Mouse Prostasin Gene Structure, Promoter Analysis, and Restricted Expression in Lung and Kidney

George M. Verghese, Z. Y. Tong, Vikash Bhagwandin, and George H. Caughey

Department of Medicine, University of Virginia, Charlottesville, Virginia; and Cardiovascular Research Institute, University of California, San Francisco, California

Human prostasin is a membrane-anchored serine peptidase hypothesized to regulate lung epithelial sodium transport. It belongs to a unique family of genes on chromosome 16p11.2/13.3. Here we describe genomic cloning, promoter analysis, and expression of prostasin’s mouse ortholog. The 4.3-kb mouse prostasin gene (prss8) has a six-exon organization identical to human prostasin. Prss8 spans two signal tagged-sites localized to chromosome 7. Multiple mRNA transcripts arise from two consensus initiator elements of a TATA-less promoter and an alternatively spliced, 5′ untranslated region intron. Reporter assay establishes that the initiator elements and a GC-rich domain comprise the core promoter and identifies 5′ flanking regions with strong enhancer and repressor activity. The 3′ untranslated region overlaps the 3′ untranslated region of the Myst1 gene oriented tail-to-tail at this locus. Prss8 is highly transcribed in pancreas, kidney, submaxillary gland, lung, thyroid, prostate, and epididymis, and is developmentally regulated. Using selective riboprobes and antibodies to mouse prostasin, we localized its expression to lung airway epithelial and alveolar type II cells and kidney cortical tubule epithelium. Mouse prostasin highly resembles its human ortholog in gene organization and tissue specificity, including strong expression in pulmonary epithelium, suggesting that mice will be useful for probing prostasin’s functions in vivo.

Serine proteases play critical roles in diverse processes such as coagulation, complement activation, receptor activation and tissue remodeling. A striking variety of serine proteases have evolved from shared ancestral genes to subserve specific functions. In prior studies we mapped a locus of mast cell tryptase genes on human chromosome 16p13.3 and identified the novel, related serine protease genes, tryptase-ε and pancreasin, in this locus (1–3). The organization of tryptase-ε and pancreasin genes is similar to human prostasin (PRSS8), testisin (PRSS21), and tryptase-ε (PRSS22) (4–7), which also map to chromosome 16p. These genes have a unique six-exon structure with a very short second exon, distinct from that of α- and β-tryptases, trypsin, and other known serine proteases (8). Tryptase-ε, prostasin, testisin, pancreasin, and tryptase-ε are ~ 40% identical to α- and β-tryptases in amino acid sequence, but have several distinguishing features, including a tryptic propeptide cleavage site, conserved cysteines predicted to form a disulfide link between propeptide and catalytic chain, and (with the exception of pancreasin and tryptase-ε) a C-terminal transmembrane domain (2, 3, 7, 9, 10). These data suggest that tryptase-ε, pancreasin, prostasin, testisin, and tryptase-ε share a recent ancestral gene and form a distinct branch of the trypsin-like serine protease family tree. These proteases are differentially expressed, indicating that their expression and activity are tightly regulated and cell-specific (2, 3, 6, 9, 11), but the mechanisms regulating expression of these genes and their protease activity are poorly defined at present.

The functions of this family of serine proteases are just beginning to be discerned and their biological substrates and inhibitors are unknown. Prostasin (also termed channel-activating protease 1 [CAP1]) activates the epithelial sodium channel (ENaC) by a protease-dependent mechanism (12). Prostasin is one of several candidate trypsin serine peptidases proposed to regulate ENaC function in human lung and kidney (13, 14). ENaC activity is tightly regulated in the lung and is critical to maintaining normal airway and alveolar surface liquid volume (15, 16). Recent studies show that apical serine protease inhibition decreases sodium transport in primary human airway epithelial cells (17). Overexpression of prostasin induces hypertension in rats, indicating that protease regulation may be an important pathway in overall body fluid homeostasis (18). In addition, prostasin and testisin genes are downregulated in some human tumors and have been proposed to be tumor suppressors (11, 19, 20). The precise mechanisms and physiological significance of these prostasin activities remain to be determined. Tryptase-ε is expressed in mast cell secretory granules and induces airway hyperreactivity in mice (2, 21). Tryptase-ε is expressed in lung epithelial cells and not in mast cells, but its functions are not yet known (7). In the present work, as a step toward deciphering the biological functions of prostasin and mechanisms regulating its expression, we report sequence and organization of the mouse prostasin gene and flanking regions, chromosomal localization, promoter function, tissue distribution, and cell-specific expression in lung and kidney.

Materials and Methods

Materials

Chemicals were obtained from Sigma (St. Louis, MO) and Fisher Scientific (Pittsburgh, PA). Oligonucleotide primers were obtained
from the University of California, San Francisco Biomolecular Resource Center, Genemed Synthesis (South San Francisco, CA) and Operon (Fremont, CA). DNA and RNA purification columns were obtained from Qiagen (Valencia, CA). Mouse lung and kidney RNA were obtained from Ambion (Austin, TX). Kits for 3′ rapid amplification of cDNA ends (RACE) and RNA ligase-mediated RACE were purchased from Life Technologies (Gaithersburg, PA) and Ambion (Austin, TX), respectively. AmpliTaq Gold reverse transcriptase was obtained from PE Biosystems (Foster City, CA). PCR was conducted with an MJ Research PTC-200 thermal cycler (South San Francisco, CA). Restriction enzymes were obtained from New England Biolabs (Beverly, MA). PCR2.1-TOPO–TA cloning vector was obtained from Invitrogen (Carlsbad, CA); pGL3 Basic and pRLTK luciferase reporter vectors were obtained from Promega (Madison, WI). RNA blots were purchased from BD Biosciences-Clontech (Palo Alto, CA). 32P-dCTP was purchased from Amersham Pharmacia Biotech (Buckinghamshire, UK). Biotin-ctCTP was obtained from Roche Diagnostics (Mannheim, Germany); tyramide signal amplification reagents were obtained from NEN (Boston, MA). Anti-rabbit secondary antibodies were obtained from Cell Signaling Technology (Beverly, MA) and Vector Laboratories (Burlingame, CA). Strepa-dixylin-Texas Red or fluorescein isothiocyanate conjugates and immunohistochemistry staining reagents were purchased from Zymed Laboratories (South San Francisco, CA). M-1 cells were obtained from American Type Cell Culture (Manassas, VA). Cell culture media was obtained from BioWhittaker (Walkersville, MD) and serum was from Sigma. Lipofectamine 2000 was from Invitrogen. Dual Glo Luciferase Assay was from Promega. Luciferase activity was measured using a Bio-Tek Instruments Flx800TBI plate reader (Winooski, VT).

DNA Amplification, Sequencing, and Analysis
PCR products were separated by TAE (40 mM Tris acetate, 1 mM EDTA) agarose gel electrophoresis and stained with ethidium bromide. PCR products were sequenced directly or gel-purified using QiAquick gel extraction columns, subcloned into pcR2.1-TOPO-TA and transformed into Escherichia coli INVoZ™. Bacteria were cultured in Luria Bertani broth supplemented with 50 µg/ml ampicillin. Plasmid DNA was purified from bacterial extracts using Qiagen Miniprep columns and screened for desired DNA inserts by plasmid restriction digest or PCR. DNA sequencing was done by the University of California Biomolecular Resource Center via standard dideoxynucleotide technique. PCR primers were designed using MacVector and Primer3 (www-genome.wi.mit.edu/cgi-bin/primer/primer3_www.cgi). Sequences were aligned and analyzed using MacVector software (Oxford Molecular Group, Beverly, MA) or online utilities from the National Center for Biotechnology Information (www.nlm.nih.gov).

Genomic Cloning of Mouse Prostasin Gene
A bacterial artificial chromosome (BAC) library of 129Sv/J mouse genomic DNA was screened by PCR for the mouse prostasin gene (prss8) by Genome Systems (St. Louis, MO) using specific primers MPFRS (5′-CTAGACTCTCCGAACTGACAC-3′) and MPFRS (5′-GTGGGCTAAGATACAACTGG-3′). BAC DNA clones in pBeloBAC11 were column purified from LB-chloramphenicol (50 µg/ml) bacterial cultures as directed by the manufacturer (Qiagen Maxi Prep). A ~4-kb sequence spanning the coding region of the mouse prostasin gene was amplified by PCR from BAC26 using primers MPFR1 (5′-GGATTCCCTCCTCTAGTCTACCACA-3′) and MPRR1 (5′-TGACTGGAATTTTGGAGGTACG-3′) and subcloned into pcR2.1 for sequencing. DNA was amplified using AmpliTaq Gold as follows: 95°C for 7.5 min; 94°C for 30 s, 62°C for 30 s, 72°C for 3 min for 35 cycles; and 72°C for 6 min. Adjacent 5′ and 3′ flanking regions were sequenced directly from pBeloBAC11-BAC26 by plasmid walking. Prostasin genomic sequence was compared with human prostasin gene using MacVector, and to GenBank databases using BLASTn at NCBI. Signal-tagged sites were identified using the NCBI Virtual PCR application. Consensus transcription factor binding sites were identified using Transcriptional Element Search Software (TESS) for comparison to TRANSFAC 4.0 database.

Full-Length cDNA Cloning and Mapping Transcription Initiation Sites by RACE
Before publication of mouse prostasin sequences in GenBank, partial sequence encoding mouse prostasin (mProstasin) cDNA was identified from the mouse expressed sequence tag (EST) database using the Basic Local Alignment Search Tool (BLAST) (www.ncbi.nlm.nih.gov) to search for sequences homologous to human Prostasin (hProstasin; L41351). cDNA clones spanning the predicted coding sequence and 3′ untranslated region (UTR) were obtained by 3′ rapid amplification of cDNA ends (RACE) using a Life Technologies kit with an oligo-dT primer. Based on EST sequences, gene-specific mProstasin primers MPRI (5′-GGATT CCTACCTGGAATTTTGGAGGTACG-3′) and MPRI (5′-TGACTGGAATTTTGGAGGTACG-3′) were used for PCR and RACE. For 3′-RACE adult mouse lung total RNA was reverse-transcribed with Superscript II. cDNA was amplified under the following conditions: 95°C for 10 min; 94°C for 30 s, 64°C for 1 min, 72°C for 2 min for 35 cycles; and 72°C for 5 min. 3′-RACE products were cloned into pcR2.1-TOPO-TA to generate the plasmid pcR2.1-mProstasin and confirmed by sequencing.

RNA ligase-mediated 5′ RACE (RLM-RACE) was done from mouse kidney polyA RNA using the Ambion FirstChoice RLM-RACE kit as per the manufacturer’s directions. Accuracy of RLM-RACE for identifying transcriptional start sites has been validated by comparison with primer extension and RNase protection assays (23). Briefly, mRNA was treated with tobacco acid pyrophosphatase to remove the 5′ cap, ligated 5′ to an RNA adapter, and reverse-transcribed using random decamer primers. 5′ cDNA end was amplified using nested forward adapter primers supplied by the manufacturer, which anneal to the 5′ adapter sequence, and nested gene-specific reverse primers MPRI-E4R (5′-CCCTGGAA GTAGCATCCACCA) and MPRI-E3R (5′-CCACCCTATTAT TTAGACCC) annealing to exons 4 and 3, respectively. PCR products were subcloned into pcR2.1-TOPO-TA and five independent clones were sequenced.

mRNA Blotting
Mouse multiple tissue polyA RNA dot blot was obtained from BD Biosciences-Clontech. A 735-bp mProstasin cDNA probe was generated by digesting pcR2.1-mProstasin with NcoI, gel purifying the mProstasin fragment, and random prime-labeling with γ32P-dCTP. The membrane was prehybridized in ExpressHyb solution (BD Biosciences-Clontech) for 1–2 h at 65°C and subsequently incubated with 1.0–1.5 × 106 cpm/ml of labeled probe for 18 h with constant agitation. Blots were washed successively in 300 mM NaCl/30 mM sodium citrate/0.5% SDS at room temperature and 15 mM NaCl/1.5 mM sodium citrate/0.1% SDS at 50°C, and exposed to Kodak XR film with one intensifying screen at −70°C.

In Situ Hybridization
Riboprobes to unique sequences in the 5′ and 3′ untranslated regions of mouse prostasin cDNA were generated by PCR and in
were subjected to identical treatment and image processing. Sections were incubated with 1:75 streptavidin:Texas Red conjugate or 1:150 streptavidin:horseradish peroxidase (HRP) conjugate for 30 min at room temperature, and 1:50 biotin tyramide solution for 10 min at room temperature. After washing with TBST sections were incubated with avidin-HRP conjugate for 30 min at room temperature and post-fixed with 4% paraformaldehyde. Endogenous peroxidase activity was quenched with methanol/3% H₂O₂. Avidin, biotin, and sections were then treated with 0.05% pepsin in 0.2N HCl for 20 min at room temperature and post-fixed with 4% paraformaldehyde. Endogenous biotin was blocked with biotin blocking system (Molecular Probes, Eugene, OR) as per manufacturer’s instructions. Sections were acetylated with 0.1 M triethanolamine with 0.25% acetic anhydride, dehydrated, and prehybridized in hybridization buffer (50% formamide/300 mM NaCl/30 mM sodium citrate/200 μg/ml salmon sperm DNA/100 μg/ml dextran sulfate/400 μg/ml yeast tRNA/2% 100× Denhardt’s solution/2% dithiothreitol) 50°C for 1 h. Riboprobes were hybridized to hybridization solution to final concentration 60 ng/ml and incubated overnight at 50°C. Slides were washed sequentially at 50°C in 0.6M NaCl/60 mM sodium citrate for 30 min, 30 mM NaCl/3 mM sodium citrate, and TBS containing 0.05% TWEEN-20 for 5 min (TBST). Sections were then incubated sequentially in TBS containing 0.05% TSA blocking reagent for 30 min at room temperature, 1:150 streptavidin:horseredh peroxidase (HRP) conjugate for 30 min at room temperature, and 1:50 biotinyl tyramide solution for 10 min at room temperature. After washing with TBST sections were incubated with 1:75 streptavidin:Texas Red conjugate or 1:150 streptavidin:fluorescein isothiocyanate (FITC) conjugate in 150 mM NaCl/50 mM Tris pH 8.0 for 30 min at room temperature. Sections were dehydrated, covered with mounting medium (Vector Laboratories, Burlingame, CA) and imaged by fluorescence microscopy. All sections were subjected to identical treatment and image processing.

Generation and Purification of Anti-mProstasin Polyclonal Antibodies

A theoretical model of mProstasin was created with SwissModel by threading the catalytic domain amino acid sequence (Ile45-Arg322) onto mast cell tryptase (GenBank accession #1A0L) crystal structure (www.expasy.ch/swissmod/SWISS-MODEL.html). Two peptides predicted to form antigenic surface loops of prostasin were selected based on our model. Peptides HTVAO11THTHSSYRE and QEOSPDIHLCNHHP, corresponding to amino acids 113-128 and 297–310 of the preproenzyme, were conjugated to keyhole limpet hemocyanin for antibody production. Peptide synthesis, conjugation, and rabbit inoculation were done by Genemed Synthesis. Specific antibodies were purified on an Affigel-15 (Bio-Rad, Richmond, CA) affinity column anhydrously coupled with 10 μmol of each antigenic peptide. Serum was heat-inactivated before loading the column. Antibodies were eluted with 100 mM glycine/150 mM NaCl (pH 2.5), immediately neutralized with 1M Tris-base, supplemented with 0.25 mg/ml bovine serum albumin, and stored at 4°C.

Immunoblotting and Immunohistochemistry

Primary rat alveolar type II epithelial cells (AT II) were isolated to 95% purity by elastase digestion as previously described and were the generous gift of Drs. Kanam Anabu and Michael Mathay (24). Total RNA was isolated using RNAeasy spin columns (Qiagen) for RT-PCR using rat prostasin (NM_138836) primers 5'-GGTTCTGGTGCTCCTCCTG-3' and 5'-ATACGTTTCCA GCGTCTAG-3'. 2 × 10⁶ freshly isolated AT II were labeled with cell-impermeant sulfo-NHS-biotin (Pierce, Rockford, IL), homogenized in 150 mM NaCl/20 mM Tris pH 8.0/1% Triton X-100, and incubated with 20 μl strepavidin-agarose beads at 4°C for 2 h. Beads were washed with 500 mM NaCl/20 mM Tris pH 8.0, resuspended in SDS-PAGE sample buffer containing dithiothreitol, separated by 12.5% SDS-PAGE, and electroblotted to polyvinylindenedifluoride (PVDF) membranes (NEN Life Sciences, Boston, MA). Membranes were prehybridized in TBS-T (150 mM NaCl, 50 mM Tris, 0.1% Tween-20, pH 7.4) for 30–45 min at room temperature and incubated for 1 h with affinity-purified anti-mProstasin immunoglobulin in the same buffer. Blots were washed with TBS-T and then incubated for 1 h at room temperature with goat anti-rabbit IgG-HRP (Cell Signaling Technology). Bound antibody was detected by enhanced chemiluminescence using Luminol reagents (Cell Signaling Technology) and blots were exposed on Hyperfilm ECL (Amersham).

Immunohistochemical studies of prostasin were done in mouse lung using affinity-purified anti-mProstasin antibody and control preimmune serum. Tissues were fixed in 4% paraformaldehyde for 24 h, embedded in paraffin, and sectioned. Sections were deparaffinized with xylene, rehydrated, and boiled for 10 min in Antigen Unmasking Solution (Vector Laboratories). Endogenous peroxidase activity was quenched with methanol/3% H₂O₂. Avidin, biotin, and nonspecific protein binding sites were blocked with avidin-biotin blocking solutions and CAS block as directed by the manufacturer (Zymed Laboratories, South San Francisco, CA). Sections were incubated with affinity-purified immune or control serum diluted 1:50 in CAS block for 18 h at 4°C in a humidified chamber. Sections were then washed with phosphate-buffered saline containing 0.01% TWEEN-20, incubated with 1:200 biotinylated goat anti-rabbit IgG (Vector) for 30 min at room temperature, then incubated with avidin-HRP conjugate for 30 min at room temperature, washed, and developed with 3,3’-diaminobenzidine chromogen solution (Zymed). Sections were dehydrated, covered with Histomount (Zymed), coverslipped, and digitally imaged. All sections were subjected to identical treatment and image processing.

prss8 Promoter Analysis

Screening the mouse high throughput genomic sequence database yielded sequences nearly identical to prss8 on contig AC093175.6. A 2,961-bp sequence spanning 5’ flank and part of the 5’ UTR was amplified from BAC26 using forward and reverse PCR primers 5'-TTGGGCATGGGCGATCTCTTGA and 5'-AGATCGTCTCGCTC AAAACCA, respectively. This amplifier was subcloned into pGL3-Basic firefly luciferase reporter vector. Fragments of this 5’ flanking sequence were generated by PCR and restriction digestion.
and cloned into pGL3-Basic. Plasmid DNA was purified for use in transfection experiments using Qiagen Hi Speed kit as per manufacturer's instructions, quantitated by spectrophotometry, and confirmed by sequencing. Phylogenetic footprinting of mouse and human prostasin 5' flanking sequences was done with Bayes Block Aligner (http://baysweb.wadsworth.org/cgi-bin/bayes_align12.pl) (25). Consensus transcription factor binding sites in mouse and human prostasin genes were identified using TESS to search the TRANSFAC4.0 database (22). TESS queries were done with the combined string and matrix search algorithm using default parameters.

Promoter function was studied in mouse M-1 kidney cells, which constitutively express prostasin as reported (26) and confirmed by PCR in our cells (data not shown). Reporter gene activity was measured using the Promega Dual Luciferase Assay as directed by the manufacturer. Cells were co-transfected with pGL3 prss8 promoter constructs and a control Renilla luciferase vector, pRL-TK, to permit normalization of data for variation in transfection efficiency and cell viability between wells. M-1 cells were maintained in 50:50 F12K:DMEM supplemented with 100 nM dexamethasone, 100 μg/ml penicillin, 100 μg/ml streptomycin, and 5% fetal calf serum at 37°C with 5% CO2. Before transfection cells were detached with trypsin and plated overnight at 100,000–150,000 cells/well in 24-well plastic culture plates. Cells were co-transfected with 2 μg pGL3 construct and 0.5 μg pRL-TK plasmid DNA using 2 μl/μg DNA of Lipofectamine 2000. Luciferase activity was assayed 48 h after transfection. Cell lysates were transferred to white 96-well luminescence plates after 15–30 min incubation and firefly luciferase activity was measured. Subsequently, 100 μl Stop-and-Glo reagent was added to each well to quench firefly luciferase activity and Renilla luciferase activity was normalized to ground luminescence from untransfected cells was subtracted. Firefly luciferase activity was measured using the Promega Dual Luciferase Assay as directed by the manufacturer. Cells were co-transfected with pGL3 prss8 promoter constructs and a control Renilla luciferase vector, pRL-TK, to permit normalization of data for variation in transfection efficiency and cell viability between wells. M-1 cells were maintained in 50:50 F12K:DMEM supplemented with 100 nM dexamethasone, 100 μg/ml penicillin, 100 μg/ml streptomycin, and 5% fetal calf serum at 37°C with 5% CO2. Before transfection cells were detached with trypsin and plated overnight at 100,000–150,000 cells/well in 24-well plastic culture plates. Cells were co-transfected with 2 μg pGL3 construct and 0.5 μg pRL-TK plasmid DNA using 2 μl/μg DNA of Lipofectamine 2000. Luciferase activity was assayed 48 h after transfection. Cell lysates were transferred to white 96-well luminescence plates after 15–30 min incubation and firefly luciferase activity was measured. Subsequently, 100 μl Stop-and-Glo reagent was added to each well to quench firefly luciferase substrate and provide Renilla luciferase substrate, incubated for 15–30 min at room temperature, and Renilla luciferase activity was measured.

Five independent experiments were done in triplicate and background luminescence from untransfected cells was subtracted. Firefly luciferase activity was normalized to Renilla luciferase activity in each sample by computing the ratio of firefly luciferase-to-Renilla luciferase activity. Data are presented as percent maximal activity. Data were compared by one-way ANOVA using SPSS 10.0 (SPSS Inc., Chicago, IL) and the Tukey test was used for post hoc analysis.

Results
Genomic Cloning and Chromosomal Localization of mProstasin Gene
Overlapping mouse ESTs AI789035, AA139498, AI646242, BE200123, and AI648945 were identified by BLASTn search for sequences homologous to human prostasin. Translation of the deduced mProstasin cDNA yielded a 339–amino acid sequence 76% identical to hProstasin. Screening a 129Sv/J mouse genomic DNA BAC library by PCR using primers to mProstasin cDNA yielded two BAC clones spanning the deduced coding region of mProstasin gene (data not shown). The coding sequence of the mProstasin gene spans ~4.2 kb and is divided into six exons (Figure 1) (GenBank accession no. AF378086). The open reading frame encodes 339 amino acids which are identical to those deduced from mProstasin cDNA and >99% (337/339 amino acids) identical to mCAP1 (AF188613). All intron splice junctions conform to the 5'-GT...AG-3' consensus sequence for RNA splicing (Figure 1) (27). Placement and phase of introns are identical to those of human prostasin. Consistent with the genomic organization of serine proproteases, the His-Asp-Ser residues comprising the catalytic triad are encoded by different exons (8). Like human prostasin, mProstasin contains a short second exon encoding six amino acids spanning the predicted signal peptide cleavage site at Ala3 (4).

Two sequence-tagged sites (GenBank accession nos. C79772 and AI313909) are present in exon 6 at bp 4653–4870 and 4849–4954, respectively. These markers localize prss8 to mouse chromosome 7 based on the Whitehead Institute radiation hybrid map (www.ncbi.nlm.nih.gov/genome/sts) (28). This region of mouse chromosome 7 is syntenic to human chromosome 16p11, where the human prostasin gene resides (4). Matrix alignment of mouse and human prostasin genes reveals several highly conserved domains in the 3' UTR and flank (Figure 2). BLASTn search for sequences related to these domains reveals that the longest (96-bp) conserved segment of the mouse prostasin gene 3' flank is 99% identical to the reverse complement of a cDNA encoding the 3' coding sequence and UTR of mouse Myst1 (NM_026370). Myst1 and its putative human ortholog MOF (AF260665) encode a MYST-family histone acetyltransferase (29). Alignment of mProstasin gene and Myst1 using BLAST2 reveals 99% identity of bp 5028–5207 of prss8 (AF378086) with the reverse complement of bp...
Full-Length mProstasin cDNA Cloning and Identification of Multiple Transcription Initiation Sites

Full-length mProstasin cDNA was cloned by RACE. 3' RACE from lung cDNA generated a 1753-bp mProstasin amplifier crossing all exon–intron junctions in the coding sequence (GenBank accession no. AF378085). The 5' cDNA end and transcriptional start sites of mProstasin were determined by RNA-ligase–mediated 5' RACE from mouse kidney polyA RNA using nested gene-specific primers in exons 4 and 3. Two distinct regions of transcription initiation were identified with transcripts originating at four positions: bp −150, −149, −221, and −215 (numbered relative to the start codon) (Figure 2). These regions conform to requirements for transcription initiation in TATA-less promoters (30). Alignment of mProstasin transcripts with genomic sequence reveals an alternatively spliced 143-bp intron within the 5' UTR of transcripts initiating at positions −221 and −215 (Figures 1 and 2). This intron spans bp −173 to −32, including the transcription initiation sites located at bp −150 and −149, and the splice junctions conform to minimum requirements for RNA splicing (27). Recently deposited mProstasin cDNA sequences, NM_133351 and BC003851, initiate near the most 5' initiator and transcribe through the 5' intron we identified. Thus, at least three 5' UTR variants of mProstasin are transcribed from two distinct transcription initiators and alternative splicing of an intron in the 5' UTR. The amino acid sequence deduced from the prss8 gene is identical to that predicted from the cDNA (AF378085) and > 99% identical (337 of 339 amino acids) to mCAP1 (AF188613). The human prostasin gene similarly lacks a TATA-box and begins transcription from an initiator element (4).

Tissue-Specific and Developmental Regulation of mProstasin Expression

To expand data regarding tissue and developmental expression of mouse prostasin, a multiple tissue mRNA dot blot was probed with radiolabeled 735-bp mProstasin cDNA probe (Figure 3). mProstasin transcription is highly regulated, with most abundant expression in kidney and pan-
creas, moderate expression in lung, submaxillary gland, prostate, thyroid, and epididymis, and lower expression in eye, liver, uterus, and smooth muscle. Furthermore, mProstasin expression increases in whole embryo throughout development, becoming weakly detectable by RNA dot-blot at Embryonic Day 15 and increasing thereafter. These data indicate that expression of the mouse prostasin gene is regulated in a tissue- and developmental stage-specific manner.

Specific cell types expressing prostasin in the lung and kidney, where ENaC is critical for normal physiologic function, were identified by in situ hybridization and immunohistochemistry. In the lung, prostasin transcripts are detected in large airway epithelial cells (Figure 4). Staining is also seen in the peripheral lung in a punctate distribution suggestive of expression in alveolar type II cells (AT II); however, these data cannot exclude the possibility that these transcripts are in alveolar macrophages, or subsets of alveolar type I cells or endothelial cells. This distribution is similar to that reported for prostasin in human airway and lung by in situ hybridization (13). In the kidney in situ hybridization shows that mouse prostasin is transcribed primarily in epithelial cells of renal cortical tubules. This pattern of gene expression is consistent with data obtained by laser capture microdissection and RT-PCR in the mouse nephron (12).

Immunoblotting and immunohistochemistry were done to further define prostasin expression in specific cell types and to confirm in situ hybridization findings. We determined by PCR and DNA sequencing that primary rat AT II cells express prostasin (data not shown). Cell surface biotinylation and immunoblotting reveals that prostasin is expressed as a ~45- to 50-kD protein on the surface of AT II cells (Figure 5). Immunohistochemical staining of adult mouse lung demonstrates specific staining in large airway epithelial cells and in corner cells of the alveoli, which are likely to be AT II cells (Figure 5). Immunohistochemical studies were not successful in the kidney due to excessive background staining.

Cloning and Analysis of prss8 5’ and 3’ Flanking Regions and Promoter

We directly sequenced 684-bp of prss8 5’ flanking sequence from BAC26. By using BLASTn to search the mouse high throughput genomic sequence database for sequences homologous to prss8 5’ flanking and UTR, we identified additional 5’ flanking by assembling highly homologous sequences on mouse chromosome 7 contig AC093175.6. We confirmed this deduced genomic sequence by PCR from BAC26 using forward primer MPRS5’-F2 (5’ TGGCTTTGAGCCATCT CTTACG), annealing to 5’ flanking sequence and reverse primer MPRS5’-R1 (5’ TGGCTTTGACAGGATCTG), annealing to the 5’ UTR. This amplimer spans all transcriptional start sites identified by RLM-RACE (Figure 2). These primers amplified a 2691-bp fragment including 2529-bp of 5’ flank (numbering relative to start codon) (Genbank accession no. AF378086). We analyzed the prss8 5’ flank for functional regulatory elements by conventional promoter reporter assays and by phylogenetic footprinting and database searches for conserved, consensus transcription factor binding sites.

Mouse prostasin has a CAAT- and GC-boxes in close proximity to the initiator elements, as is typical of other TATA-less genes (Figure 2). Matrix alignment of mouse and human prostasin 5’ flanks reveals that the 300-bp

Figure 4. In situ hybridization for mouse prostasin in normal lung and kidney. Sections hybridized with antisense riboprobe to mouse prostasin are shown in the left column, and sense riboprobe controls are shown in the right column. A and B show that mouse prostasin is highly expressed in epithelial cells lining large airways. C and D show that prostasin is also expressed in a punctate pattern in lung alveoli, suggestive of expression in alveolar type II cells. E and F reveal expression of prostasin in epithelial cells of kidney cortical tubules.

Figure 5. Immunostaining for prostasin in primary alveolar type II (AT II) epithelial cells and normal mouse lung. Left panel: Cell-surface proteins of rat primary AT II cells were labeled with N-hydroxysuccinimidobiotin and affinity purified with streptavidin-agarose beads; control cells were not biotinylated. Immunoblots were probed with affinity-purified, rabbit anti-mouse prostasin antibodies and detected by enhanced chemiluminescence. Right panel: Immunohistochemical staining of prostasin in mouse lung. Paraformaldehyde-fixed, paraffin-embedded sections were incubated with polyclonal, affinity-purified anti-mProstasin antibodies or preimmune serum; bound antibody was detected using avidin-biotin amplification and chromogenic detection with HRP and 3’,3’-diaminobenzidine substrate. All sections were processed and imaged identically. Intense staining is present in epithelial cells of proximal airway and in corner cells of alveoli that most likely represent AT II cells.
AC-rich region of human prostasin, which is interposed between the initiator and a putative GC-box, is not present in the mouse gene (data not shown). No CpG-islands are present in the 2.5-kb region adjacent to the mProstasin gene.

To identify basal elements regulating transcription of mProstasin, we generated promoter–reporter constructs in pGL3-Basic and measured promoter activity using a dual luciferase assay in transiently transfected M-1 cells. Data are normalized to Renilla luciferase activity to control for transfection efficiency, cell number and viability; data are expressed as mean ± SD percent maximal activity. Promoter activity was compared by ANOVA and Tukey’s test was used for multiple comparisons. *P < 0.05 compared with all other constructs; **activity of XbaR1 and F4R2 are not significantly different from each other but differ from all other constructs (P < 0.05).

our RLM-RACE experiments did not identify transcripts initiating in this region in kidney. F3R1 drives expression ∼ 2-fold higher than the core promoter contained in F4R1. Comparison of F3R1, KpnIR1, and XbaIR1 constructs indicates that this transcriptional activation is due primarily to elements located in the 1461-bp sequence between KpnI and XbaI sites, with a small but statistically significant effect attributable to the 105-bp fragment between F3 and KpnI. A construct containing further 5′ sequence, F2R1 (−2750 to −59) has less activity than F3R1 and KpnR1, indicating the presence of repressor(s) in this region. XbaR1 has significantly less activity than F4R1, indicating that a functional repressor(s) lay in this 160-bp sequence.

Computational analysis of the 5′ flank of prss8 was done using Bayes Block Aligner for phylogenetic footprinting and TESS for identification of consensus transcription factor binding sites. Phylogenetic fingerprinting identifies short (≥ 6 bp) conserved sequences in the gene regulatory regions (25). Three very highly conserved regions with > 80% identity and three regions of > 60% identity were identified by comparing the 2529-bp sequence upstream of the mouse prostasin 5′ UTR sequence index relative to the start codon; y-axis is probability of alignment. Three regions of > 80% and three regions of 60–70% identity are present and contain several consensus transcription factor binding sites (see RESULTS). All of these conserved regions lay between F3 and XbaI sites, a region with strong transcriptional activity, and five are clustered within a 313-bp sequence.
between bp −2063 and −1958 (F3 and KpnI), the best statistical matches are a Kruppel-like factor 4 (KLF-4) binding site at bp −1985 to −1972 and an interferon (IFN) regulatory factor-1 (IRF-1) binding site at bp −1975 to −1966, though multiple other potential sites are present. A CAAT/enhancer binding protein-β (C/EBP-β) site at bp −1966 to −1961 lies adjacent to the IRF-1 site.

The BLASTp algorithm was used to search GenBank for the serine proteases most highly related to the catalytic domain of mouse prostasin (amino acids Ile45-Arg322), and a phylogenetic tree of these sequences was constructed from multiple alignment of these sequences generated by MacVector (Figure 8). Phylogenetic tree analysis reveals that prostasin, as suggested from recent functional studies of ENaC activation (12), is indeed the ortholog of Xenopus CAP1, and that these genes form a unique branch of the prostasin/tryptase-ε gene family. This analysis reveals strikingly high conservation between rodent and human orthologs in this gene family.

**Discussion**

In these studies we describe the initial cloning and characterization of the mouse prostasin gene, identify multiple transcription start sites and a 5’ UTR splice variant, and identify critical regulatory regions in the 5’ flank that control its basal expression. We also show that the mouse prostasin gene is developmentally regulated and is specifically expressed in airway and alveolar type II epithelial cells in the lung. Mouse and human prostasin genes reside on syntenic regions of chromosomes 7 and 16, respectively, and contain several highly conserved sequences in the 5’ regulatory region. Comparing these findings with data on human prostasin reveals that mouse and human prostasin have highly conserved genomic organization, deduced amino acid sequence, and tissue- and cell-specific pattern of expression (4, 9, 12). Thus mouse and human prostasin are likely to have similar functions and mechanisms of regulation, supporting the rationale for studying function and regulation of prostasin in the lung and kidney in mouse models.

The six-exon genomic organization of prostasin is characterized by a short second exon, similar to tryptase-ε, testisin, pancreasin, tryptase-ε, and DISP (Figure 1) but distinct from other serine proteases (2, 3, 5–7, 32–34). These genes encode trypsin-like serine proteases, which preferentially cleave substrates with a basic amino acid in the P1 position. With the exception of pancreasin and tryptase-ε, these genes encode type I transmembrane proteins with a carboxy-terminal hydrophobic or GPI-anchor signal domain affixing the protease domain extracellularly to the membrane. Pancreasin and tryptase-ε lack the C-terminal hydrophobic domain and are predicted to be secreted proteins. These human genes reside on human chromosome 16p13.3/11.2, and the orthologous murine genes reside on syntenic fragments of chromosomes 7 and 17 (2, 5, 33–35). Based on these strong similarities in gene organization and chromosomal localization, sequence and predicted structural and enzymatic features, these proteases are proposed to belong to a distinct family of trypsin-like serine proteases. These genes probably arose by duplication of an ancestral serine protease gene with the unique, short second exon originating by insertion of a new phase I intron. The extensive sequence conservation between mouse and human orthologs in this gene family and their similar patterns of expression suggest a high likelihood of functional conservation in vivo (Figure 8). Indeed, mouse, rat, and human prostasin/CAP1 have been shown to activate ENaC by an unknown proteolytic mechanism (12, 13, 36).

Little is known about the mechanisms regulating gene expression, activation, membrane anchoring, and secretion of prostasin, or its biologically relevant substrates and inhibitors. In the lung we have localized expression of prostasin mRNA and protein to airway and alveolar type II epithelial cells (Figures 4 and 5). This parallels the pattern of expression in normal human lung reported by Donaldson and coworkers using in situ hybridization (13). These cell types express ENaC and are critical to maintaining normal epithelial lining fluid in the lung, an ENaC-dependent function, and thereby preserving normal gas exchange and mucociliary clearance (15, 16). In these studies we also demonstrate that prostasin is expressed in normal mouse cortical tubule epithelial cells (Figure 4), consistent with RT-PCR data of prostasin expression in the nephron (12). Narikiyo and colleagues recently reported that transcription and secretion of prostasin is induced by aldosterone in mouse M-1 renal cortical collecting duct cells, resulting in increased transepithelial sodium transport (26). However, other investigators have found that aldosterone does not regulate expression of mCAP1/prostasin in mdkCCD14 or M-1 mouse cortical collecting duct cells (31, 37). Our analysis of 2.6-kb of mouse prostasin promoter-enhancer sequence did not reveal any consensus mineralocorticoid receptor binding sites.

We have initiated studies of the mechanisms regulating
transcription of the prostasin gene by doing promoter analysis in mouse M-1 renal cortical collecting duct cells. Available mouse lung epithelial cell lines MLE-12, MLE-15, and LA-4 do not express prostasin constitutively (data not shown); therefore, we are currently unable to do parallel experiments in a mouse lung epithelial cell line to try identify regulatory regions specific for lung epithelial cell expression. As shown in Results, the core promoter of mouse prostasin lacks a TATA-box. Basal transcription is driven by two consensus initiator elements and a GC-rich domain, located 50–140 bp 5' of the initiator elements, that is likely to bind the ubiquitous factor SP1 (Figures 2 and 6). Multiple transcription start sites were identified by RLM-RACE, typical of TATA-less promoters due to the absence of a precise localizing signal for transcription initiation (38). The core promoter architecture is similar to that proposed for human prostasin, although the human gene has a 300-bp AC-rich sequence interposed between the initiator and the GC-boxes (4). Experimental data on human prostasin promoter activity is not currently available to address the effect of this AC-repeat on promoter function. The mouse prostasin gene, unlike the human gene, also has a CAAT-box in the promoter proximal region 251-bp 5' of the first transcription start site. A reporter construct containing this site (XbaIR1) unexpectedly had less transcriptional activity than the construct containing only the GC-boxes and initiators (F4R1). This finding may be explained by the presence of a repressor in the additional 160-bp sequence.

We found strong enhancer activity in a 1461-bp region between KpnI and XbaI sites and an adjacent 105-bp region between F3 and the KpnI site that cumulatively increase expression ~2-fold over the core promoter activity (Figure 6). Cloning the mouse prostasin gene enabled us to use phylogenetic footprinting to compare the 2.5-kb mouse prostasin 5' flank and 1.7-kb human prostasin 5' flank. With this analysis we identified a cluster of short, highly conserved sequences (Figure 2). Although the functions of conserved sequences (Figure 2). Although the functions of consensus transcription factor binding sites, including several sites for the ubiquitous factor SP1 and for tissue-specific regulators involved in epithelial cell differentiation such as forkhead factors HNF-3 and HNF-5 and homeodomain factor Cdx-1 (39, 40). Several putative transcription factor binding sites are also found in the 105-bp enhancer region between F3 and KpnI, the statistically most significant being sites for Krueppel-like factor 4 (KLF4) and an IFN-γ-regulated factor, IRF-1. KLF4, a member of the SPI/Krueppel-like factor family of zinc finger proteins, is an important factor in terminal differentiation of epithelial cells as well as in cell-cycle regulation in carcinogenesis (41, 42). Expression of KLF4 increases progressively during gestation beginning at mouse Embryonic Day 13 (42), a pattern that parallels developmental expression of mouse prostasin (Figure 3), suggesting that KLF4 may regulate expression of prostasin during maturation. In addition to these basal regulatory factors, the conserved sequences identified by phylogenetic footprinting also contain consensus sites for NF-κB, AP-1, and the glucocorticoid receptor, which coordinate transcriptional programs in inflammatory states in epithelial cells (43, 44).

Although not identified by phylogenetic footprinting, both mouse and human prostasin also have an AP-2 site in the promoter proximal region (data not shown) (4). Tumor necrosis factor-α and activators of protein kinase C, which activate NF-κB, AP-1, and AP-2, are known to modulate ENaC activity in lung alveolar epithelium (45, 46). IFN-γ inhibits amiloride-sensitive current and fluid transport in cultured bronchial epithelial cells (47). Further studies are necessary to determine whether the effects of tumor necrosis factor-α and IFN-γ on ENaC activity in the lung may be due in part to transcriptional regulation of prostasin. Recent reports that prostasin is upregulated in airway submucosal glands of patients with cystic fibrosis and in the kidney of patients with primary hyperaldosteronism suggest that regulation of the prostasin gene occurs in vivo and is linked to physiologically significant perturbations of epithelial ion transport (13, 26). These findings underscore the importance of further identifying the mechanisms regulating prostasin gene expression.

We identified multiple 5'-UTR variants of mouse prostasin that are generated by use of multiple transcription start sites and by alternative splicing of a 5'-UTR intron (Figure 2). Although 5'-UTR introns have not yet been described for the other genes in this family, human α- and β-tryptases genes do contain a 5'-UTR intron of unknown functional significance (1). A computational model of the mouse transcriptome, based on genomic and RIKEN cDNA databases, suggests that ~30% of genes have splice variants that most frequently occur in the 5' region of a gene (48). Variations in 5' UTR may regulate gene expression in specific cell types by altering translational efficiency and mRNA stability. An archetype for studying this phenomenon is the human growth hormone receptor gene, for which at least nine 5' UTR variants with specific expression patterns have been described (49). Whether these splice variants of prostasin are involved in translational regulation and are cell-specific remains to be determined.

Alignment of the 3'-UTR and flank of the mouse and human prostasin genes reveals several short, highly conserved sequences (Figure 2). Although the functions of these domains are not known, sequences in the 3'-UTR can regulate mRNA stability and cytoplasmic localization (50). Our data demonstrates that the 3' end of the mouse prostasin gene overlaps with the mouse Myst1 histone acetyltransferase gene, which is oriented in the opposite direction on chromosome 7. This relationship is similar between the human prostasin and MOF histone acetyltransferase gene, which are oriented tail-to-tail on chromosome 16p11.2 but do not overlap in the 3'-UTRs (29). Thus, mouse prostasin is located on a region of chromosome 7 syntenic to the human prostasin locus on 16p11.2. The apposition of these genes suggests that inactivating prss8 may also affect Myst1 expression, and must be considered in designing gene inactivation experiments.

The tissue and developmental expression pattern of mouse prostasin is very similar to human prostasin (9), being highly transcribed in ENaC-expressing tissues such as kidney, salivary gland, and lung. The progressive increase in mouse prostasin gene expression during development suggests that it has functions specific to differentiated cells.
and also follows the developmental pattern of ENaC gene expression (51). Prostasin is also expressed in tissues that do not express functional ENaC such as pancreas, prostate, and smooth muscle. Members of this gene family are proposed tumor suppressor genes (11, 20), indicating that prostasin may have important functions in addition to activation of ENaC. In summary, in these studies we describe the mouse prostasin gene locus, expand data regarding its pattern of expression and begin to characterize its regulatory regions. The high degree of similarity between mouse and human prss8 genes and their specific patterns of expression suggest that the mouse will be a useful model for studying prostasin functions and regulation. Further studies are necessary to define specific mechanisms that regulate transcription of the prostasin gene and to provide insights into its functions in normal physiology and in disease states.

Acknowledgments: This study was supported by American Lung Association RT-031-N and NIH K08-HL67920 (G.M.V.), NIH HL-24136 and a Research and Development Project of the Cystic Fibrosis Foundation (G.H.C.), and Cystic Fibrosis Research, Inc. (Z.Y.T.).

References


