

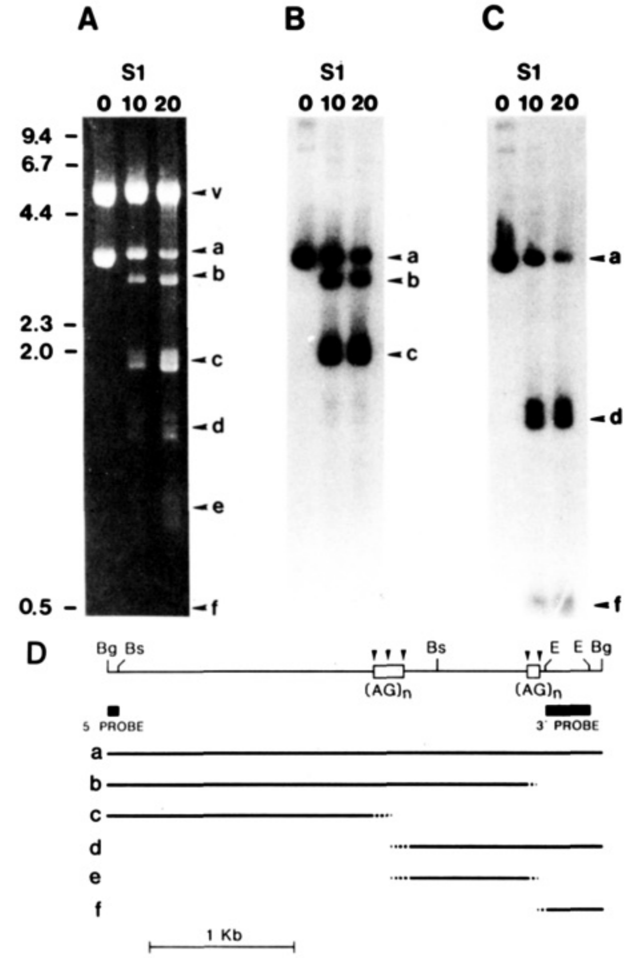

FIGURE 5: Nuclease S_1 hypersensitive regions in the murine uPA gene. The plasmid pB3.5 (see Figure 1, 3.5-kb *Bgl*II subclone) was incubated in reaction mixtures containing either 0, 10, or 20 units of nuclease S_1. The DNA samples were then digested to completion with *Bgl*II and the fragments generated analyzed by agarose gel electrophoresis. (A) Ethidium bromide stained DNA fragments of pB3.5; v indicates plasmid vector sequences, a indicates the 3.5-kb *Bgl*II insert fragment, and b–f denote fragments generated by nuclease S_1 cleavage of the insert DNA. The relative migration and size (in kilobases) of *Hin*dIII fragments of λ DNA are indicated at the left in panel A. The DNA fragments in panel A were transferred to nitrocellulose filters and hybridized to ^{32}P-labeled probes derived from either the 5′ end (B) or the 3′ end (C) of the 3.5-kb insert (see panel D). (D) Partial map of restriction endonuclease cleavage sites in the 3.5-kb insert (Bg, *Bgl*II; Bs, *Bst*EII; E, *Eco*RI) indicating the positions of the polypurine repeats [boxed regions labeled $(AG)_n$]. The DNA fragments expected from a partial cleavage within the polypurine sequences are labeled a–f.

Repetitive DNA in the uPA Gene. The presence of both B1 and B2 repeats in the murine uPA gene is consistent with the frequency that these elements occur in the murine genome (~10^5 copies) and the common occurrence of these elements in other murine genes. The B elements of the murine uPA gene, like the Alu elements in the human uPA gene, are located in introns D and J. However, the precise locations of the repetitive elements within orthologous introns of the murine and human uPA genes were not comparable. Thus, these data are consistent with the proposal that B and Alu family repeats were inserted into a repetitive element-free primordial uPA gene after murine–human speciation. The porcine gene may best reflect the primordial uPA gene in that it does not contain any discernible transposable element-like sequences.

Polypurine Sequences. A striking feature of the murine uPA gene is the presence of two large polypurine sequences. One hypothesis that remains to be tested directly is that these sequences influence murine uPA gene expression. A number of observations are consistent with this notion. First, the large 5′-flanking polypurine sequence occurs less than 50 nucleotides upstream of two common RNA polymerase II promoter elements, a putative transcription factor Sp1 binding site and the TATA box (see Results), and less than 50 nucleotides downstream of a potential cAMP regulatory element (see below). Second, our results indicate that the polypurine sequences of the murine uPA gene can assume a S_1 nuclease hypersensitive structure. Although the relationship of this type of structural alteration and gene expression is yet to be clearly defined, it is well documented that changes in gene activity are often accompanied by local changes in gene nuclease hypersensitivity when assayed in isolated chromatin (Weintraub & Groudine, 1976; Larsen & Weintraub, 1982). The non-B structure generated by polypurine sequences may extend into neighboring sequences; Kohwi-Shigematsu and Kowhi (1985) reported that polypurine elements alter the DNA conformation (described as "unpaired") of 3′-flanking sequences over a distance of at least 40 bp. If similar structural alterations occur in the murine uPA gene in vivo, then the region influenced might include the putative Sp1 binding site and TATA sequence. Finally, large polypurine sequences occur immediately 5′ to a number of mammalian genes, including the mouse α_1-antitrypsin (Krauter et al., 1986), H-2K^b (Kimura et al., 1986), and N-myc (DePinho et al., 1986) genes. Similarly, the porcine and human uPA genes have purine-rich sequences within 80 bp of their respective transcription initiation sites (~30 nucleotides; >90% A + G).

cAMP Modulation of uPA Gene Expression. Recent studies of a number of cAMP-modulated genes, including those encoding rat phosphoenolpyruvate carboxykinase (PEP-CK; Wynshaw-Boris et al., 1986; Short et al., 1986), rat preprosomatostatin (Montminy et al., 1986), human proenkephalin (Comb et al., 1986), and human chorionic gonadotropin (Silver et al., 1987), have provided direct evidence for 5′-flanking cis-acting regulatory elements which mediate cAMP induction of gene transcription. These documented cAMP regulatory sequences appear to be short (<30 bp), share a similar core sequence (~12 bp, see Table II), and impart cAMP-dependent expression to both parent and heterologous promoters independent of position and orientation. Consistent with the frequently (but not universally) observed stimulation of uPA transcription by cAMP in cultured cells (Degen et al., 1985; our unpublished results), the murine, porcine, and human uPA genes were found to contain sequences similar to the proposed cAMP regulatory elements (see Table II). However, given that only relatively short and somewhat variable sequences

Table II: Homology in the 5′-Flanking Regions of cAMP-Modulated Genes

Gene	Sequence[a]	Identity (%)
Rat PEP-CK[b]	-91 -80 CTTACGTCAGAG	100
Human Proenkephalin[c]	-92 -82 --Gx------CT	67
Rat Preprosomatostatin[d]	-49 -38 --G---------	92
Human Chorionic Gonadotropin[e]	-125 -114 T-G------TG-	67
Murine uPA	-305 -295 ---x-----A--	83
Human uPA[f]	-306 -296 --G-G--x----	75
Porcine uPA[g]	-125 AG -112 G-G-C-------	64

[a] A dash indicates identity with the PEP-CK sequence, and a x indicates the nucleotide is not present. [b] Short et al. (1986). [c] Comb et al. (1986). [d] Montminy et al. (1986). [e] Silver et al. (1987). [f] Riccio et al. (1985). [g] Nagamine et al. (1984).

have been associated with cAMP modulation, direct studies will be necessary to determine whether these and/or other uPA gene sequences are important to cAMP regulation. The availability of murine cell lines which contain defined mutations in the cAMP-dependent protein kinase [for a review, see Steinberg (1983)] and the availability of both cDNAs (Uhler et al., 1986; Lee et al., 1983) and genes (Uhler et al., 1986) for the murine cAMP-dependent protein kinase subunits will be valuable tools in clarifying the relationship of the cAMP-dependent protein kinase and murine uPA gene expression.

Acknowledgments

We thank C. Dang, M. Neustein, and D. Spector for their excellent technical assistance and J. Hagedorn for assistance in preparing the manuscript. We also thank Drs. D. Wiginton, C. Duncan, and N. Leslie for critically reading the manuscript and Dr. R. Ganschow for advice with data processing.

References

Anson, D. S., Choo, K. H., Rees, D. J. G., Giannelli, F., Gould, K., Huddleston, J. A., & Brownlee, G. G. (1984) *EMBO J. 3*, 1053–1060.

Aviv, H., & Leder, P. (1972) *Proc. Natl. Acad. Sci. U.S.A. 69*, 1408–1412.

Beers, W. H., Strickland, S., & Reich, E. (1975) *Cell (Cambridge, Mass.) 6*, 387–394.

Belin, D., Godeau, F., & Vassalli, J.-D. (1984) *EMBO J. 3*, 1901–1906.

Belin, D., Vassalli, J.-D., Combepine, C., Godeau, F., Nagamine, Y., Reich, E., Kocher, H. P., & Duvoisin, R. M. (1985) *Eur. J. Biochem. 148*, 225–232.

Breathnach, R., Benoit, C., O'Hare, K., Gannon, F., & Chambon, P. (1978) *Proc. Natl. Acad. Sci. U.S.A. 75*, 4853–4857.

Briggs, M. R., Kadonaga, J. T., Bell, S. P., & Tjian, R. (1986) *Science (Washington, D.C.) 234*, 47–52.

Comb, M., Birnberg, N. C., Seasholtz, A., Herbert, E., & Goodman, H. M. (1986) *Nature (London) 323*, 353–356.

Danø, K., Andreasen, P. A., Grøndahl-Hansen, J., Kristensen, P., Nielsen, L. S., & Skriver, L. (1985) *Adv. Cancer Res. 44*, 139–266.

Dayer, J.-M., Vassalli, J.-D., Bobbitt, J. L., Hull, R. N., Reich, E., & Krane, S. M. (1981) *J. Cell Biol. 91*, 195–200.

Degen, J. L., Estensen, R. D., Nagamine, Y., & Reich, E. (1985) *J. Biol. Chem. 260*, 12426–12433.

Degen, S. J. F., & Davie, E. W. (1987) *Biochemistry 26*, 6165–6177.

Degen, S. J. F., Rajput, B., & Reich, E. (1986) *J. Biol. Chem. 261*, 6972–6985.

De Pinho, R. A., Legouy, E., Feldman, L. B., Kohl, N. E., Yancopoulos, G. D., & Alt, F. W. (1986) *Proc. Natl. Acad. Sci. U.S.A. 83*, 1827–1831.

Fritz, I. B., Parvinen, M., Karmally, K., & Lacroix, M. (1982) *Ann. N.Y. Acad. Sci. 383*, 447–448.

Grimaldi, G., DiFiore, P., Locatelli, E. K., Falco, J., & Blasi, F. (1986) *EMBO J. 5*, 855–861.

Gross, J. L., Moscatelli, D., & Rifkin, D. B. (1983) *Proc. Natl. Acad. Sci. U.S.A. 80*, 2623–2627.

Grosveld, F. G., Lund, T., Murray, E. J., Mellor, A. L., Dahl, H. H. M., & Flavell, R. A. (1982) *Nucleic Acids Res. 10*, 6715–6732.

Hamada, H., Petrino, M. G., Kakunaga, T., Seidman, M., & Stollar, B. D. (1984a) *Mol. Cell. Biol. 4*, 2610–2621.

Hamada, H., Seidman, M., Howard, B. H., & Gorman, C. A. (1984b) *Mol. Cell. Biol. 4*, 2622–2630.

Hentschel, C. C. (1982) *Nature (London) 295*, 714–716.

Htun, H., Lund, E., & Dahlberg, J. E. (1984) *Proc. Natl. Acad. Sci. U.S.A. 81*, 7288–7292.

Hu, M. C., Sharp, S. B., & Davidson, N. (1986) *Mol. Cell. Biol. 6*, 15–25.

Kalb, V. F., Glasser, S., King, D., & Lingrel, J. B. (1983) *Nucleic Acids Res. 11*, 2177–2184.

Kimura, A., Israël, A., LeBail, O., & Kourilsky, P. (1986) *Cell (Cambridge, Mass.) 44*, 261–272.

King, D., Snider, L. D., & Lingrel, J. B. (1986) *Mol. Cell. Biol. 6*, 209–217.

Kohwi-Shigematsu, T., & Kowhi, Y. (1985) *Cell (Cambridge, Mass.) 43*, 199–206.

Kramerov, D. A., Grigoryan, A. A., Ryskov, A. P., & Georgiev, G. P. (1979) *Nucleic Acids Res. 6*, 697–713.

Krauter, K. S., Citron, B. A., Hsu, M.-T., Powell, D., & Darnell, J. E. (1986) *DNA 5*, 29–36.

Krayev, A. S., Markusheva, T. V., Kramerov, D. A., Pyskov, A. P., Scryabin, K. G., Bayev, A. A., & Georgiev, G. P. (1982) *Nucleic Acids Res. 10*, 7461–7475.

Larsen, A., & Weintraub, H. (1982) *Cell (Cambridge, Mass.) 29*, 609–622.

Lee, D. C., Carmichael, D. F., Krebs, E. G., & McKnight, G. S. (1983) *Proc. Natl. Acad. Sci. U.S.A. 80*, 3608–3612.

Malinowski, D. P., Sadler, J. E., & Davie, E. W. (1984) *Biochemistry 23*, 4243–4250.

Maxam, A. M., & Gilbert, W. (1980) *Methods Enzymol. 65*, 499–560.

Medcalf, R. L., Richards, R. I., Crawford, R. I., & Hamilton, J. A. (1986) *EMBO J. 9*, 2217–2222.

Mignatti, P., Robbins, E., & Rifkin, D. B. (1986) *Cell (Cambridge, Mass.) 47*, 487–498.

Miskin, R., Easton, T. G., & Reich, E. (1978) *Cell (Cambridge, Mass.) 15*, 1301–1312.

Montminy, M. R., Sevarino, K. A., Wagner, J. A., Mandel, G., & Goodman, R. H. (1986) *Proc. Natl. Acad. Sci. U.S.A. 83*, 6682–6686.

Mount, S. M. (1982) *Nucleic Acids Res. 10*, 459–472.

Nagamine, Y., & Reich, E. (1985) *Proc. Natl. Acad. Sci. U.S.A. 82*, 4606–4610.

Nagamine, Y., Sudol, M., & Reich, E. (1983) *Cell (Cambridge, Mass.) 32*, 1181–1190.

Nagamine, Y., Pearson, D., Altus, M. S., & Reich, E. (1984) *Nucleic Acids Res. 12*, 9525–9541.

Nagamine, Y., Pearson, D., & Grattan, M. (1985) *Biochem. Biophys. Res. Commun. 132*, 563–569.

Ossowski, L., & Reich, E. (1983) *Cell (Cambridge, Mass.) 35*, 611–619.

Ossowski, L., Biegel, D., & Reich, E. (1979) *Cell (Cambridge, Mass.) 16*, 929–940.

Pulleyblank, D. E., Haniford, D. B., & Morgan, A. R. (1985) *Cell (Cambridge, Mass.) 42*, 271–280.

Queen, C., & Korn, L. J. (1984) *Nucleic Acids Res. 12*, 581–599.

Riccio, A., Grimaldi, G., Verde, P., Sebastio, G., Boast, S., & Blasi, F. (1985) *Nucleic Acids Res. 13*, 2759–2771.

Ryskov, A. P., Ivanov, P. L., Kramerov, D. A., & Georgiev, G. P. (1983) *Nucleic Acids Res. 11*, 6541–6558.

Shapiro, S. G., Schon, E. A., Townes, T. M., & Lingrel, J. B. (1983) *J. Mol. Biol. 169*, 31–52.

Sharp, P. A. (1981) *Cell (Cambridge, Mass.) 23*, 643–646.

Short, J. M., Wynshaw-Boris, A., Short, H. P., & Hanson, R. W. (1986) *J. Biol. Chem. 261*, 9721–9726.

Silver, B. J., Bokar, J. A., Virgin, J. B., Vallen, E. A., Milsted, A., & Nilson, J. G. (1987) *Proc. Natl. Acad. Sci. U.S.A. 84*, 2198–2202.

Southern, E. M. (1975) *J. Mol. Biol. 98*, 503–517.
Steinberg, R. A. (1983) *Biochem. Actions Horm. 11*, 25–65.
Steinmetz, M., Stephan, D., Dastoornikov, G. R., Gibb, E., & Romanuik, R. (1985) in *Immunological Methods* (Lefkovits, I., & Pernis, B., Eds.) Vol. III, pp 1–19, Academic Press, New York.
Strickland, S., Reich, E., & Sherman, M. I. (1976) *Cell (Cambridge, Mass.) 9*, 231–240.
Uhler, M. D., Carmichael, D. F., Lee, D. C., Chrivia, J. C., Krebs, E. G., & McKnight, G. S. (1986) *Proc. Natl. Acad. Sci. U.S.A. 83*, 1300–1304.
Unkeless, J. C., Danø, K., Kellerman, G. M., & Reich, E. (1974) *J. Biol. Chem. 249*, 4295–4305.
Vassalli, J.-D., Hamilton, J., & Reich, E. (1977) *Cell (Cambridge, Mass.) 11*, 695–705.
Virji, M. A. G., Vassalli, J.-D., Estensen, R. D., & Reich, E. (1980) *Proc. Natl. Acad. Sci. U.S.A. 77*, 875–879.
Weintraub, H., & Groudine, M. (1976) *Science (Washington, D.C.) 193*, 848–856.
Wu, C. (1980) *Nature (London) 286*, 854–860.
Wynshaw-Boris, A., Short, J. M., Loose, D. S., & Hanson, R. W. (1986) *J. Biol. Chem. 261*, 9714–9720.
Yoshitake, S., Schach, B. G., Foster, D. C., Davie, E. W., & Kurachi, K. (1985) *Biochemistry 24*, 3736–3750.

Gene Structure of Cytochrome P-450(M-1) Specifically Expressed in Male Rat Liver[†]

Nobuhiro Morishima,[‡] Hidefumi Yoshioka,[§] Yujiro Higashi, Kazuhiro Sogawa,[∥] and Yoshiaki Fujii-Kuriyama*[,∥]
Department of Biochemistry, Cancer Institute, Japanese Foundation for Cancer Research, Toshima-ku Tokyo 170, Japan
Received May 6, 1987; Revised Manuscript Received July 31, 1987

ABSTRACT: Cytochrome P-450(M-1) [P-450(M-1)] is specifically expressed in adult male rat liver [Yoshioka, H., Morohashi, K., Sogawa, K., Miyata, T., Kawajiri, K., Hirose, T., Inayama, S., Fujii-Kuriyama, Y., & Omura, T. (1987) *J. Biol. Chem. 262*, 1706–1711]. Isolation and analysis of the gene for P-450(M-1) revealed that the coding region of the gene is interrupted by eight introns and is dispersed over a 35-kilobase pair region of chromosomal DNA. Intron insertion sites of the P-450(M-1) gene are located at equivalent positions to those of cytochrome P-450b and P-450e, which are phenobarbital-inducible. Sequence analysis of the 5′-upstream region of the P-450(M-1) gene shows that there is a homologous sequence to glucocorticoid regulatory elements (GRE) identified in glucocorticoid-responsive genes.

Cytochrome P-450 (P-450)[1] is a group of monooxygenases that catalyze oxidation of a variety of both endogenous and exogenous substrates (Sato & Omura, 1978; Lu & West, 1980). These enzymes contain a heme moiety as a prosthetic group and are related with one another as revealed by the comparison of their amino acid sequences (Gotoh et al., 1983). The sequence similarity of P-450s suggests that they have diverged from a common ancestral enzyme in the course of evolution. In addition to diversity of substrate specificity, constituents of the P-450 superfamily exhibit various modes of expression of their own, for example, in temporal, tissue-specific, sex-dependent, or inducer-specific manners (Sato & Omura, 1978; Lu & West, 1980). Because of the variety in the mode of expression, P-450 should provide a suitable system for the study on the regulation of gene expression.

P-450(M-1) is present in microsomes of adult rat livers and catalyzes testosterone 16α-hydroxylation (Matsumoto et al., 1986; Morgan et al., 1985a). The expression of the enzyme is male- and age-specific. Recently, cDNA clones for P-450(M-1) were isolated in our laboratory (Yoshioka et al., 1987). Although sequence analysis of the cDNA clones revealed that P-450(M-1) has a high degree of sequence similarity to the coding sequence of phenobarbital-inducible P-450s, Northern blot analysis clearly showed that P-450(M-1) mRNA is specifically synthesized in adult male rat livers as a constitutive form. Morgan et al. (1985b) have demonstrated that the expression of the male-specific P-450 of rat livers is under the control of the growth hormone secretion pattern, a highly pulsatile secretion of the hormone in the male and a more constant level of the hormone in the female rats. The regulation mechanism of the hormonal axis (pituitary–liver) mediated by the growth hormone, however, has not yet been fully unraveled at a molecular level. As an essential step toward investigating the regulation mechanism of male- and age-specific expression of P-450(M-1), we have isolated and characterized the P-450(M-1) gene.

Experimental Procedures

Materials. Restriction endonucleases were purchased from Takara Shuzo Co. (Kyoto, Japan), New England Biolabs (Beverly, MA), and Bethesda Research Laboratories (Rockville, MD). *Escherichia coli* DNA polymerase I (large fragment), T4 DNA ligase, bacterial alkaline phosphatase, and polynucleotide kinase were obtained from Takara Shuzo Co.

[†] This work was supported in part by grants-in-aid for Scientific Research from the Ministry of Education, Science and Culture of Japan, research grants from the Ministry of Health and Welfare of Japan, and funds obtained under the Life Science Project from the Institute of Physical and Chemical Research.
* Author to whom correspondence should be addressed.
[‡] Present address: Laboratory of Microbiology, Institute of Physical and Chemical Research, Wako-shi, Saitama 351-01, Japan.
[§] Present address: Division of Molecular Genetics, Institute for Comprehensive Medical Science, Fujita-Gakuen Health University, Toyoake-shi, Aichi 470-11, Japan.
[∥] Present address: Department of Chemistry, Faculty of Science, Tohoku University, Sendai, Miyagi, Japan.

[1] Abbreviations: P-450, cytochrome P-450; kb, kilobase pair(s); bp, base pair(s); SSC, 0.15 M NaCl containing 15 mM sodium citrate; GRE, glucocorticoid regulatory elements.

0006-2960/87/0426-8279$01.50/0