Structure and Promoter Activity of the LpS1 Genes of Lytechinus pictus

DUPLICATED EXONS ACCOUNT FOR LpS1 PROTEINS WITH EIGHT CALCIUM BINDING DOMAINS

(Received for publication, January 15, 1991)

Mengqing Xiang, Tong Ge, Craig R. Tomlinson, and William H. Klein

From the Department of Biochemistry and Molecular Biology, University of Texas, Austin, Texas 78705 and the Department of Biology, University of Houston, Houston, Texas 77204

The LpS1 genes of the sea urchin Lytechinus pictus are activated early in development in aboral ectoderm cells. They therefore have ontogenic properties similar to their counterparts in Strongylocentrotus purpuratus, the Spec genes. Both gene families encode proteins belonging to the calmodulin superfamily as evidenced by the presence of distinct EF-hand (helix-loop-helix) domains. The presence of eight EF-hand domains in LpS1 proteins suggests that the LpS1 genes arose from a duplication of an ancestral Spec-like gene. The LpS1 genes were further analyzed to increase our understanding of the mechanisms underlying their evolution and activation in aboral ectoderm cells. Genomic DNA blot analysis showed two LpS1 genes, LpS1α and LpS1β, which did not appear to be closely linked. LpS1 genomic clones were isolated by screening an L. pictus genomic library with an LpS1 cDNA clone, and partial gene structures for both LpS1α and LpS1β were constructed. These revealed internal duplication of the LpS1 genes that accounted for the eight EF-hand domains in the LpS1 proteins. Duplication of exon 1 in both genes suggested four different LpS1 proteins could be derived from the LpS1 genes. Primer extension to map the transcriptional initiation sites of the LpS1 genes and sequencing analysis showed there was little in common among the 5′-flanking regions of the LpS1 and Spec genes except for the presence of a binding site for the transcription factor USF. A sea urchin gene-transfer expression system showed that 762 base pairs (bp) of 5′-flanking DNA and 17 bp of 5′-untranslated leader sequence of the LpS1β gene were sufficient for correct temporal and spatial expression of reporter chloramphenicol acetyltransferase and lacZ genes in sea urchin embryos. Deletions at the 5′ end to either 511 or 368 bp resulted in a 3–4 fold decrease in chloramphenicol acetyltransferase activity and disrupted the exclusive activation of the lacZ gene in aboral ectodermal cells. Based on a lineage analysis among the LpS1 and Spec gene families and other related genes, we propose a model in which LpS1 genes evolved from a series of duplications of an ancestral Spec-like gene.

Calmodulin and its many relatives form a large, complex superfamily with highly diverse functions and widespread distribution. The superfamily consists mainly of low molecular weight intracellular calcium-binding proteins related to each other through a conserved helix-loop-helix motif, the EF-hand (Kretsinger et al., 1988). It is thought that the superfamily evolved from a primordial EF-hand-containing polypeptide by tandem gene duplication, sequence divergence, and dispersion (Demain, 1982). Many of these proteins play important roles in modulating intracellular calcium levels and in regulating cellular events via calcium (Norman et al., 1987). Much is already known about their evolution, expression, and function (Moncrief et al., 1990).

In sea urchins, the calcium-binding Spec proteins represent interesting examples of the evolution and specialization of superfamily members. These proteins accumulate in embryos and larvae, but not in adults, and are found only in cell lineages giving rise to the aboral ectoderm, a squamous epithelium covering the surface of the late stage embryo and larva (Bruskin et al., 1982; Lynn et al., 1983; Carpenter et al., 1984; Klein et al., 1990). The exact function of the Spec proteins is unknown, but they may play a role in elevating calcium ion concentration in the aboral ectoderm, perhaps by enhancing transport of calcium from the sea water to the blastocoel, where high calcium concentrations are required for skeletogenesis (Klein et al., 1991).

We have been studying the Spec proteins and their genes from two different sea urchin species, Strongylocentrotus purpuratus and Lytechinus pictus, with the notion that a comparative analysis would be useful for gaining insights into Spec protein function. It should also be helpful for elucidating the mechanisms by which the Spec genes are activated specifically in aboral ectoderm cells.

In *S. purpuratus*, there are seven or eight Spec genes encoding approximately 10 polypeptides as visualized by two-dimensional gel analysis (Bruskin et al., 1982; Hardin et al., 1988). The *S. purpuratus* Proteins range in size from 14,000 to 17,000 Da and contain four helix-loop-helix domains per molecule (Bruskin et al., 1982; Carpenter et al., 1984; Klein et al., 1991). Recently, we initiated experiments to map the transcriptional regulatory elements for three Spec genes using a sea urchin embryo gene-transfer system. Promoter regions from all three genes are capable of activating a CAT reporter gene at the appropriate time in development, but only the Spec 2a promoter yields aboral ectoderm specificity when fused with a lacZ reporter gene (Gan et al., 1990a, 1990b).  

The abbreviations used are: CAT, chloramphenicol acetyltransferase; bp, base pair(s); kb, kilobase pair(s); Pipes, 1,4-piperazine-dicarboxylic acid; X-Gal, 5-bromo-4-chloro-3-indolyl-β-D-galactoside.

* This study was supported by National Institutes of Health Grant HD22619 (to W. H. K.). The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be 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/EMBL Data Bank with accession number(s) M62848.
Spec2a, 1516 bp of 5'-flanking DNA plus 18 bp of 5'-untranslated leader sequence are sufficient for aboral ectoderm expression, while 5.6 kb of Spec 1 5'-flanking DNA plus 120 bp of 5'-untranslated leader sequence result in preferential expression in mesenchymal cell types (Gan et al., 1990a).

All the characterized Spec genes contain an enhancer-like element in their 5'-flanking DNA that appears to be partly responsible for the activation of transcription at the correct time but does not confer aboral ectoderm specificity by itself (Gan et al., 1990b). The Spec gene promoter regions also contain a high affinity binding site for USF, a transcription factor originally found associated with the upstream region of the major late genes of adenovirus (Sawadogo and Roeder, 1985; Tomlinson et al., 1990). In sea urchins, USF binding activity is found only in ectodermal cells; posttranslational inactivation may occur in other cell types (Tomlinson et al., 1990).

In contrast to S. purpuratus, L. pictus has only three Spec-like proteins, with molecular sizes of 34,000 rather than 14,000-17,000 Da (Xiang et al., 1988). A cDNA clone, called LpS1, encodes at least one of these proteins and contains eight helix-loop-helix domains rather than four, suggesting that, in L. pictus, the Spec-like proteins are tailed only duplicated versions of their S. purpuratus counterparts (Xiang et al., 1988). The LpS1 proteins are highly diverged from the S. purpuratus Spec proteins, showing no more similarity to them than to other seemingly distant members of the superfamily (Xiang et al., 1988; Moncrief et al., 1990; Klein et al., 1991). Nevertheless, LpS1 genes are activated at the same developmental stage as Spec genes, and LpS1 messages accumulate only in aboral ectoderm cells, implying an identical functional role for LpS1 and Spec proteins in the two species (Xiang et al., 1988).

We employed the LpS1 cDNA to ask several specific questions regarding the evolution and expression of the Spec-like genes in the two species. We show that there are two LpS1 genes in the L. pictus genome, resulting from a recent gene duplication. The two LpS1 genes could result in as many as four different LpS1 proteins. The coding exons of the LpS1 genes have been duplicated, explaining the eight helix-loop-helix domains in the LpS1 proteins. We also show that 762 bp of 5'-flanking and 17 bp of 5'-untranslated leader DNA are sufficient for activating the LpS1 genes at the appropriate time and in the correct cell type. In addition, sequence comparisons among LpS1 and Spec 5'-flanking regions show that the USF binding site is a major conserved element.

**EXPERIMENTAL PROCEDURES**

_Southern Blot Analysis of Genomic DNA_—Genomic Southern blot analysis was done as described (Hardin et al., 1988). _L. pictus_ sperm DNA was digested with an appropriate restriction enzyme, separated on an agarose gel, transferred to N-Hydroy filter, and hybridized to the 1.1-kb EcoRI fragment of the LpS1 cDNA clone (Xiang et al., 1988). Final wash conditions were 68°C and 0.15 M NaCl.

_Pulsed Field Gel Electrophoresis—_L. pictus sperm were washed twice in ice-cold phosphate-buffered saline, 1 mM dithiothreitol, and resuspended in phosphate-buffered saline to a concentration of 1.6 × 10^6 sperm/ml. An equal volume of 1% low-melting agarose (FMC Bioproducts) was added to the sperm for a total volume of 3 ml. The temperature of the sperm suspension and agarose was adjusted to 42°C, and the two were mixed completely, pipetted into a plus mold (Bio-Rad), and incubated at 4°C for 5 min. The plugs were transferred to 50 ml of 0.5 M EDTA, pH 8.0, 0.1% lauroylsarcosine, and 1 mg/ml proteinase K (ESF buffer) and incubated at 50°C for 48 h. The ESF buffer was decanted and the plugs washed twice for 1 h with 50 ml of TE (10 mM Tris-HCl, 1 mM EDTA, pH 7.5) containing 1 mM phenylmethanesulfonyl fluoride at room temperature. The plugs were washed an additional time in 50 ml of TE without phenylmethanesulfonyl fluoride for 30 min and stored at 4°C in 10 ml of 0.5 M EDTA.

The DNA was prepared for restriction enzyme digestion by washing the plugs two times with 50 ml of TE for 30 min at room temperature, and 30 min at 4°C in 3 volumes of restriction enzyme buffer. The buffer was replaced and replaced with 3 volumes of fresh restriction enzyme buffer. Approximately 40 units of the appropriate restriction enzyme were added, and digestion was carried out for 24 h at the appropriate temperature.

The plugs were loaded directly into the wells of a 1% agarose gel (Bio-Rad CHEF-DR II Megabase DNA pulsed field electrophoresis system) and electrophoresed for ramp times of 30-40 s at 200 V. The gels were stained, photographed, and blotted as described by Sambrook et al. (1989). Hybridizations and washes were performed as in Tomlinson and Klein (1990).

_Screening of an L. pictus Genomic Library—_An L. pictus ARMB13 genomic library (Hardin et al., 1987) was screened and rescreeened using the 0.52- and 1.1-kb EcoRI fragments of the LpS1 cDNA clone (Xiang et al., 1988) at a final wash condition of 68°C, 0.15 M NaCl. Sixteen positive clones were obtained and subsequently mapped by digestion with restriction enzymes.

_Sequencing of LpS1 Genomic Clones—_All positive clones were hybridized to the three EcoRI fragments of the LpS1 cDNA clone. The smallest hybridizing fragments were subcloned into Bluescript. Oligonucleotides corresponding to various exons of the LpS1 gene were synthesized and used to sequence both the LpS1α and LpS1β hybridizing fragments by the chain termination method (Sanger et al., 1977).

_Primer Extension Analysis—_A 17-bp primer complementary to +106 to +90 in the LpS1α or LpS1β gene was synthesized and end-labeled with [γ-32P]ATP. Specific activity of L. pictus α-catenin (Promega, Erase-a-base kit). The 511-CAT construct was linearized with EcoRI, and 20 μg of L. pictus total cDNA RNA from plateau stage embryos was annealed to 15 ng of labeled primer in 5 μl of annealing buffer (40 mM Pipes, pH 6.4, 1 mM EDTA, and 0.4 M NaCl) for 4 h at 42°C. The reaction was then carried out for 2 h at 42°C in a 50-μl solution containing 50 mM Tris-HCl (pH 8.3), 75 mM KCl, 3 mM MgCl2, 0.5 mM each dATP, dCTP, and dTTP, 80 units of RNasin (Promega Biotec, Madison, WI), and 200 units of Moloney murine leukemia virus reverse transcriptase (BRL). The reaction products were precipitated with 2.5 volumes of ethanol and separated on an 8% sequencing gel.

_Plasmid Constructs for Microinjection Assays—_All the LpS1-CAT constructs were derived from pSVo·CAT (S), which was a derivative of pSVo·CAT (Gorman et al., 1982; Fylytanis et al., 1987). The 762-CAT construct was obtained by subcloning the HincII-EcoRI fragment of LpS1β from +17 to +762 bp into pSVo·CAT(S) in a forward orientation. A corresponding construct in the reverse orientation was also obtained. Deletion mutants were obtained following the EcoRI deletion protocol (Promega, Erase-a-base kit). The 511-CAT construct contained LpS1β upstream sequences from +17 to −511 bp, and the 368-CAT constructs contained sequences from +17 to −368 bp.

The LpS1α-CAT constructs were generated from the pNL vector (Gan et al., 1990a). To create the 762-CAT, the LpS1β HincII-EcoRI fragment containing 5'-flanking sequences from +17 to −762 bp was subcloned into the pNL vector in a forward orientation through SpeI linker ligation. To generate the 3700-CAT construct, an upstream 2.9-kb EcoRI-PstI fragment of LpS1α containing sequences from −763 to −9700 bp was inserted into the PstI site of the 762-CAT construct. In a forward orientation, the PstI linker ligation. The deleted HincII-EcoRI fragments were also subcloned into the pNL vector in a reverse orientation to generate the 511-CAT and 368-CAT constructs.

_Microinjection of Sea Urchin Eggs or Zygotes_—The collection of sea urchin gametes and microinjection of eggs was the same as previously described (McMahon et al., 1985). Injections were performed with either unfertilized or fertilized eggs. Each egg was injected with approximately 2 pl of a 40% glycerol solution containing 2000 molecules of the linearized plasmid constructs plus sea urchin sperm carrier DNA at a molar ratio of 1:5. After microinjection, L. pictus embryos were cultured at 18°C in 10 mM Tris-HCl, 1 mM EDTA, and 0.1 M NaCl, with 0.5 M each dATP, dCTP, dGTP, and dTTP, 80 units of RNasin (Promega Biotec, Madison, WI), and 200 units of Moloney murine leukemia virus reverse transcriptase (BRL). The reaction products were precipitated with 2.5 volumes of ethanol and separated on an 8% sequencing gel.

_CAT Assays and DNA Determinations—_Microinjected embryos
were harvested at the desired developmental stages and mixed with approximately 1500 uninjected embryos. One-half of these embryos were used to measure CAT activity (McMahon et al., 1985); the other half were used to determine the plasmid DNA levels. Slot blot experiments were performed (Flytzanis et al., 1987), hybridized with a CAT DNA probe, and washed in 0.15 M NaCl at 68 °C.

X-Gal Staining and Microscopy—X-Gal staining for β-galactosidase activity was performed as described by Gan et al. (1990a). Several hundred *L. variegatus* or *L. pictus* fertilized eggs were injected with the LpS1b-lacZ constructs, harvested at gastrula through prism stages, and fixed with 1% glutaraldehyde in Ca²⁺-free sea water at room temperature for about 15 min. After fixation, the embryos were rinsed twice with staining solution (10 mM NaPO₄, pH 7.0, 150 mM NaCl, 1 mM MgCl₂, 5 mM K₃Fe(CN)₆, 5 mM K₄Fe(CN)₆, and 0.2% X-Gal) and stained for periods ranging from overnight to 2 days at room temperature. Stained embryos were observed by Nomarski interference optics on an inverted Nikon Diaphot microscope and photographed on Kodak Gold 100 film (ASA 100).

**RESULTS**

**Two LpS1 Genes**—In previous work we characterized an LpS1 cDNA clone isolated from an *L. pictus* gastrula cDNA library using an oligonucleotide probe corresponding to a consensus helix-loop-helix motif (Xiang et al., 1988). Three EcoRI fragments of 0.52, 0.44, and 1.1 kb, generated from the cDNA insert, represented the 5′-untranslated leader sequence plus the first half of the protein coding sequence, the second half of the protein sequence, and the 3′ end of the coding sequence plus the 3′-untranslated trailer sequence, respectively. Preliminary experiments using these EcoRI fragments as probes for genomic Southern blot analysis suggested that *L. pictus* contained only a single LpS1 gene unlike the multiple *Spec* gene family in *S. purpuratus* (Tomlinson and Klein, 1990).

To determine the number of LpS1 genes more precisely, we performed a series of quantitative slot blot and genomic Southern blot experiments with the EcoRI fragments of LpS1. Slot blot analysis was consistent with the notion of one LpS1 gene but could not eliminate the possibility of a second (data not shown). However, genomic Southern blots with the LpS1 1.1-kb EcoRI fragment clearly indicated a second LpS1 gene (Fig. 1). The 1.1-kb EcoRI fragment contained the 3′-untranslated trailer sequence and corresponded to the final exon of the LpS1 gene, based on comparison with the *Spec* genes of *S. purpuratus*. Below we demonstrate this also to be the case with genomic LpS1 clones. The number of hybridizing bands therefore indicates the number of 3′ LpS1 exons in the genome.

Sperm DNA from two individuals was probed with the 1.1-kb EcoRI fragment. In a representative blot, all lanes from both individuals showed two hybridizing bands, one strong and one weak, with the exception of the lane labeled X in individual 1 (Fig. 1). The simplest interpretation is that the strong band corresponds to the LpS1 gene from which the cDNA clone was derived and the weaker band to a second LpS1 gene. For clarity, we call these two genes LpS1α and LpS1β, respectively. In lane X of individual 1, a restriction fragment length polymorphism can be seen in the LpS1α gene. This individual appeared to be heterozygous at the LpS1α locus; one allele contained the 3′-untranslated trailer exon within an 8.5-kb XbaI fragment, and a second allele, not present in individual 2, contained the exon within a 6.0-kb XbaI fragment. Both the 8.5- and 6.0-kb XbaI fragments were cut by HindIII to produce the 4.5-kb HindIII-XbaI fragment visualized in the X/H lanes of both individuals (Fig. 1). The experiments shown in Fig. 1 strongly implied the existence of two LpS1 genes.

We used pulsed field gel electrophoresis to determine whether the LpS1α and LpS1β genes were physically linked. Sperm DNA from three individuals was digested with either SalI, NotI, or SfiI and subjected to pulsed field gel electrophoresis. Hybridization with the 0.52-kb EcoRI fragment of the LpS1 cDNA clone, which represents the 5′-half of the cDNA, showed two or three hybridizing bands in all digestions, ranging in size from 40 to 450 kb (Fig. 2). These results are consistent with the presence of two polymorphic LpS1 genes in the *L. pictus* genome. From the intensity of the hybridizing fragments, we could predict which contained the LpS1α or LpS1β gene (Fig. 2). However, since we did not observe both genes on a single SalI, NotI, or SfiI fragment, it was not possible to determine conclusively whether the genes were linked. NotI and SfiI are enzymes with rare cutting sites appearing on an average of every 200–300 kb in the *L. pictus* genome (data not shown). The probability of both NotI and SfiI cutting between two linked LpS1 genes is low, and the fact that they did suggest that the LpS1α and LpS1β genes were not closely linked.

**Partial Structures of the LpS1α and LpS1β Genes**—We used the 0.52-, 0.44-, and 1.1-kb EcoRI fragments to screen an *L. pictus* genomic library for the LpS1 genes. DNA isolated from the positive phage was mapped with several restriction enzymes, and genomic fragments hybridizing to different regions of the LpS1 cDNA were subcloned and sequenced using oligonucleotide primers generated from the LpS1 cDNA sequence. The genomic clones fell into three nonoverlapping regions as shown in Fig. 3. Two regions corresponded to the 5′- and 3′-most ends of the LpS1α gene, judging by sequence identity between the LpS1 cDNA clone and the genomic

**Fig. 1.** Determination of LpS1 gene copy number in *L. pictus* by genomic Southern blot analysis. 10 μg of *L. pictus* sperm DNA from individuals 1 and 2 were digested with EcoRI (R), BamHI (B), XbaI (X), or XhoI and HindIII (X/H), electrophoresed, transferred, and hybridized with the 1.1-kb EcoRI fragment of the LpS1 cDNA clone. The blot was washed in 0.15 M NaCl at 68 °C. Sizes of the hybridizing bands are indicated on the right.

**Fig. 2.** Linkage analysis of LpS1 genes by pulsed field gel electrophoresis. *L. pictus* sperm DNA from individuals 1, 2, and 3 were digested with SalI, NotI, and SfiI, subjected to pulsed field gel electrophoresis, transferred to filters, and hybridized with the 0.52-kb EcoRI fragment of the LpS1 cDNA clone. The blot was washed in 0.15 M NaCl at 66 °C. Size markers are indicated on the right.
fragments (data not shown; Xiang et al., 1988). The third group of genomic clones corresponded to all but the final two exons of the second gene, LpSlβ. Its sequence was different from LpSlα (see below). Despite repeated attempts, we were unable to isolate the middle portion of the LpSlα gene or the 3' exons of the LpSlβ genes. Their putative positions are shown parenthetically in Fig. 3.

Several noticeable features emerged from analysis of the LpSlα and LpSlβ gene structures. Both genes contained two exon 1s, and the downstream exon 1s (labeled 1* in Fig. 3) were truncated by six codons at their 3' ends, allowing each LpSl gene to encode two slightly different proteins by differential promoter utilization. The LpSlβ gene, for which most of the reading frame exons have been cloned, contained exons 2-5 and shared 99% sequence identity with the 5' half of the LpSlα cDNA clone. The LpSlβ gene had similar exon/intron junctions to those found in the S. purpuratus Spec genes (Hardin et al., 1988, 1988). Following these exons were exons corresponding to the second half of the LpSlα cDNA. We had previously shown that the LpSl cDNA contained a duplicated reading frame encoding a protein with twice the molecular weight of the Spec proteins (Xiang et al., 1988). Because these latter exons have sequence similarity with exons 2-5 and because they have similar exon/intron junctions, we labeled them 2', 3', 4', and 5' (exon 5' sequence was obtained from the LpSlα gene) to indicate that they were duplicated versions of exons 2, 3, 4, and 5, respectively (Fig. 3). It can be readily seen from the partial structures of LpSlα and LpSlβ that the LpSl duplication event was the result of an internal duplication of coding exons 2-5.

A partial sequence of the LpSlβ gene is shown in Fig. 4A. In most genes belonging to the calmodulin superfamily, including the S. purpuratus Spec genes, the first exon ends with the initiator methionine codon (Hardin et al., 1985; 1988; Perret et al., 1988). However, in the LpSlα and LpSlβ genes, all four first exons contain coding sequence beyond the AUG codon. The upstream exon 1s of both LpSlα and LpSlβ contained codons that matched the LpSlα cDNA clone precisely (Fig. 4B), while the downstream exon 1s were truncated at the second codon, a serine. If LpSlα proteins were generated from these downstream exon 1s, they would be missing amino acid residues 3-8. Fig. 4B shows the predicted amino acid sequence for LpSlα (derived from the cDNA clone; Xiang et al., 1988) and LpSlβ (derived from the genomic sequence). Although the proteins were obviously similar, it is interesting to note that the most significant divergence existed outside of the calcium binding helix-loop-helix domains (underlined in Fig. 4B).

In order to map the 5' ends of the LpSlα and LpSlβ genes, we generated an oligonucleotide primer within the first exon (bracketed in Fig. 4A) and performed a primer extension analysis. Two major extended fragments could be observed, 105 and 106 bp in length (Fig. 5). The last base of the longer product corresponded to an A residue 99 bp from the initiator codon, and this base was arbitrarily assigned +1 on Fig. 4A. Since there were two exon 1s for both LpSl genes and all four of these exons have nearly identical sequences, we could not distinguish among them with the primer extension technique. Sequences upstream of the 5' end of all four exons showed more than 99% identity for 326 bp; then exons 1 and 1* of the LpSlα gene diverged completely from exons 1 and 1* of the LpSlβ gene. In the latter gene, sequences upstream from exons 1 and 1* were nearly the same for 762 bp from their 5' ends, and then they diverged. Thus, it was possible that transcripitonal elements sufficient for appropriate temporal and spatial expression lay upstream of both first exons of both LpSl genes. The 5'-flanking sequences contained a putative TATA box at −17 to −22 bp and a USF binding site at −526 to −531 bp (Fig. 4A). The USF site was found to bind USF specifically in vitro, albeit with weaker affinity than the corresponding USF sites within the Spec gene promoters. A comparison of 5'-flanking DNA among the LpSl and Spec genes showed no other convincing similarities. In particular, there was no homology with the S. purpuratus RSR element, a highly conserved 600-bp region flanking the Spec genes and containing an enhancer-like element within it.

**Figure 3. Partial structures of LpSl genes.** The two long horizontal lines represent genomic DNA regions containing LpSlα and LpSlβ genes. The overlapping lines below depict the 16 genomic clones obtained by screening an L. pictus genomic library with the LpSl cDNA clone. Above are restriction enzyme sites. B, BamHI; H, HindIII; Hc, HincII; N, NdeI; P, PstI; R, EcoRI; S, SalI; Xb, XbaI; Xh, XhoI. Underlined R reflects the restriction fragment length polymorphism for EcoRI sites. The numbered boxes indicate exons and their relative positions. The open boxes represent untranslated regions, while the filled boxes represent translated regions. Parentheses indicate genomic DNA regions that are not cloned.

* M. Xiang and W. Klein, unpublished results.
gene on the pSVO plasmid. The LpS1\textbeta-CAT plasmid was injected into *L. pictus* or *L. variegatus* eggs and CAT activity monitored at different developmental stages. We have shown previously that CAT activity resulting from pSVO is virtually absent in *L. pictus* and *L. variegatus* eggs and embryos without appropriate promoter elements (Gan et al., 1990b). LpS1\textbeta-CAT activity produced in hatched blastulae and later stage embryos of *L. pictus* was quantitatively comparable with what we had found previously with Spec promoters (Fig. 6; Gan et al., 1990b). Less than one-tenth of this activity could be seen when the LpS1 fragment was inserted into pSVO in the reverse orientation (data not shown). Moreover, no detectable activity was observed with cleavage stage embryos. Though the amount of plasmid DNA was not as high in cleavage stage embryos as in later stage embryos, owing to less DNA amplification, sufficient DNA was present that any expression should have been detected. In the experiment shown in Fig. 6, the ratio of LpS1\textbeta-CAT DNA at the hatched blastula stage to that in the cleavage stage was about 10. Assuming that the levels of DNA in these experiments were nonsaturating and thus proportional to CAT activity, we should have observed at the cleavage stage 10% of the level of CAT activity of the hatched blastula stage. However, we could detect no more than 0.1% (Fig. 6A). In Fig. 6B, it can be seen that CAT activity remained fairly constant through the remainder of *L. pictus* embryogenesis. Similar results were obtained with *L. variegatus* embryos (data not shown). Earlier work using nuclear run-on analysis has shown that the LpS1 genes are activated at the end of the cleavage stage and remain active (Tomlinson and Klein, 1990). The results presented in Fig. 6 were in good agreement with these nuclear run-on experiments and suggested that the LpS1-CAT plasmid behaved appropriately with regard to temporal activation.

To monitor the embryonic cell types capable of utilizing the LpS1 promoter elements, we fused either the same DNA fragment (~762 to +17 bp) or one containing from about -3700 to +17 bp to lacZ. Injections were performed on *L. pictus* or *L. variegatus* fertilized eggs, and late gastrula/early prism stage embryos were harvested and stained for \(\beta\)-galactosidase activity with X-Gal. It should be noted that the injected DNA was incorporated into only a fraction of the cleaving nuclei during development, and although the incor-
The complementary sequence is shown in Fig. 4. The TATA box corresponding to the 106-bp major primer extension product. The products of primer extension (I') and sequencing reactions (G, primer was radiolabeled at its 5' end) from pluteus stage embryos. The plasmid DNA content of total ectodermal RNA isolated from pluteus stage embryos. The vertical line at the 64-cell stage is random with respect to cell type, the DNA at the 64-cell stage (HB), uninjected embryos; and 890 embryos were harvested at the 64-cell stage (C); 600 at hatched blastula stage (HB); and 220 at the remaining stages Bg, uninjected embryos; St, uninjected embryos with 0.1 unit of bacterial chloramphenicol acetyltransferase.

The principal reason for characterizing the LpS1 gene was to make comparisons with the S. purpuratus Spec genes in the hopes of identifying conserved regions important for expression and function. There are seven or eight Spec genes, and four of them (Spec 1, 2a, 2c, and 2d) have been cloned and characterized (Hardin et al., 1985, 1988). Spec 2a and 2c are the most similar in sequence, and Spec 1 is more similar to them than Spec 2d (Carpenter et al., 1984; Hardin and Klein, 1987; Moncrief et al., 1988). The highly diverged Spec 2d is only weakly expressed and appears to encode a nonfunctional product (Hardin et al., 1985, 1988; Klein et al., 1991). All four genes have an 800-bp conserved sequence block, designated RSRA, upstream of their translational start sites, we have
Fig. 7. Spatial expression of 3700-lacZ and 762-lacZ in L. variegatus embryos. Gastrula stage embryos injected with 3700-lacZ (A–C) or 762-lacZ (D–F) were fixed and assayed for β-galactosidase activity by X-Gal staining. Shown in this figure are six representative embryos with stained aboral ectoderm cells. In all panels but C, embryos are shown with their oral ectoderm on the top and future anus on the bottom. Panel C shows an embryo in cross-section. Panel E, oe, oral ectoderm; ae, aboral ectoderm; pm, primary mesenchyme; sm, secondary mesenchyme; en, endoderm.

Table I

| Constructs | Total stained embryos | Correct* | Incorrect* | Percent correct* |
|------------|-----------------------|----------|------------|-----------------|
| 3700-lacZ  | 129                   | 128      | 1          | 99              |
| 762-lacZ   | 239                   | 215      | 24         | 91 ± 6.2 (n = 4) |
| 511-lacZ   | 147                   | 113      | 34         | 73 ± 10.7 (n = 3) |
| 368-lacZ   | 80                    | 56       | 24         | 67 ± 6.6 (n = 2) |

* Embryos were scored as correct spatial expression if only aboral ectoderm cells were stained.
* Embryos were scored as incorrect if unambiguous staining was seen in cell types other than aboral ectoderm, regardless of whether aboral ectoderm cells were stained or not.
* If applicable, the average value is expressed with a standard deviation. n is the number of experiments.

suggested it contains cis elements important for Spec gene activation (Gan et al., 1990b). In Spec 2a, this block is continuous but in Spec 1, 2c, and 2d, inserted DNA interrupts the conserved block, displacing the RSR portion and parts of the A region a few kilobases upstream (Gan et al., 1990b). The 5' half of the RSR region contains a transcriptional element that appears to be a temporal enhancer, and the A region contains the above mentioned USF binding site, which binds sea urchin USF with high affinity (Gan et al., 1990b; Tomlinson et al., 1990). We have shown USF activity is present only in ectoderm cells but have not yet demonstrated a role for USF in the transcriptional activation of the Spec genes (Tomlinson et al., 1990).

The two LpS1 genes have diverged in sequence from the Spec genes. Sequence comparisons within the protein-coding regions show only weak similarity, mostly in the calcium binding domains (Xiang et al., 1988). In addition, the 5'-flanking and 3'-untranslated regions are totally dissimilar. Within the limits of the sequence analysis, LpS1α or LpS1β are no more closely related to one Spec gene than to another, suggesting that the LpS1 gene pair in L. pictus arose separately from the Spec gene family. We hypothesize that a common echinoid ancestor to S. purpuratus and L. pictus contained a single Spec-like gene. Most likely, this gene arose from a duplication event of an ancestral calmodulin gene, since calmodulin is the most ancient member of the superfamily yet shares a similar exon/intron structure with the Spec genes. This ancestral Spec gene somehow obtained or evolved the regulatory elements required for aboral ectoderm expression. Sometime after the branching of the Strongylocentrotus and Lytechinus genera, subsequent duplications produced the Spec gene family and LpS1 gene pair. However, an alternative view, difficult to rule out, is that the Spec and LpS1 genes did not have a shared ancestral Spec-like gene, but that each family arose and acquired the appropriate regulatory elements independently. Though independent acquisition of aboral ectoderm specificity would seem a less probable scenario, the
FIG. 9. Spatial expression of 368-lacZ in *L. pictus* embryos.
The embryos are oriented as in Fig. 7 except that panel C shows an embryo in cross-section. A and B, aboral ectoderm cells are stained. C, one secondary mesenchyme cell is stained. D, one primary mesenchyme cell is stained.

The fact that there is no more sequence similarity between LpS1 genes and Spec genes than to other members of the superfamily prevents this model from being ruled out (Moncrieff et al., 1990).

Another way of determining ancestral relationship among highly diverged genes is by comparisons of gene structures. Perret et al. (1988) analyzed many gene structures within the calmodulin superfamily and demonstrated that most intron placements within the protein encoding regions were conserved within a given group (Fig. 10A). Variable intron placements also occurred, and this variability can serve as an indicator of each lineage's relatedness. For example, there is a variable intron placement near or within the third calcium binding loop in the calmodulin, Spec, myosin light chain, and parvalbumin group. It can be seen in Fig. 10A that the Spec and LpS1 genes have the identical intron placement within the third loop, while other members of this group and other groups in the superfamily do not (Fig. 10A; Perret et al. (1988)). This suggests that the Spec and LpS1 genes are more closely related to each other than to other superfamily members, supporting the hypothesis that these genes shared a more recent ancestor.

The postulated ancestral Spec-like gene associated with the *Lytechinus* lineage must have undergone a series of duplications, presumably generated by unequal crossovers. Fig. 10B shows a model whereby the early LpS1 gene was internally duplicated, first duplicating exons 2, 3, 4, and 5, then exon 1. This gene was then externally duplicated to produce the extant LpS1 gene pair.

Differences between the *L. pictus* and *S. purpuratus* gene families are substantial: there is a high degree of sequence divergence, there are seven or eight Spec genes but only two LpS1 genes, and the LpS1 genes are internally duplicated.

These differences imply that there is little selective pressure being exerted on these genes in terms of sequence, copy number, or size. Nevertheless, all the genes have maintained the identical spatial expression pattern, and most of the proteins, as deduced from their sequences, have retained their ability to bind calcium.

The *LpS1* mRNAs and Proteins—The LpS1α and LpS1β genes could produce four distinct LpS1 proteins, if each gene initiated transcription both at exon 1 and exon 1'. The transcript generated from the more upstream promoter would always have exon 1' removed by splicing, since there is no acceptor splice site flanking exon 1'. That such an event happens is implied by the fact that a single LpS1 cDNA clone has been characterized that corresponds precisely to the LpS1α gene containing exon 1 but missing exon 1' (Xiang et al., 1988). We have made no attempts to find cDNA clones for the other predicted LpS1 mRNAs, but based on the similarity of 5'-flanking sequences and the expression data presented here with LpS1β 5'-flanking DNA, it is likely that the LpS1β gene is actively transcribed. Whether LpS1 messages exist beginning with exon 1' is unknown but certainly possible, since similar promoter elements could exist upstream of these exons. In an earlier study we showed the presence on two-dimensional gel electrophoresis of three LpS1 proteins (Xiang et al., 1988). However, we do not know the relationship between these proteins and the LpS1α and LpS1β genes.

Regulatory Regions of *LpS1* Genes—The results presented in this study demonstrate that 762 bp of 5'-flanking sequence plus 17 bp of 5'-untranslated leader sequence from the LpS1β gene are sufficient for proper aboral ectoderm expression. Our results also suggest the presence of a spatial repressor element within the region from −762 to −511 bp, since aboral ectoderm expression is partially lost when this region is deleted. Relevant to LpS1β gene expression are the studies of the aboral ectoderm-specific CycIIa actin gene of *S. purpuratus*. Davidson and co-workers (Hough-Evans et al., 1990) have shown that two negative spatial elements, P3A and P7II, control spatial expression of this gene. The consensus sequence for the P3A and P7II elements are ... C/T X C/T GCAC A/T ... and ... G/C T A/T ACCCT T/C , respectively (Thiebaut et al., 1990). Neither of these sites appear anywhere in the upstream region (base pairs −1 to −762) of the LpS1β gene, suggesting that other cis elements control LpS1 spatial expression.

In an earlier study we showed that 1516 bp of 5'-flanking sequence plus 18 bp of 5'-untranslated leader sequence from the Spec 2a gene were sufficient for proper expression, although similar regions from the Spec 1 and 2c genes were not sufficient. Relevant to the present study, the Spec 2a promoter region was capable of driving a reporter lacZ gene with appropriate aboral ectoderm specificity not only in *S. purpuratus* but also in *L. pictus* and *L. variegatus*. Thus, it seems reasonable to argue that, at least for the LpS1β and Spec 2a genes, there should exist common regulatory elements recognized by identical transcription factors. This argument is based partly on the belief that it is unlikely that cis elements required for aboral ectoderm-specific expression arose independently in these two related genes. However, sequence comparisons between LpS1β and Spec 2a show little to suggest common cis elements other than the USF binding site, whose function is unclear. The repetitive RSR element, which is a prominent feature of Spec 2a promoter region, is not present upstream of the LpS1 genes and is probably unique to *S. purpuratus*. Bandshift analysis with 5'-flanking DNA fragments from LpS1β and Spec 2a show many specific protein-DNA inter-

---

5 M. Xiang, L. Gan, and W. H. Klein, unpublished results.
actions, approximately 10–15 for each gene. In several instances the same proteins appear to be interacting with fragments from both genes. These latter proteins are likely candidates for transcription factors specifically involved in controlling the expression of the Spec and LpS1 genes in aboral ectoderm cells. However, only after many of the specific cis elements and trans factors are precisely defined will a picture emerge regarding the relatedness of the two gene promoter regions. Ultimately, comparisons of the regulatory features between the Spec 2a and LpS1β genes required for aboral ectoderm specificity should provide useful information on how cis regulatory elements on ancestrally related genes have changed over evolutionary time and how trans factors have evolved in response to these changes.

Acknowledgments—We thank our colleagues Robin Barrett, Lin Gan, Leah Goldberg, Mark Kozlowski, Robert Noiva, and Judith Venuti for their advice, technical assistance, and encouragement. We also thank San Juanita De La Cerda for secretarial assistance.

REFERENCES

Bruskin, A. M., Bedard, P. A., Tyner, A. L., Showman, R. M., Brandhorst, B. P., and Klein, W. H. (1982) Dev. Biol. 91, 312–324

Carpenter, C. D., Bruskin, A. M., Hardin, P. E., Keast, M. J., Anstrom, J., Tyner, A. L., Brandhorst, B. P., and Klein, W. H. (1984) Cell 36, 665–671

Demaine, J. G. (1982) in Calcium and Cell Function (Cheung, W.-Y., ed) Vol. II, pp. 111–144, Academic Press, New York

Flytzanis, C. N., Britten, R. J., and Davidson, E. H. (1987) Proc. Natl. Acad. Sci. U. S. A. 84, 151–155

Gan, L., Wessel, G. M., and Klein, W. H. (1990a) Dev. Biol. 142, 346–359

Gan, L., Zhang, W., and Klein, W. H. (1990b) Dev. Biol. 139, 186–196

Gorman, C. M., Moffat, L. F., and Howard, B. H. (1982) Mol. Cell. Biol. 2, 1044–1051

Hardin, P. E., and Klein, W. H. (1987) J. Mol. Evol. 25, 126–133

Hardin, S. H., Carpenter, C. D., Hardin, P. E., Bruskin, A. M., and Klein, W. H. (1985) J. Mol. Biol. 186, 243–255

Hardin, S. H., Keast, M. J., Hardin, P. E., and Klein, W. H. (1987) Biochemistry 26, 3518–3523

Hardin, P. E., Angerer, L. M., Hardin, S. H., Angerer, R. C., and Klein, W. H. (1988) J. Mol. Biol. 202, 417–431

Hough-Evans, B. R., Britten, R. J., and Davidson, E. H. (1988) Dev. Biol. 129, 198–206

Hough-Evans, B. R., Franks, R. R., Zeller, R. W., Britten, R. J., and Davidson, E. H. (1990) Development (Camb) 110, 41–50

Klein, W. H., Tomlinson, C. R., Zhang, W., Xiang, M., Kozlowski, M., Gan, L., Nguyen, T., and Wessel, G. M. (1990) in Developmental Biology (Davidson, E. H., Ruderman, J. V., and Posakony, J. W., eds) pp. 123–134, Wiley-Liss, New York

Klein, W. H., Xiang, M., and Wessel, G. M. (1991) in Novel Calcium-binding Proteins (Heizmann, C., ed) Springer-Verlag, Heidelberg, in press

Kretsinger, R. H., Moncrief, N. D., Goodman, M., and Czelusniak, J. (1988) in The Calcium Channel: Structure, Function and Implications (Morad, M., Naylor, W., Korda, S., and Schramm, M., ed) pp. 16–35, Springer-Verlag, Berlin

Lee, J. J., Shott, R. J., Rose, S. J., III, Thomas, T. L., Britten, R. J.,...
L. pictus LpS1 Gene Structure and Promoter Activity

and Davidson, E. H. (1984) J. Mol. Biol. 172, 149-176
Lynn, D. A., Angerer, L. M., Bruskin, A. M., and Klein, W. H. (1983) Proc. Natl. Acad. Sci. U. S. A. 80, 2659-2660
McMahon, A. P., Flytzanis, C. N., Hough-Evans, B. R., Katula, K. S., Britten, R. J., and Davidson, E. H. (1985) Dev. Biol. 108, 420-430
Moncrief, N. D., Kretzinger, R. H., and Goodman, M. (1990) J. Mol. Evol. 30, 522-562
Norman, A. W., Vanaman, T. C., and Means, A. R. (1987) Calcium-binding Proteins in Health and Disease, Academic Press, San Diego, CA
Perret, E., Lomri, N., Thomasset, M. (1988) J. Mol. Evol. 27, 351-364
Sambrook, J., Fritsch, E. F., and Maniatis, T. (1989) Molecular Cloning: A Laboratory Manual, pp. 650-659, 2nd Ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY
Sanger, F., Nicklen, S., and Coulson, A. R. (1977) Proc. Natl. Acad. Sci. U. S. A. 74, 5463-5467
Sawadogo, M., and Roeder, R. G. (1985) Cell 43, 165-175
Thiebaud, R., Goodstein, M., Calzone, F. J., Thézé, N., Britten, R. J., and Davidson, E. H. (1990) Genes & Dev. 4, 1999-2010
Tomlinson, C. R., and Klein, W. H. (1990) Mol. Reprod. Dev. 25, 328-338
Tomlinson, C. R., Kozlowski, M. T., and Klein, W. H. (1990) Development (Camb.) 110, 259-272
Xiang, M., Bédard, P. A., Wessel, G., Filion, M., Brandhorst, B. P., and Klein, W. H. (1988) J. Biol. Chem. 263, 17173-17180