Identification of an Enhancer Required for the Expression of a Mouse Major Urinary Protein Gene in the Submaxillary Gland

HYEUNG JIN SON,^{1†} KAREN SHAHAN,² MARIANNE RODRIGUEZ,² EVA DERMAN,²
AND FRANK COSTANTINI^{1*}

Department of Genetics and Development, College of Physicians and Surgeons, Columbia University, New York, New York 10032,¹ and The Public Health Research Institute of the City of New York, New York, New York 10016²

Received 24 January 1991/Accepted 20 May 1991

The *MUP1.5b* gene was previously found to be expressed specifically in the submaxillary gland and at high levels when introduced into mice as a transgene including 4.7 kb of 5'-flanking DNA and 0.3 kb of 3'-flanking DNA. To localize regulatory elements responsible for this tissue-specific pattern of expression, we tested the expression of three additional *MUP1.5b* transgenes including various amounts of 5'-flanking DNA. These experiments indicated that sequences between -1.85 and -3.46 kb from the transcription initiation site were required for high-level expression in the submaxillary gland. The presence of regulatory elements in this region was also suggested by the detection of a DNase I-hypersensitive site, seen only in submaxillary gland nuclei, at position -2.5 kb upstream from the *MUP1.5a* gene, a member of the same *MUP* gene subfamily and virtually identical to the *MUP1.5b* gene. Further evidence for enhancer activity was provided by the ability of the 1.6-kb DNA fragment including sequences between -1.85 and -3.46 kb to stimulate the expression of an otherwise inactive *MUP1.5b*-chloramphenicol acetyltransferase fusion gene specifically in the submaxillary gland. The nucleotide sequence of this 1.6-kb DNA fragment was found to be identical for the *MUP1.5a* and *MUP1.5b* genes. Together, these results provide the first localization of a *cis*-acting regulatory sequence involved in the differential tissue-specific expression of the *MUP* gene family.

The mouse major urinary proteins (MUPs) are a family of low-molecular-weight, secreted proteins (12, 19) whose functions are unknown but which are closely related in sequence to a group of proteins that bind small, hydrophobic ligands (e.g., retinol-binding protein, lactoglobulin, and apolipoprotein D [15]). The MUPs are synthesized in a variety of secretory tissues (19, 31) and are encoded by at least six distinct *MUP* mRNA species, which have been identified by sequence analysis of cDNA clones isolated from these tissues (11, 20, 30, 31). *MUP I* and *MUP III* are synthesized only in the liver, *MUP II* is synthesized in the liver and the mammary gland, *MUP IV* is synthesized in the lachrymal and parotid glands, *MUP V* is synthesized in the submaxillary and (at low levels) sublingual glands, and *MUP VI* is synthesized in the parotid gland (30). These mRNAs are encoded by a family of approximately 35 linked genes, some of which are functional and expressed in a variety of tissue-specific patterns (5, 7, 9, 19, 20, 25, 32, 33) and others of which are pseudogenes (8) or silent genes (34). Because all the *MUP* genes are highly conserved in DNA sequence, even within introns and flanking regions (7, 20, 33), the diversity of expression patterns among the different *MUP* genes is likely to have resulted from relatively small changes in regulatory DNA sequences during the evolution of this multigene family. Therefore, the *MUP* genes represent an interesting system in which to investigate the mechanisms of tissue-specific gene expression. However, to date, nothing is known about the regulatory sequences responsible for dif-

ferential tissue-specific expression among members of the *MUP* gene family.

The present study concerns the regulation of a *MUP* gene subfamily consisting of two recently duplicated and virtually identical genes, *MUP1.5a* and *MUP1.5b*, in BALB/cByJ mice (34). These two genes are >99.9% similar, differing at only three positions over the entire 4-kb transcription unit; moreover, their 5'-flanking regions are virtually identical for at least several kilobases upstream from the genes, as shown by extensive restriction mapping (34; unpublished data) and the sequence analysis reported here. However, only the *MUP1.5a* gene encodes the *MUP V* mRNA found primarily in the submaxillary gland, and the *MUP1.5b* gene is not expressed in normal mouse tissues. Nevertheless, in mice in which a *MUP1.5b* transgene was inserted at new chromosomal loci, the *MUP1.5b* gene was expressed at high levels with a tissue specificity indistinguishable from that of the endogenous *MUP1.5a* gene. Thus, we have concluded that the *MUP1.5a* and *MUP1.5b* genes both contain *cis*-acting regulatory elements capable of specifying submaxillary gland expression but that the latter is silenced by a long-range position effect at its normal locus (34).

In this paper, we report experiments designed to localize the regulatory elements responsible for the tissue-specific transcription of the *MUP1.5* genes. An analysis of a series of *MUP1.5b* transgenes with progressive 5' truncations has identified a 1.6-kb upstream region that is required for high-level expression in the submaxillary gland. The presence of regulatory elements has been further supported by the observation of a tissue-specific DNase I-hypersensitive (HS) site within the corresponding region of the *MUP1.5a* gene. Finally, studies with *MUP1.5*-CAT fusion genes either containing or lacking this upstream region have confirmed the presence of an enhancer required for expression in the submaxillary gland.

* Corresponding author.

† Present address: Laboratory of Molecular Neurobiology, The W. M. Burke Medical Research Institute, Cornell University Medical College, White Plains, NY 10605.

FIG. 1. *MUP1.5b* gene and transgene constructs. (a) Partial restriction map of the *MUP1.5b* gene showing restriction sites used in the construction of 5'-truncated genes and fusion genes, as well as *PstI* sites used for the DNase I-HS studies. Numbers indicate the distance from the transcription initiation site in kilobases. Solid boxes indicate exons; the arrow indicates a transcription initiation site; the solid oval indicates the location of the DNase I-HS site in the *MUP1.5a* gene. (b) *MUP1.5b* transgenes with progressively truncated 5'-flanking DNA, all terminating at the 3' *HindIII* site. The *MUP1.5b* transgenes were produced with a gene that had been marked by insertion of a 10-bp oligonucleotide into the first exon (34). (c) *MUP1.5b*-CAT fusion genes. Solid lines indicate *MUP1.5b* sequences included in fusion genes; dashed lines indicate *MUP1.5b* 5'-flanking sequences absent from fusion genes; open boxes indicate the CAT gene.

MATERIALS AND METHODS

DNA constructs. The production of construct pE and the insertion of an oligonucleotide to mark the gene and its mRNA have been described previously (34). Construct pD consisted of a *BamHI*-*HindIII* fragment excised from pE. Construct pG was produced by joining a 1.3-kb *SstI*-*BamHI* fragment to the 5' end (*BamHI* site) of pD; a *SalI* linker was added at the *SstI* site, allowing the entire 6.2-kb transgene fragment to be excised as a *SalI*-*HindIII* fragment. For the production of construct pH, a 1.6-kb *EcoRI*-*SstI* fragment was added to the *SstI* site of pG and a *SalI* linker was added to the *EcoRI* site to allow the 7.8-kb pH fragment to be excised as a *SalI*-*HindIII* fragment. pDCAT was constructed by isolating the chloramphenicol acetyltransferase (CAT) gene (1.6-kb *BamHI*-*BamHI* fragment) from plasmid ElmCAT (26), cloning it into the *BamHI* site of pUC19, and inserting the 0.57-kb *BamHI*-*SmaI* fragment, including the *MUP1.5b* promoter into the *SmaI* site in the polylinker upstream from the CAT gene. For the production of pDC-1.3, the 1.3-kb *HindIII*-*EcoRI* fragment of the *MUP1.5b* gene was treated with the Klenow DNA polymerase to blunt the ends and subcloned into the blunt-ended *KpnI* site upstream of pDCAT. The 3.6-kb pDC-1.3 transgene was excised by double digestion with *HindIII* and *SstI*. pDC-1.6 was similarly produced, except that the 1.6-kb *EcoRI*-*SstI* fragment of the *MUP1.5b* gene was used.

Production and identification of transgenic mice. The *MUP1.5b* transgenes (Fig. 1b) and CAT fusion genes (Fig. 1c) were excised from plasmid vectors, purified by agarose gel electrophoresis, and subjected to cesium chloride centrifugation (22) or phenol extraction (28). Purified DNA was microinjected at a concentration of 2 to 4 $\mu\text{g}/\text{ml}$ into the pronuclei of (CBA/J $\times$ C57BL/6J) F_2 mouse zygotes as described previously (22). DNA was isolated from the tails

of potential transgenic mice at 2 weeks of age and analyzed by Southern hybridization. For construct pD, mice were screened by Southern hybridization with a *MUP1.5* gene-specific probe (34), which hybridized to the endogenous *MUP1.5* genes as well as to the *MUP1.5b* transgene, yielding restriction fragments of different sizes. The copy number of the transgenes was estimated by comparison of the transgene bands with the endogenous *MUP1.5* gene bands. For constructs pG and pH and the three CAT constructs, mice were screened by the polymerase chain reaction with primers specific for the transgenes.

MUP mRNA measurements. Transgenic mice were sacrificed at approximately 3 weeks of age (the age at which the MUP V mRNA is found at the highest level [30]), and RNA was isolated from the submaxillary gland or other tissues. RNase protection measurements of *MUP1.5b* mRNA were performed as described previously (34).

CAT determinations. Transgenic mice were sacrificed at 3 to 4 weeks of age, and tissues were frozen at -70°C . Tissue samples were thawed and homogenized in 0.5 ml of 250 mM Tris HCl (pH 7.8) at 4°C with a Polytron homogenizer. The extracts were heated at 65°C for 5 min to inactivate endogenous deacetylase activity. A 50- to 100- μl portion of each extract was used for the CAT assay as described previously (16) in a reaction mixture including 2 μl of [^{14}C]chloramphenicol (NEN-408A, CAT assay grade), 35 μl of 0.5 M Tris (pH 7.8), 20 μl of 4 mM acetyl coenzyme A, and the tissue extract. Reaction mixtures were incubated for 5 h and assayed by ascending thin-layer chromatography and autoradiography for 72 h. The autoradiograms were scanned with a densitometer.

Preparation of nuclei. Nuclei were prepared essentially as described previously (4). Finely minced tissue was homogenized at 4°C in solution A (15 mM Tris-HCl [pH 7.4], 0.5

FIG. 2. Low-level, nonspecific expression of the pD transgene. Seven independent founder transgenic mice carrying transgene pD (Fig. 1b) were identified by Southern analysis, and RNA was isolated from the submaxillary glands at 3 weeks of age. (A) RNase protection analysis of MUP mRNA from the submaxillary glands of five pD-carrying transgenic mice (lanes 1, 12, 15, 34, and 35); lane E, submaxillary gland RNA from a positive control mouse expressing the pE construct (transgenic line 14); lane M, molecular weight markers (32 P-labeled *Hpa*II fragments of pBR322 DNA). The transgenes and the labeled RNA probe used to detect MUP mRNA all contained a 10-nucleotide insertion in the first exon, while the endogenous *MUP1.5* genes lacked this insertion. Thus, hybridization of the probe to transgenic MUP mRNA protected probe fragments of approximately 110 nucleotides from RNase digestion, while hybridization to endogenous *MUP1.5* mRNA resulted in shorter (67- to 76-nucleotide) protected fragments (30). (B) Analysis of submaxillary gland RNA from two additional pD-carrying founder transgenic mice, numbers 49 and 54, together with mouse number 35 for comparison; lane C, control submaxillary gland RNA from a nontransgenic mouse; lane E is as in panel A. (C) Analysis of RNA from submaxillary gland (Sm), lachrymal gland (La), sublingual gland (SI), liver (Li), and kidney (K) tissues of transgenic mouse number 35; lane M is as in panel A. Ten micrograms of total RNA was used for each lane in panels A, B, and C. Numbers at left and right represent nucleotides.

mM spermine, 0.15 mM spermidine, 60 mM KCl, 15 mM NaCl, 12 mM β -mercaptoethanol, 0.5 M sucrose, 2 mM EDTA, 0.2 mM EGTA, 1.0 mM phenylmethylsulfonyl fluoride [PMSF]) with a motor-driven Potter homogenizer; approximately 1 g of tissue was used per 10 ml of solution A. The homogenate was diluted with an equal volume of solution B (15 mM Tris-HCl [pH 7.4], 0.5 mM spermine, 0.15 mM spermidine, 60 mM KCl, 15 mM NaCl, 12 mM β -mercaptoethanol, 2.1 M sucrose, 2 mM EDTA, 0.2 mM EGTA, 1.0 mM PMSF), layered over a cushion of solution B, and centrifuged at 20,000 rpm for 45 min in an SW27 rotor at 4°C. Pelletted nuclei were washed twice in 2 ml of solution C (25 mM Tris-HCl [pH 7.4], 0.5 mM spermine, 0.25 mM spermidine, 60 mM KCl, 25 mM NaCl, 12 mM β -mercaptoethanol, 0.34 M sucrose, 0.1% Triton X-100, 0.2 mM EDTA, 0.2 mM EGTA, 1.0 mM PMSF), and the purified nuclei were frozen in aliquots in 25% glycerol–50 mM Tris (pH 8.0) at -70°C .

Mapping of DNase I-HS sites. For each digestion, 3×10^6 nuclei were resuspended in 100 μl of DNase I buffer (15 mM Tris-HCl [pH 7.4], 0.5 mM spermine, 0.15 mM spermidine, 60 mM KCl, 15 mM NaCl, 12 mM β -mercaptoethanol, 0.5 M sucrose, 0.2 mM EDTA, 0.2 mM EGTA, 1.0 mM PMSF). The reactions were initiated by the addition of DNase I (Worthington) to a final concentration of 5 to 30 U/ml and by the addition of MgCl_2 to 50 mM, in that order, on ice. The DNase I digestions were allowed to proceed at 30°C for 20 min and were terminated by the addition of 100 μl of a solution containing 20 mM EDTA, 0.4% sodium dodecyl

sulfate (SDS), 0.4 M NaCl, and 200 μg of proteinase K per ml. Following incubation at 37°C , the DNA was further purified by phenol-chloroform extraction and ethanol precipitation. Control nuclei were treated as described above, except that no DNase was added.

For Southern analysis of the DNase I-digested DNA, hybridizations were performed in a buffer containing $5\times$ SSC ($1\times$ SSC is 0.15 M NaCl plus 0.015 M sodium citrate), 0.1 M piperazine-*N,N'*-bis(2-ethanesulfonic acid) (PIPES) (pH 7), 0.1% Ficoll, 0.1% polyvinylpyrrolidone, 0.2% SDS, 10 mM EDTA, 0.1 mg of yeast tRNA per ml, and 50% formamide at 50°C . Washes were done in a buffer containing $1\times$ SSC and 0.2% SDS at 65°C .

DNA sequence analysis. The upstream 1.6-kb *Eco*RI-*Sst*I fragment of the *MUP1.5a* gene was used to prepare a set of nested deletions, which were sequenced on one strand by the dideoxy chain termination method (29). The second strand of the *MUP1.5a* gene and the corresponding strand of the *MUP1.5b* gene were sequenced with synthetic oligonucleotide primers derived from the sequence of the first strand.

RESULTS

Far-upstream DNA sequences required for high-level, tissue-specific expression of a *MUP1.5b* transgene. The *MUP1.5b* transgene that had displayed submaxillary gland-specific expression (34) included 4.7 kb of 5'-flanking DNA and 0.3

TABLE 1. Expression of *MUP1.5b* transgenes

Transgene (kb of 5'-flanking DNA)	Transgenic line or mouse ^a	Transgene copy number ^b	mRNA level (transgenic/endogenous) ^c
pE (4.7)	1	10-20	3.7
	14	6-8	2.9
	9	2-3	0
	16	6-8	0
pH (3.46)	20	60-80	4.3
	9	30-40	3.1
	35	30-40	0.5
	47	2-3	0.5
	13	40-50	0.1
	10	8-12	0
	48	6-8	0
pG (1.85)	3	150-200	0.2
	1'	8-10	0
	3'	80-100	0
	5'	6-8	0
	5	40-50	0
	6'	10-20	0
	8	8-10	0
	10	80-100	0
	13	10-20	0
	15	6-8	0
	15'	6-8	0
pD (0.53)	35	80-120	0.2
	15	50-100	0.1
	34	20-40	0.1
	1	30-40	0
	12	10-20	0
	49	30-40	0
	54	30-40	0

^a pE transgenic lines were derived from four independent founder transgenic mice; the mRNA levels indicated are average levels from measurements for several transgenic progeny of each line (data are from reference 34). Data for pH, pG, and pD mice were all obtained from analysis of the founder transgenic mice.

^b The transgene copy number was estimated from the intensity of bands in Southern analyses relative to that of the endogenous *MUP1.5* gene present at four copies per diploid genome (two copies of *MUP1.5a* and two copies of *MUP1.5b*).

^c mRNA levels were estimated from the data in Fig. 2 (pD transgene), Fig. 5 (pH transgene), and similar RNase protection assays not shown (pG transgene). The intensity of the bands representing transgenic MUP mRNA was compared with that of the bands representing endogenous MUP1.5 mRNA by densitometry of the autoradiograms.

kb of 3'-flanking DNA (Fig. 1b, transgene pE). To determine whether an upstream enhancer was required for this pattern of expression or whether promoter and/or intragenic elements were sufficient, we produced transgenic mice carrying a truncated gene containing only 0.53 kb of 5'-flanking DNA (Fig. 1b, transgene pD). The *MUP* transgenes were marked by the insertion of an oligonucleotide into the 5'-untranslated region to distinguish their mRNA from endogenous MUP V mRNA. Seven founder transgenic mice were produced; all carried multiple intact copies of transgene pD (Table 1), as determined by Southern blot analysis (data not shown).

Expression of the transgenic MUP mRNA and endogenous MUP V mRNA was measured in the submaxillary glands of these animals with an RNase protection assay. Three of the mice (no. 34, 35, and 15) expressed the transgenic MUP mRNA at low levels, approximately 0.1 to 0.2 times the endogenous MUP V mRNA level, and four did not express the transgene (Fig. 2A and B, and Table 1). These results contrast with the results obtained with trans-

FIG. 3. Mapping of DNase I-HS sites. Purified nuclei prepared from the submaxillary glands (Submax) and kidneys of 3-week-old BALB/cByJ mice were digested with increasing amounts of DNase I (0, 1, or 3 U), and the DNase I-digested DNA was deproteinized as described in Materials and Methods. From each DNase I-digested sample, 20 μ g of DNA was restricted with *Pst*I, electrophoresed through an 0.8% agarose gel, blotted, and hybridized to a *MUP1.5a*- and *MUP1.5b* gene-specific riboprobe (33, 34) derived from the first exon. Under the conditions of hybridization and washing used (see Materials and Methods), this riboprobe hybridizes to the *MUP1.5a* and *MUP1.5b* genes (the 5.6-kb *Pst*I fragments) as well as to another unidentified *MUP* gene (the 2.5-kb *Pst*I fragment); with more stringent washes, cross-hybridization to the 2.5-kb fragment is eliminated (33). The 3.5-kb band, seen only after DNase I digestion of submaxillary gland nuclei, represents cleavage by DNase I at an HS site (Fig. 1a).

gene pE (34), which was expressed at three to four times the endogenous level in two of four transgenic lines examined (Table 1; for example, see lanes E in Fig. 2A and B). Thus, sequences more than 530 bp upstream from the cap site appeared to be required for high-level expression of this gene in the submaxillary gland. In addition, the transgenic MUP mRNA seen in the pD-expressing mice displayed more heterogeneous 5' termini than were seen with constructs containing more 5'-flanking DNA (Fig. 2A, lane E; see also Fig. 5).

To determine whether the promoter and intragenic sequences might be sufficient for low-level, tissue-specific expression, we examined the distribution of transgenic MUP V mRNA in several other tissues of the transgenic mouse expressing the highest level of pD, number 35. We found that the transgene was expressed in at least two inappropriate tissues, the liver and the kidney, at levels only slightly below that in the submaxillary gland (Fig. 2C). Therefore, sequences located between 0.53 and 4.7 kb from the cap site are required for submaxillary gland-specific as well as high-level expression.

Tissue-specific DNase I-HS site in the *MUP1.5a* 5'-flanking region. The location of DNase I-HS sites in chromatin has often provided an accurate prediction of the position of enhancers or other regulatory DNA sequences (3, 13, 17, 18). Therefore, to help locate an enhancer within the upstream flanking region of the *MUP1.5* genes, we tested for the presence of tissue-specific DNase I-HS sites. Nuclei from the submaxillary glands or kidneys of normal BALB/cByJ mice were treated with DNase I, DNA was purified and digested with *Pst*I, and Southern analysis was performed. The probe was a riboprobe which was derived from the first exon (33, 34) and which hybridizes to the same 5.6-kb *Pst*I fragment from the *MUP1.5a* and *MUP1.5b* genes, as well as to a weaker 2.5-kb band apparently due to cross-hybridization with another unidentified *MUP* gene. A single DNase-dependent band of 3.5 kb was detected with DNA from the submaxillary gland but not with DNA from the kidney (Fig.

```

      10      30      50      70
AGTGTGGGA CAACTCTTAC TCAGTCTAGC CTATATCCTT TTTGCAACCT TGCAGTCAAC AGTCTTCAAA
      90      110      130
AGATAAAATT TATAGTGCTA ACATTAATGA TGTATGTTGG TTACCATGCA GGATACCATC TTCTGTTTTC
      150      170      190      210
TGTTGAGCCC AACCAAGACA GTATCCCTCA ATCAGTTTTT CTGTACTTTG GTCCCCGTGT CAGGCACAAA
      230      250      270
CTCTACTGAT GTTGGGAGCT ATTAAGGCAA CGCTATTGTC CTGATCTCTG AATTGGGCCT CTCCCCCTGA
      290      310      330      350
GAAGAAAAGG GGGTCAAAAG CGGGCCACCA GATGCACCGC TCTGAGAACA ACCACAGATT GTTCCAGCCC
      370      390      410
TAAGTCAGCA CTGGATGTCC TGACCTCAAG ATGTACCCTG ATACCACCAA GTTCCTGCTT CCCCAGAGTAA
      430      450      470      490
TCGCCAAAAG AAAAATTTCT GCTGCCCAT CCCTCCTGCT GAGGTGCATT TCCTCTGCCC CATTCCTCCT
      510      530      550
GCTGAAAGGC ACTTCCCCTT TTGCTTGTGC ATTTAAACCT CGGGCCTGGC TAGTACATT GGGGTCTTGA
      570      590      610      630
TGCAATCCAG AACGGTTTCT GTGTCGTTAT TCCCACAAGA CCCTCGTCTC TCTCTACTCC CCCCCCCCCC
      650      670      690
AATTTGGTTA TTAGGGGATG GGAAGAACAG CAAGAGAAAT AAGGAGCCAA GACAAAGTTC TCTGATCAAG
      710      730      750      770
GCTCAATCTT TAATGTATGG CACTGAATAC ATATGGGAAA ACCCAAGGAC CTATGCCTTT GTTTCAGCTG
      790      810      830
GGTCTGTGTG CTGTGTTTAG TTAGATTGAG TTGAAAAATA GGCACCTTCAT CAGCTCAGCA TGTGGCAGTT
      850      870      890      910
CTCAGTTGTC TGACAGGAAG ATATAATTGT ATCAAAGCAA GACATGTAGC TCCAGTGTAC TAGACTTGTT
      930      950      970
AAGCCAGTTC AAGGCTAAGG GGTGGGGGTA GGGGTACAC TCTGACACTG ACACAGCAGG TGTTCAGAA
      990      1010      1030      1050
CTAATGTTGA GCCAGAGTCA GCAGGATGTA CTCTCACTTT GGCTGCTAGA GAGACTACAG AATATTTCTT
      1070      1090      1110
CTCTGGGACA TGAATCAAAT TAGCCTGTGT TCATTGATGT TCTTCTTGGC TCCAAATATA CCTGACATCA
      1130      1150      1170      1190
TATAAGTGGC TTGAATGTGT CCACAGGATT TCTGAATGAG TTATTTGTTT AGTTAGTTGA TGGTGAGAGA
      1210      1230      1250
CAGGACAAGG TCATTAAATC ATTTTCCTAT GAGTCAGTTT CAGAAATGCT AGCTAGTTAC AAATAGGCTG
      1270      1290      1310      1330
TTTGTAGGTT ACACGTAAAC TCCTTAAACC CCATCCACAA ACACATTGTC TAGTCAATTC TTCACTGCTT
      1350      1370      1390
GAGCTTTTCA GTCACATGAA CAGTCTACAC ATGCTAAGGA GTGTTAACAT GCTTAGTGAC AGGGAAGGTG
      1410      1430      1450      1470
CCTCAGCTGA AGTCAGGAGT GGAAGGTACT CATTGCTGCT TTCTGCTGTA GAAGCAACTT GAAGTTCATT
      1490      1510      1530
ACTAGAAAAA CAATAGTCCA TAGCCAGATG AGGTGCTACA ACCATAATC CCCGAGTTCA GGAAGCTGGA
      1550      1570      1590
GCAGGAGGAT TGCCACAGAT CTAAGACCAT TCAGAAGCTC AGTGTAAGAA CAAGTCCAGC TCAGAGCTC

```

FIG. 4. Nucleotide sequence of an *EcoRI*-*SstI* fragment from the upstream region of the *MUPI.5a* gene containing the DNase I-HS region. The nucleotide sequence of the *MUPI.5b* gene is identical in this region. The A at position 1 is the first nucleotide following the *EcoRI* site, and position 1609 is the last nucleotide of the *SstI* site. Sequences with homology to consensus binding sites for known nuclear factors are mentioned in the text.

3). This result indicated that either the *MUPI.5a* or the *MUPI.5b* gene or both contain a tissue-specific DNase I-HS site 2.5 kb upstream from the transcription initiation site (Fig. 1a). As the *MUPI.5b* gene is within a DNase-resistant region of chromatin (11a), the DNase I-HS site that we observed is presumably in the 5'-flanking region of the active *MUPI.5a* gene.

The nucleotide sequences of a 1.6-kb region upstream from the *MUPI.5a* and *MUPI.5b* genes, including the DNase I-HS site, are identical. To determine whether the nucleotide sequences of the DNase I-HS region differed between the active *MUPI.5a* gene and the silent *MUPI.5b* gene, we sequenced the entire 1.6-kb *EcoRI*-*SstI* fragment upstream from each gene. The nucleotide sequences, presented in Fig. 4, were identical for the two *MUPI.5* genes. Therefore, any regulatory elements associated with the DNase I-HS site (which we estimate to map between nucleotides 785 and 1085 of the 1,609-nucleotide sequence) must be present in the silent as well as in the active gene.

Truncation of 5'-flanking sequences, including the DNase I-HS site, abolishes the expression of a *MUPI.5b* transgene.

We next tested the hypothesis that the DNase I-HS region contains positive regulatory elements required for *MUPI.5* gene expression. For this purpose, we used two additional *MUPI.5b* transgene constructs, one truncated at the *EcoRI* site beyond the DNase I-HS region (pH) and one truncated at the proximal *SstI* site and lacking the DNase I-HS region (pG) (Fig. 1b). Of seven founder transgenic mice carrying the pH transgene, two (no. 9 and 20) expressed the gene in the submaxillary gland at high levels, three to four times that of the endogenous gene, two (no. 35 and 47) expressed the gene at low to moderate levels, and three showed no expression (Fig. 5 and Table 1). The levels of expression and the fractions of transgenic mice expressing this gene were not significantly different from those observed with the longer pE transgene, suggesting that sequences between the *EcoRI* site at -3.46 kb and the *HindIII* site at -4.7 kb do not include any essential regulatory elements.

In contrast, none of 11 founder mice carrying the shorter pG transgene expressed the gene at high levels in the submaxillary gland, similar to the results obtained with construct pD (Table 1; data not shown). Thus, the removal

FIG. 5. High-level expression of *MUP1.5b* transgene pH in mouse submaxillary glands. Seven independent founder transgenic mice carrying transgene pH (lanes 9 to 48) were identified by Southern analysis, and RNA isolated from the submaxillary glands at 3 weeks of age was used in RNase protection assays (Fig. 2). Lane C, submaxillary gland RNA from a nontransgenic control mouse; lane E, submaxillary gland RNA from a transgenic mouse of line pE14; m, molecular weight markers; Tr, protected fragments representing mRNA from the *MUP1.5b* transgene; En, protected fragments representing mRNA from the endogenous *MUP1.5a* gene. Numbers at right represent nucleotides.

of sequences between the *Sst*I site at -1.85 kb and the *Eco*RI site at -3.46 kb, including the DNase I-HS site, appeared to eliminate an essential positive regulatory element.

The 1.6-kb *Eco*RI-*Sst*I fragment, including the DNase I-HS site, is required for tissue-specific expression of a *MUP1.5b*-CAT fusion gene. To confirm the presence of a tissue-specific enhancer in the vicinity of the DNase I-HS site, we next tested three additional constructs (Fig. 1b), all containing the *MUP1.5b* promoter fused to the CAT reporter gene (16). Construct pDCAT included only MUP promoter sequences (up to -530) fused to the CAT gene, while the other two constructs contained, in addition, an upstream MUP DNA fragment, either the 1.6-kb *Eco*RI-*Sst*I fragment (including the DNase I-HS site and putative enhancer [pDC-1.6]) or the more distal 1.3-kb *Hind*III-*Eco*RI fragment, which appeared to be dispensable for expression of the *MUP1.5b* gene (pDC-1.3).

Of 12 founder transgenic mice carrying the pDCAT gene, none showed any detectable CAT gene expression in the submaxillary gland (Fig. 6a). This result was consistent with our observation that promoter sequences up to -530 were not sufficient for tissue-specific expression of a *MUP1.5b*

transgene. The addition of the upstream 1.3-kb *Hind*III-*Eco*RI fragment had no discernible effect, as all eight founder transgenic mice carrying pDC-1.3 failed to express the CAT gene in the submaxillary gland (Fig. 6b). However, the addition of the 1.6-kb *Eco*RI-*Sst*I fragment resulted in CAT gene expression in the submaxillary gland in 5 of 14 founder transgenic mice examined (Fig. 6c). Densitometry of the monoacetylated chloramphenicol spots indicated that the levels of CAT activity in positive samples were 15 to 25 times the background level. In two of these pDC-1.6 transgenic mice (no. 10 and 11), CAT activity was also measured in four other tissues, and in both cases CAT gene expression was restricted to the submaxillary gland. Although the level of CAT activity was very low (see Discussion), these results demonstrate that the 1.6-kb *Eco*RI-*Sst*I fragment includes an enhancer that can stimulate transcription from the otherwise inactive *MUP1.5b* promoter in a tissue-specific pattern.

DISCUSSION

The *MUP* gene family is unusual in that its members have highly conserved DNA sequences yet display quite divergent patterns of tissue-specific, developmental, and hormonal regulation. The use of transgenic mice provides a powerful method by which to experimentally define the locations of *cis*-acting regulatory elements responsible for the differential expression of various *MUP* genes. To date, transgenic studies with normal *MUP* genes or *MUP* fusion genes (1, 21, 34), as well as members of the homologous rat α -2u globulin gene family (10), have established that several members of the gene family, including those expressed in the liver as well as those expressed in the submaxillary gland, are controlled by closely linked *cis*-acting elements. However, none of these elements or the factors with which they interact have yet been identified. In this paper, we have provided the first evidence for the location of an enhancer responsible for the tissue-specific expression of a subfamily of *MUP* genes.

By testing the expression of a series of *MUP1.5b* transgenes with progressive truncations of 5'-flanking DNA, we defined a 1.6-kb region that contains an essential positive regulatory element. Transgenes that included this region (pE and pH, including 4.7 and 3.46 kb of 5'-flanking DNA, respectively) were frequently expressed at high levels in the mouse submaxillary gland, while those lacking this region (pD and pG) were inactive or were expressed only at very low levels. The facts that even the two larger transgenes failed to be expressed in approximately half of the transgenic mice produced and that their expression did not correlate with transgene copy number (Table 1) suggest that even the largest transgene that we have tested is sensitive to chromosomal position effects; i.e., the enhancer is not sufficient to confer position independence.

The 1.6-kb DNA fragment, suggested by the truncation experiments to contain an enhancer, was found to confer a submaxillary gland-specific pattern of expression when added to a fusion gene containing the *MUP1.5b* promoter (-530 bp) and a CAT reporter gene. Neither the fusion gene alone nor the fusion gene plus a more distal upstream *Hind*III-*Eco*RI fragment showed any expression in the submaxillary gland. This result provided additional support for the presence of an enhancer that is active in the submaxillary gland. It should be noted that our experiments have not ruled out the possibility that elements within the -530 MUP promoter fragment may be required, in addition to upstream

FIG. 6. Expression of *MUP1.5b*-CAT fusion genes in transgenic mice. Transgenic mice carrying the fusion genes pDCAT (a), pDC-1.3 (b), and pDC-1.6 (c) were identified by Southern analysis (data not shown), and submaxillary gland tissue extracts from the 3-week-old founder transgenic mice were assayed for CAT activity (acetylation of [14 C]chloramphenicol). Each numbered lane in panels a, b, and c represents the submaxillary gland from an independent founder transgenic mouse. Panel d shows assays of CAT activity in five tissues from pDC-1.6 transgenic mouse numbers 10 and 11: Sm, submaxillary gland; SI, sublingual gland; La, lachrymal gland; Li, liver; K, kidney. Ch, unacetylated chloramphenicol substrate; A, monoacetylated chloramphenicol, indicative of CAT activity. The other bands seen in some samples (migrating between Ch and A) are degraded chloramphenicol; these were a consequence of the long incubation times required to see CAT activity and were not products of CAT, as they were also obtained with control submaxillary gland extracts from nontransgenic mice (panel e, lanes 1 to 7). Lane C in panels a, b, and d contains positive control submaxillary gland extract from transgenic mouse pDC-1.6-10.

enhancer sequences, for specific expression in the submaxillary gland.

The levels of CAT activity in the transgenic submaxillary gland appeared quite low on the basis of the incubation times required to detect CAT activity. The significance of this observation is unclear, because it is difficult to compare levels of CAT activity with levels of MUP mRNA measured in other experiments. The low level of CAT activity may be attributed in part to the fact that MUP V mRNA in the submaxillary gland is normally found at a relatively low level, in comparison with MUP expression in other tissues, such as the liver (25, 30, 32). On the other hand, it suggests that intragenic *MUP1.5* sequences which were absent from the CAT fusion gene may be required for normal levels of expression in the submaxillary gland; these might include additional transcriptional elements or sequences affecting MUP mRNA stability.

Consistent with the functional evidence for an upstream enhancer, we observed a tissue-specific DNase I-HS site within the corresponding region of the endogenous *MUP1.5a* gene. This site is likely to represent the site at which regulatory factors in submaxillary gland nuclei bind to the enhancer. Our preliminary experiments indicated that a

DNA fragment corresponding to nucleotides 766 to 1043 of the *EcoRI*-*SstI* fragment contains protein binding sites, and experiments to define these sites and the respective factors are in progress. A computer analysis of this region revealed several DNA sequences with homology to known consensus sequences and representing potential binding sites. For example, a consensus CTF/NF-1 binding site (TGGCA [23, 24, 27]) and an imperfect copy of this sequence are present beginning at positions 833 and 851, respectively (Fig. 4). At position 931 is the sequence GGTGGG, which differs by 1 nucleotide from the Sp1 consensus binding sequence GGCGGG (6). At position 820 is the sequence TCAGCTCA, which differs by 1 nucleotide from the estrogen-thyroid hormone response element half-site TCAGGTCA (14). The sequence TGACACA at position 959 differs by 1 nucleotide from the consensus serum response element TGACTCA (35). Finally, the partially overlapping sequences TGAGC CAG and AGAGTCAG at positions 988 and 994, respectively, each differ by 1 nucleotide from the consensus phorbol ester response element TGAGTCAG (2). Of course, the functional significance of these sequence homologies remains to be determined.

As we have shown before (34), the *MUP1.5b* gene is

transcriptionally silent, and unpublished data (11a) indicate that it is in a DNase-I-insensitive conformation. Because the functional evidence for an enhancer was obtained with the normally silent *MUP1.5b* gene, it was important to compare the sequences of the two *MUP1.5* genes in this region. DNA sequence analysis revealed that the corresponding 1.6-kb *EcoRI-SstI* fragments upstream from the *MUP1.5a* and *MUP1.5b* genes were indeed identical. This result provides further support for our previous hypothesis that the endogenous *MUP1.5b* gene is silenced as a consequence of its chromosomal position, not as a result of mutations in promoter or enhancer elements.

The identification of an enhancer involved in the submaxillary gland-specific expression of the *MUP1.5* gene subfamily now provides an opportunity to compare the regulatory elements of *MUP* genes expressed in different tissues with structural as well as functional criteria. The corresponding region upstream from one of the liver-specific *MUP* genes, encoding the MUP III mRNA, has been sequenced, and it is approximately 85% similar to that of the *MUP1.5* genes, differing only by scattered single-base substitutions, insertions, and deletions (28a). This result raises several interesting questions, such as whether the same region of the liver-specific gene also displays enhancer activity, whether this or a different element assumes the role of enhancer for the liver-specific gene, whether changes in other elements (e.g., the promoter) are responsible for liver versus submaxillary gland expression, and which of the sequence changes are responsible for altered tissue specificity.

ACKNOWLEDGMENTS

We thank Raymond MacDonald for providing plasmid ElmcAT. This research was supported by Public Health Service grants CA22376 (to F.C.) and AG07583 (to E.D.), by grant CD315 from the American Cancer Society (to E.D.), and by a Faculty Research Award from the American Cancer Society (to F.C.).

REFERENCES

- Al Shawi, R., J. Burke, C. T. Jones, J. P. Simons, and J. O. Bishop. 1988. A Mup promoter-thymidine kinase reporter gene shows relaxed tissue-specific expression and confers male sterility upon transgenic mice. *Mol. Cell. Biol.* 8:4821-4828.
- Angel, P., M. Imagawa, R. Chiu, B. Stein, H. J. Rahmsdorf, C. Jonat, P. Herrlich, and M. Karin. 1987. Phorbol ester-inducible genes contain a common *cis* element recognized by a TPA modulated *trans*-acting factor. *Cell* 49:729-739.
- Banerji, J., L. Olson, and W. Schaffner. 1983. A lymphocyte-specific cellular enhancer is located downstream of the joining region in immunoglobulin heavy chain genes. *Cell* 33:729-740.
- Bellard, M., G. Dretzen, F. Bellard, P. Oudet, and P. Chambon. 1982. Disruption of the typical chromatin structure in a 2500 base-pair region at the 5' end of the actively transcribed ovalbumin gene. *EMBO J.* 1:223-230.
- Bishop, J. O., G. G. Selman, J. Hickman, L. Black, R. D. P. Saunders, and A. J. Clark. 1985. The 45-kb unit of major urinary protein gene organization is a gigantic imperfect palindrome. *Mol. Cell. Biol.* 5:1591-1600.
- Briggs, M. R., J. T. Kadonaga, S. P. Bell, and R. Tjian. 1986. Purification and biochemical characterization of the promoter-specific transcription factor, Sp1. *Science* 234:47-52.
- Clark, A. J., A. Chave-Cox, X. Ma, and J. O. Bishop. 1985. Analysis of mouse major urinary protein genes: variation between the exonic sequences of group 1 genes and a comparison with an active gene without group 1 both suggest that gene conversion has occurred between MUP genes. *EMBO J.* 4:3167-3171.
- Clark, A. J., P. Ghazal, R. W. Bingham, D. Barrett, and J. O. Bishop. 1985. Sequence structures of a mouse major urinary protein gene and pseudogene compared. *EMBO J.* 4:3159-3165.
- Clark, A. J., J. Hickman, and J. O. Bishop. 1984. A 45-kb domain with two divergently orientated genes is the unit of organization of the murine major urinary protein genes. *EMBO J.* 3:2055-2064.
- da Costa Soares, V., R. M. Gubits, P. Feigelson, and F. Costantini. 1987. Tissue-specific and hormonally regulated expression of a rat α_{2u} globulin gene in transgenic mice. *Mol. Cell. Biol.* 7:3749-3758.
- Derman, E. 1981. Isolation of a cDNA clone for mouse urinary proteins: age- and sex-related expression of mouse urinary protein genes is transcriptionally controlled. *Proc. Natl. Acad. Sci. USA* 78:5425-5429.
- Derman, E. Unpublished data.
- Finlayson, J. S., R. Asofsky, M. Potter, and C. C. Runner. 1965. Major urinary protein complex of normal mice: origin. *Science* 149:981-982.
- Gillies, S. D., S. L. Morrison, V. T. Oi, and S. Tonegawa. 1983. A tissue-specific transcriptional enhancer element is located in the major intron of a rearranged immunoglobulin heavy chain gene. *Cell* 33:717-728.
- Glass, C. K., J. M. Holloway, O. V. Devary, and M. G. Rosenfeld. 1988. The thyroid hormone receptor binds with opposite transcriptional effects to a common sequence motif in thyroid hormone and estrogen response elements. *Cell* 54:313-323.
- Godovac-Zimmerman, J. 1988. The structural motif of beta-lactoglobulin and retinol-binding protein: a basic framework for binding and transport of small hydrophobic molecules? *Trends Biochem. Sci.* 13:64-66.
- Gorman, C. M., G. T. Merlino, M. C. Willingham, I. Pastan, and B. H. Howard. 1982. The Rous sarcoma virus long terminal repeat is a strong promoter when introduced into a variety of eukaryotic cells by DNA-mediated transfection. *Proc. Natl. Acad. Sci. USA* 79:6777-6781.
- Gross, D., and W. Garrard. 1988. Nuclease hypersensitive sites in chromatin. *Annu. Rev. Biochem.* 57:159-197.
- Grosveld, F., G. B. van Assendelft, D. R. Greaves, and G. Kollias. 1987. Position-independent, high-level expression of the human beta-globin gene in transgenic mice. *Cell* 51:975-985.
- Hastie, N. D., W. A. Held, and J. J. Toole. 1979. Multiple genes coding for the androgen-regulated major urinary proteins of the mouse. *Cell* 17:449-457.
- Held, W. A., J. F. Gallagher, C. M. Hohman, N. J. Kuhn, B. M. Sampsel, and R. G. Hughes, Jr. 1987. Identification and characterization of functional genes encoding the mouse major urinary proteins. *Mol. Cell. Biol.* 7:3705-3712.
- Held, W. A., J. J. Mullins, N. J. Kuhn, J. F. Gallagher, G. D. Gu, and K. W. Gross. 1989. T antigen expression and tumorigenesis in transgenic mice containing a mouse major urinary protein/SV40 T antigen hybrid gene. *EMBO J.* 8:183-191.
- Hogan, B., F. Costantini, and E. Lacy. 1986. Manipulating the mouse embryo: a laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
- Jones, K. A., J. T. Kadonaga, P. J. Rosenfeld, T. J. Kelly, and R. Tjian. 1987. A cellular DNA binding protein that activates eukaryotic transcription and DNA replication. *Cell* 48:79-89.
- Jones, K. A., K. R. Yamamoto, and R. Tjian. 1985. Two distinct transcription factors that bind to the HSV thymidine kinase promoter *in vitro*. *Cell* 42:559-572.
- Knopf, J. L., J. F. Gallagher, and W. A. Held. 1983. Differential, multihormonal regulation of the mouse major urinary protein gene family in the liver. *Mol. Cell. Biol.* 3:2232-2240.
- Kruse, F., C. T. Komro, C. H. Michnoff, and R. J. MacDonald. 1988. The cell-specific elastase I enhancer comprises two domains. *Mol. Cell. Biol.* 8:893-902.
- Nowock, J., U. Borgmeyer, A. W. Puschel, R. A. W. Rupp, and A. E. Sippel. 1985. The TGGCA protein binds to the MMTV-LTR, the adenovirus origin of replication, and the BK virus enhancer. *Nucleic Acids Res.* 13:2045-2061.
- Perez-Stable, C., and F. Costantini. 1990. Roles of fetal γ -globin promoter elements and the adult β -globin 3' enhancer in the stage-specific expression of globin genes. *Mol. Cell. Biol.* 10:1116-1125.

- 28a. **Rodriguez, M., and E. Derman.** Unpublished data.
29. **Sanger, F., S. Nicklen, and A. R. Coulson.** 1977. DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. USA* **74**:5463–5467.
30. **Shahan, K., M. Denaro, M. Gilmartin, Y. Shi, and E. Derman.** 1987. Expression of six mouse major urinary protein genes in the mammary, parotid, sublingual, submaxillary, and lachrymal glands and in the liver. *Mol. Cell. Biol.* **7**:1947–1954.
31. **Shahan, K., M. Gilmartin, and E. Derman.** 1987. Nucleotide sequences of liver, lachrymal, and submaxillary gland mouse major urinary protein mRNAs: mosaic structure and construction of panels of gene-specific synthetic oligonucleotide probes. *Mol. Cell. Biol.* **7**:1938–1946.
32. **Shaw, P. H., W. A. Held, and N. D. Hastie.** 1983. The gene family for major urinary proteins: expression in several secretory tissues of the mouse. *Cell* **32**:755–761.
33. **Shi, Y., M. Rodriguez, K. Shahan, and E. Derman.** 1989. Subfamily of submaxillary gland-specific Mup genes: chromosomal linkage and sequence comparison with liver-specific Mup genes. *Nucleic Acids Res.* **17**:6191–6203.
34. **Shi, Y., H. J. Son, K. Shahan, M. Rodriguez, F. Costantini, and E. Derman.** 1989. Silent genes in the mouse major urinary protein gene family. *Proc. Natl. Acad. Sci. USA* **86**:4584–4588.
35. **Treisman, R.** 1986. Identification of a protein-binding site that mediates transcriptional response of the *c-fos* gene to serum factors. *Cell* **46**:567–574.