The Pem Homeobox Gene

ANDROGEN-DEPENDENT AND -INDEPENDENT PROMOTERS AND TISSUE-SPECIFIC ALTERNATIVE RNA SPlicing*

Sourindra Maiti, Jessica Doskowiak, Shulin Li, Ron P. Nhim, J. Suzanne Lindsey, and Miles F. Wilkinson

From the Department of Immunology, University of Texas, M.D. Anderson Cancer Center, Houston, Texas 77030 and the Molecular Microbiology and Immunology Department, Oregon Health Sciences University, Portland, Oregon 97201

The Pem gene encodes an atypical homeodomain protein, distantly related to Prd/Pax family members, that we demonstrate is regulated in a complex transcriptional and post-transcriptional manner. We show that the rat Pem genomic structure includes three 5'-untranslated (5'-UT) exons and four coding exons, three of which encode the homeodomain. Several alternatively spliced transcripts were identified, including one that skips an internal coding exon, enabling this mRNA to express a novel form of the Pem protein. Other alternatively spliced mRNAs were characterized that possess different 5'-UT regions, including a muscle-specific transcript. The different 5'-UT termini present in Pem transcripts conferred different levels of translatability in vitro. Two promoters containing multiple transcription initiation sites were identified: a distal promoter (Pd) in the first 5'-UT exon and a proximal promoter (Ps) located in the "intron" upstream of the first coding exon. The Ps was active in placenta, ovary, tumor cell lines, and to a lesser extent in skeletal muscle. In contrast, transcripts from the Ps were only detectable in testis and epididymis and were only expressed in epididymis in the presence of testosterone. To our knowledge no transcription factors have previously been identified that exhibit androgen-dependent expression in the epididymis.

Androgens are of paramount importance to spermatogenesis in the testis and sperm maturation in the epididymis. Testosterone alone maintains spermatogenesis in gonadotropin-deficient animals, including hypophysectomized rats and mutant hypogonadal animals, including hypophysectomized rats and mutant hypogonadal mice (1–3). Evidence suggests that testosterone drives spermatogenesis by acting on Sertoli cells and peritubular cells, both of which express androgen receptors (4, 5). Sertoli cells perform numerous functions critical for spermatogenesis by virtue of their intimate contact with differentiating germ cells within the seminiferous tubule. Androgens have been known to be critical for epididymal function since early in this century. They regulate the proliferation and differentiation of somatic cells in the epididymis and control the micro-environment of the maturing spermatozoa by regulating the synthesis of adhesion proteins in the epididymis, the secretion of proteins into the luminal fluid that are in contact with the spermatozoa, and the transport of ions and small organic molecules across the epididymal epithelium (6, 7). Secreted proteins under androgen control in the epididymis include steroid metabolizing enzymes, polyamine synthesis enzymes, detoxification enzymes, oxidation reduction enzymes, hydrolases, and proteases (6–8).

The transcription factors that orchestrate androgen-dependent events in the testis and epididymis have not been identified, although the androgen receptor clearly plays a major role in such responses. Homeobox transcription factors are candidates to regulate spermatogenesis and sperm maturation, since they are known to regulate many other developmental events. The distinguishing feature of homeobox proteins is a conserved DNA-binding motif 60 amino acids in length, termed a homeodomain. The homeodomain is comprised of three α-helices; sequence specificity is conferred by key residues in the third helix that direct binding to base contacts in the major groove of DNA. The best understood homeobox proteins are those encoded by the hox/hom, prd/pax, and POU gene families (9). Studies in Drosophila melanogaster, Xenopus laevis, and mice have shown that members of these classical homeobox gene families are required for discrete events during development. For example, studies in null mutant mice have demonstrated that the Pax-6 gene activates a regulatory cascade necessary for eye development (10), the Oct-2 POU homebox gene promotes late stages of B-cell maturation (11), and Hox genes specify axial identity during embryogenesis (12).

Homeobox transcription factor genes have been shown to be expressed in the male reproductive system, but none have been shown to be androgen-regulated. Many of these homeobox genes are expressed in germ cells of the testis. For example, the POU homeobox gene sperm-1 is expressed transiently prior to meiosis in germ cells (13), and Hoxa-4 is expressed specifically in postmeiotic germ cells (14, 15). Hoxb-4 is expressed by both germ cells and somatic cells in the testis, while Hoxd-4 is expressed by Leydig cells (15). Both Hoxb-4 and Hoxd-4 are expressed by other adult organs besides testis (16, 17). Little is known about the expression pattern of transcription factors in the epididymis. To our knowledge the only transcription factor genes identified as expressed in the epididymis are the homeobox gene Pax-2 (18) and the ETS-like transcription factor PEA3 (19).

In a search for developmentally regulated genes, we used the subtraction hybridization technique to isolate several cDNAs corresponding to novel genes (20), including the homeobox gene, Pem (21, 22). The Pem homeodomain shares modest sequence identity with prd/pax homeodomains (22, 23), but its

*This work was supported by National Institutes of Health Grants HD 27233 and T32 EY07123. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore by hereby marked "advertisement" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

The nucleotide sequence(s) reported in this paper has been submitted to the GenBank™/EBI Data Bank with accession number(s) U52034.

§ To whom correspondence should be addressed: U.T. M.D. Anderson Cancer Center, Dept. of Immunology, Box 180, 1515 Holcombe Blvd., Houston, TX 77030. Tel.: 713-794-3526; Fax: 713-745-0846; E-mail: Miles_Wilkinson@sgm.mda.uth.tmc.edu.

17536
primary amino acid sequence is sufficiently unique to warrant classification in a different subfamily. The Pem gene is expressed in a unique pattern during embryogenesis. Tissues that contribute to the extraembryonic compartment express mouse Pem; the gene remains highly expressed in the placenta and yolk sac until term (21, 24). The in vivo expression pattern of Pem in endodermal tissue is mimicked in pluripotent stem cell lines that differentiate in vitro; F9 embryonal carcinoma stem cells induced to differentiate into either the visceral or endodermal lineage up-regulate Pem mRNA expression (22) and accumulate Pem protein specifically in the outer layer of cells that possess characteristics of extraembryonic endoderm (24, 25). Pem gene expression is also dramatically up-regulated in normal diploid embryonic stem cells induced to differentiate in vitro (22), although it is not known which specific differentiated cell types activate Pem gene expression.

In this communication, we report that Pem gene expression is not restricted to embryogenesis. We show that in prepubertal and adult rats, the Pem gene is specifically transcribed in both male and female reproductive tissue and to a lesser extent in skeletal muscle. Transcriptional analysis revealed that Pem transcripts are derived from two promoters and undergo complex alternative splicing events that are regulated in a tissue-specific manner. The alternative splicing events alter both the 5'-UTR and coding regions of Pem mRNA. We demonstrate androgen-dependent expression of Pem transcripts from the promoter used exclusively by male reproductive tissue. The complex and androgen-dependent pattern of Pem expression by alternative promoter usage and alternative splicing has important implications for the possible role of the Pem homeobox gene in development.

MATERIALS AND METHODS

Isolation and Characterization of Rat Pem CDNA and Genomic Clones—To obtain rat Pem CDNA clones, a mouse Pem cDNA probe was used to screen 5 × 10^6 plaques (26) from a Rat-1 fibroblast cDNA library in a ZAP (kindly provided by David Pribnow). Southern blot analysis revealed that 31 independent Rat-1 cDNA clones contained inserts that strongly hybridized with the mouse Pem probe. Many of the cDNA clones were sequenced at their 5' termini to determine the approximate site of transcriptional initiation (see Results). DNA sequence analysis was performed by the dideoxy method according to the manufacturer's instructions (U.S. Biochemical Corp.).

Three genomic libraries and a P1 rat genomic library were screened to obtain rat Pem genomic clones that correspond to the rat Pem CDNAs that we had isolated. We were not able to isolate any clones for a functionally pseudogene, and instead sequenced many independent subcloned PCR products from other transgenic rat lines and sequences were used for long PCR amplification. Sequence analysis of the resulting fragment (clone 2) revealed that it overlapped clone 1 by 57 nucleotides, as expected. Together, clones 1 and 2 comprised the entire rat Pem coding region and all intervening introns.

The genomic organization of rat Pem was confirmed by analysis of two other subcloned PCR products (clones 3 and 4) derived from other primer sets. Clone 3 was generated using an oligonucleotide that included the start codon ATG (oligo E) in combination with oligo B. This 4.5-kilobase PCR product was subcloned and shown to possess restriction sites predicted from sequence analysis of clones 1 and 2. Clone 4 was a PCR product that was generated using an oligo (oligo F) that bound fortuitously to a region in the the third exon. Sequence analysis of this 0.5-kilobase PCR product showed that it contained the sequences predicted from clone 2.

To obtain sequences upstream of exon 2, we employed the method of Siebert et al. (27). The "Rat Promoter Finder Kit" used for this upstream walk (a generous gift from Clontech) was used according to the manufacturer's instructions, except that the rat TTH DNA polymerase XL (from Perkin-Elmer) was used in place of the enzyme provided by this kit. The primary PCR was performed with primer AP1 and oligo F for 7 cycles under the following cycle parameters: 94°C for 5 s and 60°C for 5 min. The secondary PCR was done with primer AP2 and oligo G for 40 cycles under the following cycle parameters: 94°C for 5 s and 63°C for 40 cycles. Sequence analysis of the subcloned PCR product (clone 5) showed it contained an intron (IVS1) and exon sequences (exons 1 and 2) identical to known rat Pem cDNA sequences.

RNA Isolation and RNase Protection Analysis—Total RNA from tissues was prepared as described previously by either guanidinium isothiocyanate lysis and centrifugation over a cesium chloride cushion (28) or single-step acid guanidinium thiocyanate-phenol extraction (29). For RNA protection analysis, we prepared (32P)UTP-labeled RNA probes with T7, T3, or SP6 RNA polymerase. The probes used for RNase protection analysis contained the following exon and intron sequences: probe A, IVS2B (142 nt); E3 (82 nt); IVS3 (167 nt); E4 (140 nt) = 0.350 nt; probe B, E1 (30 nt); IVS2-E3-IVS3 (complete); E4 (140 nt) = 0.230 nt; probe C, IVS2B (125 nt); E3 (82 nt); E4 (140 nt) = 0.350 nt; probe D, E1 (30 nt); E2 (61 nt); E3 (82 nt); E4 (140 nt) = 0.310 nt; probe E, E1 (49 nt); E2 (61 nt); E3 (30 nt) = 0.140 nt; probe F, E1 (29 nt); E2 (61 nt); E3 (30 nt) = 0.120 nt; probe G, E1 (40 nt); M exon (45 nt); E2 (61 nt); E3 (82 nt); E4 (140 nt) = 0.370 nt; probe H, E1 (18 nt); IVS1 (3-terminal 36 nt); E2 (61 nt); E3 (72 nt) = 0.189 nt; probe I, E1 (29 nt); IVS2B (30 nt); endonuclease from the IVS2B (complete border) = 0.86 nt; probe J, E4 (69 nt); E5 (46 nt); E6 (220 nt) = 0.335 nt.

In some experiments, a glyceraldehyde-3-phosphate dehydrogenase probe was included in the annealing reaction as a positive control. For probe synthesis, we used the in vitro transcription protocol as described (30). Probes were purified in a 8 × urea, 6% polyacrylamide denaturing gel. After exposure to film, the appropriately sized bands were excised from the gel and placed in individual Eppendorf tubes. The tubes were washed with an RNase-free pestle in 100 μl of diethylpyrocarbonate-treated water. To each sample, 600 μl of proteinase K-containing solution (0.3 μ NaCl, 0.5% SDS, 10 mM Tris (pH 7.5), 200 μg/ml proteinase K, and 20 μg/ml RNAse) was added, vortexed, and incubated at 37°C for 15 min. After vortexing and pulse-spin, the suspension was filtered through a 0.45-μm filter (Amicon) by passing 200 μl more proteinase K solution through the filter to increase the recovery of probe. Each sample was then extracted with 200 μl of phenol/chloroform. One microliter was used to determine radioactive counts per minute, and the rest was ethanol-precipitated and stored at −70°C.

RNase protection analyses were performed as described (30), with minor modifications. Briefly, sample RNA or tRNA (negative control) was co-precipitated with the appropriate gel-purified (32P)UTP-labeled probes. The pellet was resuspended in 30 μl of annealing buffer (40 μl PIPES (pH 6.4) 0.4 μ NaCl, 1 mM EDTA, 80% formamide) and allowed to hybridize overnight at 44°C. Unhybridized RNA was digested with RNase A and RNase T1, fianal at 37°C at concentrations of 1-3 μg/ml and 5 μg/ml, respectively, unless otherwise noted. RNAs were then removed by treatment with proteinase K and extraction with phenol/chloroform/soybean alcohol. After ethanol precipitation, the RNA pellet was resuspended in 0.2% formamide loading buffer, denatured at 85°C, electronosed in a 8 × urea, 6% polyacrylamide gel. A set of RNA size markers generated from the Century ladder template (Ambion) was included in all gels.

Primer Extension Analysis—Primer extension was carried out essentially as described by McKnight et al. (31) using total cellular RNA (30 μg) from rat placenta and (32P)-end-labeled primers. The labeled oligo (2 ng) and the RNA mixture were ethanol-precipitated and resuspended in a total volume of 18 μl of water and incubated on ice for 5 min with intermittent vortexing, followed by the addition of 2 μl of 10 × annealing buffer (3 μ NaCl, 0.4 μM Tris (pH 8.0) and 1 mM EDTA), brief vortexing, and incubation at 65°C for 10 min. Following the annealing
reaction, the tubes were transferred to the temperature of the extension reaction (42, 46, or 52°C). To each tube the following was added: 4 μl of 10 mM extension buffer (1 M Tris [pH 8.3], 120 mM MgCl2, 100 mM dithiothreitol), 0.8 μl of 25 mM dNTPs, 14 μl of double-distilled H2O, and 5 units of either avian myeloblastosis virus (for 42°C) or Moloney murine leukemia virus reverse transcriptase (for 46 and 52°C). After a 60-min extension reaction, the template RNA was degraded by incubation with 1 μl of RNase A (stock 10 mg/ml) at 37°C for 1 h. The sample was then precipitated by the addition of 132 μl of stop mixture (2.5 M NH4OAc, 10 mM EDTA) and 500 μl of ethanol. The products were separated by electrophoresis in an 8 M urea, 6% polyacrylamide gel.

Reverse Transcriptase-PCR (RT-PCR) and 5′-Rapid Amplification of cDNA Ends (5′-RACE)—RT-PCR was performed as described (32) using total cellular RNA (1 μg) from adult epididymis and skeletal muscle. 5′-RACE was performed according to the manufacturer’s instructions (Life Technologies, Inc.). In brief, cDNA was generated using a primer complementary to a region within exon 4, the cDNA was c-Dailed with terminal transferase, and then PCR was performed using oligo D and an “anchor primer” complementary to the 3′-UT.

In Vitro Transcription and Translation—Three 147C3-based plasmids were prepared that contained precisely the same rat pem open reading frame preceded by the 5′-UT region in A-, M-, and T-transcripts (Fig. 2). The length of the Pem 5′-UT region in the A-, M-, and T-transcripts was 93 nt, 138 nt, and 109 nt, respectively. The plasmids were linearized with Hin dIII, and RNA was synthesized in vitro according to the manufacturer’s instructions (Promega Corp.). The RNA was translated in vitro using [35S]methionine and reticulocyte lysates in a 25-μl reaction for 1 h at 30°C according to the manufacturer’s instructions (Promega), and the products were analyzed by SDS-polyacrylamide gel electrophoresis.

Animals—Untreated, sham-operated, and hypophysectomized Sprague-Dawley rats were obtained from Charles River Laboratories. Animals were housed in the Oregon Health Sciences University animal care facility and cared for according to approved protocols. Hypophysectomized animals received 5% glucose water ad libitum. Animals were killed by CO2 asphyxiation, and organs were immediately re-
moved, homogenized, and frozen at −70 °C until RNA was extracted. The effectiveness of hypophysectomy was determined by assessing testosterone levels in serum with a standard chromatographic procedure (33). For the testosterone implant experiments, hypophysectomized rats (12 days post-treatment) were anesthetized at one atmosphere isofluorane and the implants (silastic tubing 3 cm long filled with testosterone propionate) were placed subcutaneously along the upper back and neck in collaboration with Dr. John Resko (Oregon Health Sciences University), who has shown that these implants generate a serum concentration of 4 ng/ml testosterone. The levels of testosterone and dihydrotestosterone (DHT) were determined in contralateral epididymides (weighed, homogenized in phosphate-buffered saline, and frozen at −70 °C until analysis). The sham-operated animals had 6 pg of DHT/mg of tissue, hypophysectomized animals had 1 pg of DHT/mg of tissue, and all testosterone-treated animals had 3–4 pg of DHT/mg of tissue (assayed 2–8 days after introduction of the implants). All androgen assays were done in the laboratory of Dr. David Hess at the Oregon Regional Primate Research Center (Beaverton, OR).

RESULTS

The Pem Gene—As described under "Materials and Methods," use of the long PCR method permitted isolation of several overlapping DNA fragments that corresponded to the entire rat Pem gene. Sequence analysis and comparison of these sequences with the known mouse and rat Pem cDNA sequences (Ref. 22 and see below) allowed us to deduce the genomic organization of the Pem gene (Figs. 1 and 2). The exon/intron splice junctions in the Pem gene conform to the consensus sequences for 5′ (CAGGTRAGT) and 3′ (YNYAG) splice sites (the invariant dinucleotides at the termini of the intron consensus sequences are underlined). The homeodomain region of the Pem gene is interrupted by two introns, positioned precisely in the same location as in the D. melanogaster prd class homeobox gene aristless (al) (34). The location of the second
intron interrupting the Pem homeodomain (IVS5) is identical to the location of the intron in the homeodomain region of several other prd/pax class homeobox genes, including the gsc, S8, otx, unc-4, ceh-8, and ceh-10 genes, but is in a different position than the introns in most other known homeobox genes (9). This provides further support that the Pem gene is a distant relative of the Prd/Pax homeobox gene sub-family.

We previously showed that the Pem homeodomain exhibits up to 35% sequence identity with prd/pax family member homeodomains (35).

In the studies described below, we demonstrate the usage of two promoter regions in the Pem gene and show that transcripts derived from these promoters undergo alternative splicing events. Fig. 2 summarizes the results of these studies. Below, we will first provide evidence for the existence of the proximal promoter (Pp) and then show its androgen-dependent regulation and developmental expression pattern in male reproductive tissue. Then we will define and analyze the distal promoter (Pd), which is primarily expressed in ovary and placenta, and to a lesser extent in male reproductive tissue and skeletal muscle. Throughout the analysis, we also make use of two cell lines that transcribe the Pem gene from the Pd: the Rat-1 immortalized fibroblast cell line and the MCA8994 rat hepatoma cell line. We previously showed that immortalized and tumor cell lines from multiple cell lineages express the Pem gene (21).

A Proximal Promoter Active in Epididymis and Testis—Pem transcripts derived from the Pp were first identified from epididymis RNA by the PCR-based approach 5'-RACE. Sequencing of subcloned 5'-RACE products revealed that the 5' termini extended to several sites within the intron upstream of exon 3 (positions -59, -86, -94, -117, and -125 nt in Fig. 1). The different lengths of these termini could be due to multiple transcription start sites within the Pp, or they could have resulted from incomplete cDNA synthesis by reverse transcriptase.

RNase protection analysis was employed to determine whether multiple transcription start sites were present in the Pp. Fig. 3 shows the results with probe A, which contains the Pp region. Epididymis RNA protected four major fragments (bands 1–4) that correspond to transcriptional initiation sites at the following approximate positions relative to the initiator ATG: -126, -109, -75, and -68. Different ribonuclease A and T1 concentrations (over a 4-fold range) did not affect the migration of bands 1–4 (data not shown), so these multiple bands do not represent partial ribonuclease cleavage fragments, and instead correspond to multiple sites of transcriptional initiation. Use of multiple transcriptional initiation sites is typical for mammalian promoters that lack an upstream TATA box, and indeed, no TATA box is present upstream of the initiation sites in the Pp (Fig. 1). In order to determine if any other transcription start sites are used in epididymis further upstream within the intron, we also used a probe that contained all of intron 2 (probe B). No additional bands were obtained with this probe (data not shown), indicating that there are no other transcription start sites within the Pp.

The transcription initiation sites within the Pp, that we defined in epididymis were also used in testis (see below) but were not active in placenta, ovary, skeletal muscle, or two immortalized cell lines that express the Pem gene, Rat-1 and MCA8994. Instead, these tissues and cell lines expressed Pd-derived transcripts, and thus they protected band 5 after an-
nealing with probe A (Fig. 3B and data not shown). Placental RNA not only protected band 5 but also a much less abundant fragment that was slightly larger than band 1. The origin of this band is not known; it does not appear to represent transcription from the Pp region, since another probe that included the Pp region (probe C) failed to detect Pp-derived transcripts in placenta (data not shown).

The Proximal Promoter Is Androgen-dependent—Since the functional competence of the epididymis depends on the presence of androgens (6–8), we tested whether transcripts derived from the Pp depended on testosterone for expression. A probe that included the Pp region (probe C) was used for this analysis (Fig. 4A). Hypophysectomy caused a precipitous drop in the levels of Pp-derived transcripts (Fig. 4B and C, band 1 in lanes labeled HPX), whereas animals that underwent sham treatment maintained Pp-derived transcript expression in the epididymis. Introduction of exogenous testosterone in hypophysectomized animals restored expression of these transcripts (lanes labeled HPX1+T and HPX2+T in Fig. 4C represent two different animals). Testosterone induced expression from the Pp after 2 days of treatment (Fig. 4C); expression was maintained for at least 8 days in other animals tested (data not shown). These mRNAs derived from the Pp were designated “T-transcripts” (Fig. 2), since they are inducible by testosterone.

Developmental Shift from the Distal Promoter to the Testosterone-dependent Promoter—Tests accumulated three major Pem mRNAs represented by bands 1, 2, and 3 in Fig. 5B. Band 1 corresponds to A-transcripts, derived from the Pp, the predominant transcript also expressed by placenta, ovary, and muscle (Fig. 2, and see below). Band 2 represents an mRNA that also appears to be transcribed from the Pp, but its size suggests that it may have undergone an alternative splicing event between exons 1 and 2 (although other explanations are possible). Band 3 represents T-transcripts derived from the Pp.

The ratio of transcripts derived from the Pp and Pd varied at different developmental stages in the testis. At early time points after birth (days 5–33), the predominant transcripts were derived from the Pp (Fig. 5B, bands 1 and 2), T-transcripts (band 3) became more prominent by day 44 and later accumulated to similar levels as Pp-derived transcripts (days 78 and 104). This developmental shift was observed with a Pp probe was confirmed by analyses with a Pp probe (compare the ratio of band 1 with band 2 on day 21 with day 78 (Fig. 5D)). Note that the epididymis expressed much higher levels of T-transcripts than did the testis: even young animals (23 and 30 days old) accumulated high levels of T-transcripts in epididymis (Fig. 5D).

Multiple Transcription Initiation Sites Used by the Distal Promoter—The sites of transcriptional initiation from the Pd were analyzed by three different approaches: 1) sequence analysis of the 5' termini of cDNA clones; 2) RNase protection analysis; and 3) primer extension analysis. The Rat-1 fibroblast cell line was chosen for cDNA library screening since it expresses high levels of Pem transcripts, as do many other immortalized and tumor cell lines (21). Examination of the 5' termini of 15 independent Rat-1 Pem cDNA clones showed a range of termini that were clustered in exon 1 (5–29 nt of exon 1 were included in these cDNA clones).

RNase protection analysis was performed with a probe complementary with this putative promoter region in exon 1 (probe E; Fig. 6A). This analysis was done with RNA from the Rat-1 cell line and placental tissue, since they both express high
shown). We conclude that multiple transcription initiation sites (Fig. 1). A 5′-UT Exon Unique to Skeletal Muscle—RNase protection analysis revealed that the Pp was not only transcriptionally active in reproductive tissue but also in skeletal muscle. We examined the 5′ termini of muscle Pem transcripts by the 5′-RACE method. Sequence analysis of the subcloned 5′-RACE products from muscle showed that the 5′ termini of these products was in the Pp in exon 1. Surprisingly, two of the six subcloned PCR products possessed a novel 45-nt sequence inserted in the 5′-UT region between exons 1 and 2. This sequence corresponded to a 45-nt sequence that we identified in genomic DNA between exons 1 and 2 that is flanked by canonical 5′ and 3′ splice sites (Figs. 1 and 2). Thus, this novel sequence is an alternatively spliced exon (termed the M exon) that is included in Pem transcripts in skeletal muscle (Fig. 2). The M exon possesses twoinitiator AUG codons (Fig. 1). The first AUG is followed by a termination codon five codons downstream. The second AUG is in-frame with the Pem protein reading frame and would dictate a seven-amino acid N-terminal extension to the Pem protein. However, the sequences surrounding both AUGs exhibit poor matches with the Kozak consensus sequence GCCRCCAUGG (36). They possess neither the critically important purine at position −3 nor the G at position +4.

To examine the regulation of M exon inclusion, a Pem probe containing the M exon (probe G) was prepared for RNase protection analysis (Fig. 7A). Fig. 7B shows that skeletal muscle RNA protected two bands (1 and 2), which represent M exon+ and M exon− transcripts, respectively. It is not known if the M exon+ transcripts are present in a specific subpopulation of cells in muscle or if they are present in all cell types in this tissue. No difference in the relative expression levels of the M+ and M− transcripts was noted in skeletal muscle from male and female animals (Fig. 7B). In contrast to skeletal muscle, we could not detect M+ transcripts in any other tissues or cell lines tested, including placenta, epididymis, testis, ovary, and Rat-1 cells (Fig. 7B and data not shown). Thus, the inclusion of the M exon by alternative splicing appears to be regulated in a tissue-specific manner.

Alternative Splicing of 5′-UT Exons in Placenta and Rat-1 Cells—Further Pem alternatively spliced mRNAs were revealed from sequence analysis of Rat-1 CDNA clones. Although most Rat-1 CDNA clones corresponded to A transcript (Fig. 2), two of the CDNA clones were derived from alternatively spliced mRNAs, termed B- and C-transcripts (Fig. 2). B-transcript is derived by an alternative splice acceptor in IVS1 (Fig. 1), which results in the inclusion of 38 nt from IVS1. RNase protection analysis with a probe prepared from this variant cDNA (probe H) showed that the B transcript was expressed at about 5-fold lower levels than the A transcript in placenta and the Rat-1 and MCA8994 cell lines (data not shown). A C transcript obtained by use of an alternative splice acceptor in IVS2 (Fig. 1), resulting in the inclusion of 61 nt from IVS2. RNase protection analysis showed that Rat-1 cell and placenta expressed C transcript (Fig. 7, C and D). The B- and C transcripts encode the same Pem protein as the more abundant A transcript, since the inclusion of additional 5′-UT sequences in the B- and C-transcripts did not introduce an initiator AUG upstream of the initiator AUG in exon 3. The regulatory significance of the alternative splice acceptors in IVS1 and IVS2 is not known.

An Exon-skipped Transcript Encodes a Novel Protein—Transcripts that skip exon 4 (ΔE4 transcripts) were revealed by sequence analysis of RT-PCR products generated from epididymis RNA. These ΔE4 transcripts originated from either the Pp or the Pp as shown by sequence analysis of subcloned epididymal RT-PCR products generated using any of the following upstream of the multiple start sites in the Pp (Fig. 1).
oligo combinations: C + B, J + I, or K + L (Fig. 1). RNase protection analysis with a probe that spanned exon 4 and the adjacent exons (probe J; Fig. 8A) showed that ΔE4 transcripts were not only expressed in epididymis but also in placenta and Rat-1 cells (Fig. 8B). In vitro translation of the ΔE4 transcript generated a smaller protein (Pem-E) than that translated from a normally spliced Pem transcript (Fig. 8C). Pem-E shares the first 26 amino acids of the amino terminus with the known Pem protein but contains 55 novel amino acids in the carboxyl terminus (Fig. 8D). Thus, Pem-E would lack the homeodomain (present in the carboxyl region of Pem) but would contain instead the most highly conserved region of the Pem protein that is present in the amino terminus (35).

Translational Regulation by Alternative 5'-UT Regions—We compared the translatability of three alternatively spliced Pem transcripts that differed only in their 5'-UT region: 1) A transcript; 2) M transcript expressed exclusively in muscle; and 3) T transcript expressed exclusively in testis and epididymis. The three different RNAs were generated in vitro and quantitated by both optical density and by visual inspection of 2-fold serial dilutions in agarose gels, and then equal amounts of the RNAs were translated in vitro using reticulocyte lysates. Multiple experiments with independent RNA samples demonstrated that T transcript was translated less efficiently than A transcript (4-fold lower average translation rate; Fig. 9). M transcript was also translated less efficiently than A transcript, although the average reduction was only 2-fold. We conclude that unique sequences present in the 5'-UT regions of these three Pem transcripts can alter the rate of translation in vitro.

**DISCUSSION**

In this report, we characterized the genomic structure of the Pem gene, defined two promoters used in a tissue-specific manner, and demonstrated that Pem transcripts undergo alternative RNA splicing events. We showed that an androgen-dependent promoter, Pp, is used exclusively in male reproductive tissue, while the other promoter, Pd, is expressed in female reproductive tissue and at low levels in skeletal muscle (Fig. 2). We found that several different modes of splicing regulation are exerted on Pem transcripts: 1) alternative exon inclusion; 2) alternative exon skipping; and 3) alternative splice acceptor usage (Fig. 2).

We showed that mRNAs transcribed from the Pp (T transcripts) require androgens for expression (Fig. 4), which is likely to explain why the Pp promoter is active in testis and epididymis, and is not used detectably in placenta, muscle, or ovary. The temporal pattern of T transcript expression during development differed in testis and epididymis. In prepubertal animals, T transcript levels were very high in epididymis but low in testis. T transcript levels were high as early as day 23 post partum in epididymis, while in testis T transcripts remained barely detectable until day 44 and only reached levels similar to that of Pd-derived transcripts at later developmental times (Fig. 5). The explanation for why T transcripts are regulated differently in testis and epididymis is not known. Since we showed that T transcript expression in epididymis requires testosterone, it is likely that the early postnatal expression of this transcript in epididymis is due to the known presence of androgens in the lumen of the epididymis at this development-
tal time point (37, 38). Less clear is why T-transcript levels are low in testis. Expression of T-transcripts in the testis in vivo requires testosterone, based on experiments in EDS-treated rats, but the available androgens in the testis may be insufficient to trigger strong expression. Androgen-binding protein, which is secreted by Sertoli cells and considered to act as an “androgen sink” in the testis and as an androgen-carrier protein to the epididymis, is known to be expressed at very high levels in rats (50–100-fold higher than in mice; Ref. 39) and thus may depress Pp transcription in rat testis. The specific androgens that are present in the testis at different developmental time points may also influence Pp expression. For example, the decline in the intratesticular levels of 5α-reductase after day 40 (40) would cause a switch in the ratio of dihydrotestosterone to testosterone and thus may influence transcription from the Pp.

The Pem gene is unusual in that it contains two promoter regions and three 5′-UT exons upstream of the coding region. As a result of alternative promoter usage and alternative splicing of these 5′-UT exons, Pem transcripts possess at least five different 5′ termini (the number of variants is even greater if one considers the multiple transcriptional initiation sites within both the Pp and the Pd). Since the use of some of these different 5′-UT sequences is regulated in a tissue- or androgen-dependent manner, it is tempting to speculate that they function in a regulatory capacity. We tested the effect of different Pem 5′-UT termini on translatability in vitro (Fig. 9) and found that Pem transcripts from the Pp (T-transcripts) were translated less efficiently than transcripts from the Pd (A-transcripts). Perhaps the male reproductive cell types that express Pp-derived transcripts down-regulate the level of Pem protein that is translated because deleterious effects would be caused by Pem protein overexpression. In skeletal muscle, a unique 5′-UT exon (the M exon) is included in Pem transcripts that is excluded in all other tissues (Fig. 7). We found that inclusion of the M exon depressed translation somewhat (Fig. 9), but since the effect was not dramatic, the M exon may regulate events other than translation. For example, it is known that 5′-UT sequences can regulate mRNA stability (45). It will be of interest to determine if the M exon plays a role in the dramatic induction of Pem transcripts in 10T½ mesenchymal stem cells when they commit to the muscle cell lineage (22). Secondary structure analysis by computer suggested that the different 5′ termini present in A+, M+, and T-transcripts possess different secondary structures that may be responsible for the different rates of translation. Tissue-specific factors may be present in vivo that differentially bind to these secondary structure regions and thereby regulate the translation rate of Pem mRNAs.

17544

**Fig. 8.** An exon-skipped transcript encodes a novel form of Pem protein. A, schematic diagram showing the region encompassed by probe J and the regions of the probe protected by RNA from different cellular sources. B, RNase protection analysis using probe J and total cellular RNA from epididymis (7 µg), placenta (5 µg), Rat-1 cells (3 µg), or tRNA (20 µg). C, SDS-polyacrylamide gel electrophoresis analysis of reticulocyte lysate-translated Pem and Pem-E protein using in vitro synthesized RNA containing the Pem and Pem-E reading frames as templates. D, predicted amino acid sequence of Pem-E encoded by the ΔE4 transcript. Amino acids 1–26 are in common with the Pem amino-terminal domain, Gly-27 is encoded by a codon across the exon 3/5 junction, and amino acids 28–81 are novel amino acids encoded by a different reading frame in exons 5 and 6.

---

2 S. Maiti, M. Griswold, and M. F. Wilkinson, unpublished observations.

3 J. Doskow and M. F. Wilkinson, unpublished observations.
that possess these different 5' termini. Although many studies have demonstrated that translation is highly regulated in germ cells (46, 47), little is known about translational regulation in somatic cells of the testis and epididymis, where Pem transcripts are expressed.4

We identified an alternatively spliced transcript (ΔE4) that encodes a novel form of the Pem protein (Pem-E). Many transcription factors, including homeobox transcription factors, are known to be expressed as multiple isoforms as a result of tissue-specific alternative RNA splicing (41). Pem-E shares the amino terminus with the classical Pem protein but lacks the homeodomain and thus may not bind DNA. Since the amino-terminal region of Pem is, by far, the most conserved region of this protein based on comparison of the primary amino acid sequence of mouse and rat Pem (35), this region may possess functional attributes. For example, the amino-terminal region may serve as a binding interface that permits Pem and Pem-E to bind other proteins. Many homeobox proteins have been shown to use amino acids outside of the homeodomain to interact with other transcription factors, including other homeobox proteins (9, 42, 43). The importance of regions outside of the homeodomain for biological function is underscored by a recent study showing that a mutant Ftz protein completely lacking the homeodomain correctly regulates downstream target genes in vivo, probably because it is still able to bind to other transcription factors (44). Pem-E may act as an inhibitor protein that competes with classical Pem for interaction with another transcription factor, but by virtue of its inability to bind to DNA, it would exert a dominant negative effect. By analogy, the Id inhibitor protein possesses a helix-loop-helix motif and thus can dimerize with other helix-loop-helix proteins, such as myoD, but because Id lacks a DNA-binding domain it prevents these interacting helix-loop-helix proteins from activating the transcription of downstream target genes (41).

Most known examples of alternative transcriptional and posttranscriptional events in male reproductive tissue are known to occur in the germ cells (47). For example, the c-mos, c-abl, pim-1, cytochrome c, cyclin D3, superoxide dismutase, hoxa-4, proopiomelanocortin, and SRY genes use alternative promoters in germ cells of the testis that differ from the promoters used in somatic cells (48–55). One hypothesis to explain the common usage of alternative promoters in germ cells is that it results from the changes in the chromatin structure needed to produce spermatozoa. This restructuring would not occur in somatic cells of the testis and epididymis, and thus transcriptional regulation unique to these tissues is not necessarily expected. The Pem gene is expressed by somatic cells of the testis and epididymis,4 and thus it will be of interest to determine how and why it is regulated in such a complex manner at both the transcriptional and post-transcriptional level.

Since the Pem gene encodes a homeodomain-containing protein, it is reasonable to suppose that the Pem protein is a transcription factor that regulates specific events during male gametogenesis. The finding that the Pem gene depends on androgen for expression in the epididymis suggests that, in turn, Pem may regulate androgen-dependent events in the epididymis. To our knowledge no transcription factors have previously been shown to be androgen-regulated in the epididymis. The homeobox transcription factor, Pax-2, is clearly regulated by a distinct mechanism, since it is expressed in the epididymis of tmf mice, which lack androgen receptors (18).

Several candidate downstream genes are known to require androgens for expression in the epididymis (directly or indirectly) and thus may be regulated by Pem, including those encoding 5α-reductase, carboxypeptidase metalloprotein D/E (AEG, CRISP-1), the retinol binding protein B/C (ESPI), the glutathione peroxidase-like protein GPX, the glutamyltranspeptidase GGT, nerve growth factor, and E-cadherin (7, 8). The epididymis has multiple functions, many of which depend on the presence of androgens and thus may be regulated by Pem: (i) induction of spermatozoa motility capability, (ii) spermatozoa membrane alterations that permit fertilization competence, (iii) changes in the methylation status of spermatozoa genes, and (iv) spermatozoa storage (6–8, 56). Since the Pem gene is specifically expressed in the distal corpus/proximal cauda portion of the epididymis,4 the site where spermatozoa gain motility capability and membrane alterations necessary for fertilization competence (57, 58), it will be of interest to determine whether Pem regulates these final maturation events.

Acknowledgments—We express gratitude to Drs. Thomas Cooper, Gail Cornwall, Gilbert Cote, Michael Griswold, Marvin Meistrich, and Walter Tribley for critical reading of the manuscript.

REFERENCES
1. Santulli, R., Spirando, R. L., Awonyi, C. A., Ewing, L. L., and Zirken, B. R. (1990) Endocrinology 126, 95–101.
2. Sun, Y., Wreford, N. G., Robertson, D. M., and de Kreter, D. M. (1990) Endocrinology 127, 1215–1223.
3. Singh, J., O'Neill, C., and Handelsman, D. J. (1995) Endocrinology 136, 5311–5321.
4. Sar, M., Hall, S. H., Wilson, E. M., and French, F. S. (1993) in The Sertoli Cell (Russell, L. D., and Griswold, M. D., eds) pp. 510–516, Cache River Press, Clearwater, FL.
5. Norton, J. N., and Skinner, M. K. (1989) Endocrinology 124, 2711–2719.
6. Robarre, B., and Hermo, L. (1988) in The Physiology of Reproduction (Knobil, E., and Neil, J., eds) pp. 999–1080, Raven Press, New York.
7. Orgebin-Crist, M.-C. (1996) in Pharmacology, Biology, and Clinical Applica-

4 J. S. Lindsey and M. F. Wilkinson, submitted for publication.
ions of Androgens, (Bhasin, S., and Sweadnilf, R. S., eds), Wiley-Liss, Inc., NY, in press.
8. Cornwall, G., and Hann, S. R. (1995) J. Androl. 16, 379–383.
9. Duboule, D. (1994) Guidebook to the Homeobox Genes, Oxford University Press, Oxford.
10. Gruss, P., and Walther, C. (1992) Cell 69, 719–722.
11. Corcoran, L. M., Karvelas, M., Nossal, G. J. V., Ye, Z.-S., Jacks, T., and Baltimore, D. (1993) Genes Dev. 7, 570–582.
12. Krumlauf, R. (1994) Cell 78, 191–201.
13. Andersen, B., Pearse, R. V., II, Schlegel, P. N., Schlegel, P. N., Cichon, Z., Schonemann, M. D., Bardin, C. W., and Rosenfeld, M. G. (1993) Proc. Natl. Acad. Sci. U. S. A. 90, 11084–11088.
14. Wolgemuth, D. J., Viviano, C. M., and Watrin, F. (1991) Ann. N. Y. Acad. Sci. 637, 300–312.
15. Watrin, F., and Wolgemuth, D. J. (1993) Dev. Biol. 156, 136–145.
16. Featherstone, M. S., Baron, A., Gaunt, S., Mattei, M. G., and Duboule, D. (1993) Dev. Biol. 156, 136–145.
17. Graham, A., Papalopulu, N., Lorimer, J., McVey, J. H., Tuddenham, E. G. D., and MacLeod, C. L. (1994) Proc. Natl. Acad. Sci. U. S. A. 91, 637, 11084–11088.
18. Fickenscher, H. R., Chalepakis, G., and Gruss, P. (1993) DNA Cell Biol. 12, 381–391.
19. Xin, J.-H., Cowie, A., Lachance, P., and Hassell, J. A. (1992) Genes Dev. 6, 481–496.
20. MacLennan, C. L., Fang, A. M., Seel, B. S., Walls, L., and Wilkinson, M. F. (1990) Cell Growth & Differ. 1, 271–279.
21. Wilkinson, M. F., Kleeman, J., Richards, J., and MacLennan, C. L. (1990) Dev. Biol. 141, 451–455.
22. Sasaki, A. W., Doskow, J. K., MacLennan, C. L., Rogers, M. B., Gudas, L. J., and Wilkinson, M. F. (1991) Mech. Dev. 34, 155–164.
23. Rayle, R. E. (1991) Dev. Biol. 146, 255–257.
24. Lin, T.-P., Labosky, P. A., Grabel, L. B., Kozak, C. A., Pitman, J. L., Kleeman, J., and MacLennan, C. L. (1994) Dev. Biol. 166, 170–179.
25. Labosky, P. A., Weir, M. P., and Grabel, L. B. (1993) Dev. Biol. 159, 232–244.
26. Szczechowicz, J., Frischauf, A. M., and Muroyama, M. (1988) Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.
27. Siebert, P. D., Chenchik, A., Kellog, D. E., Lukyanov, K. A., and Lukyanov, S. A. (1985) Nucleic Acids Res. 13, 1087–1088.
28. Wilkinson, M. F. (1991) in Essential Molecular Biology: A Practical Approach (Brown, T. A., ed) pp. 69–87, Oxford University Press, New York.
29. Chomczynski, P., and Sacchi, N. (1987) Anal. Biochem. 162, 156–159.
30. Ausubel, F. M., Brent, R., Kingston, R. E., Moore, D. D., Smith, J. A., Seidman, J. G., and Struhl, K. (1987) Current Protocols in Molecular Biology, sections 3.8 and 4.7, John Wiley & Sons, Inc., New York.