CHARACTERIZATION AND DEVELOPMENTAL REGULATION OF A GENE EXPRESSED SPECIFICALLY IN THE SKELETOGENIC LINEAGE OF THE SEA URCHIN EMBRYO

Thesis by

Henry Michael Sucov

In Partial Fulfillment of the Requirements

for the Degree of

Doctor of Philosophy

California Institute of Technology

Pasadena, California

1989

(Submitted June 7, 1988)

ACKNOWLEDGEMENTS

It is my pleasure to thank all those who have made my time at Caltech what it has been, including my adviser, Eric Davidson, all the members of the Davidson lab both past and present, and everyone else in the Biology Department.

ABSTRACT

The sea urchin embryonic skeleton, or spicule, is deposited by mesenchymal progeny of four precursor cells, the micromeres, which are determined to the skeletogenic pathway by a process known as cytoplasmic localization. A gene encoding one of the major products of the skeletogenic mesenchyme, a prominent 50 kD protein of the spicule matrix, has been characterized in detail. cDNA clones were first isolated by antibody screening of a phage expression library, followed by isolation of homologous genomic clones. The gene, known as SM50, is single copy in the sea urchin genome, is divided into two exons of 213 and 1682 bp, and is expressed only in skeletogenic cells. Transcripts are first detectable at the 120 cell stage, shortly after the segregation of the skeletogenic precursors from the rest of the embryo. The SM50 open reading frame begins within the first exon, is 450 amino acids in length, and contains a loosely repeated 13 amino acid motif rich in acidic residues which accounts for 45% of the protein and which is possibly involved in interaction with the mineral phase of the spicule.

The important *cis*-acting regions of the SM50 gene necessary for proper regulation of expression were identified by gene transfer experiments. A 562 bp promoter fragment, containing 438 bp of 5' promoter sequence and 124 bp of the SM50 first exon (including the SM50 initiation codon), was both necessary and sufficient to direct high levels of expression of the bacterial chloramphenicol acetyltransferase (CAT) reporter gene specifically in the skeletogenic cells. Removal of promoter sequences between positions -2200 and -438, and of transcribed regions downstream of +124 (including the SM50 intron), had no effect on the spatial or transcriptional activity of the transgenes.

Regulatory proteins that interact with the SM50 promoter were identified by the gel retardation assay, using bulk embryo mesenchyme blastula stage nuclear proteins. Five protein binding sites were identified and mapped to various degrees of

resolution. *Two sites* are homologous, may be enhancer elements, and at least one is required for expression. Two additional sites are also present in the promoter of the aboral ectoderm specific cytoskeletal actin gene CyIIIa; one of these is a CCAAT element, the other a putative repressor element. The fifth site overlaps the binding site of the putative repressor and may function as a positive regulator by interfering with binding of the repressor. All of the proteins are detectable in nuclear extracts prepared from 64 cell stage embryos, a stage just before expression of SM50 is initiated, as well as from blastula and gastrula stage; the putative enhancer binding protein may be maternal as well.

TABLE OF CONTENTS

INTRODUCTION

The creation of differentiated tissues from a fertilized totipotent egg represents the central process of developmental biology. The differentiated state is described by a specific set of morphological characteristics, and is achieved by a specific pattern of gene expression unique for that tissue. An understanding of the mechanisms which result in the spatially confined expression of tissue specific genes provides a means to elucidate the processes which cause cells to become specified and committed to a specific program of differentiation.

Induction and Cytoplasmic Localization

Historically, embryologists have described two processes by which lineage determination is known to occur - induction and cytoplasmic localization. Induction is the process by which a cell is influenced in undertaking a choice of fate by the environment it finds itself in. The responsive cell may be influenced by physical contacts made with either other cells or extracellular matrices, or by the presentation of soluble morphogens, either growth factors or hormones (reviewed in Gurdon, 1987). By definition, cells which are determined by induction are not initially autonomous in their ability to differentiate, but have a period of plasticity during which the choice of fate can be specified externally. Induction can potentially occur at any time from the two cell stage through post-embryonic development.

In the evolution of multicellular organisms, and the concurrent specialization of groups of cells into tissues, the establishment of a means of communication between cells was no doubt necessary in order to regulate the proper allotment of cells into particular tissues. Cell interactions, then, are an ancient and fundamental aspect of metazoan cell specification. Every known animal system

utilizes induction to some extent, and in most organisms which have been studied as developmental systems it is the predominant means of cellular determination.

Cytoplasmic localization is exclusively an early acting means of determination, by which maternally derived morphogens (generally considered to be either protein or messenger RNA) synthesized during oogenesis are prelocalized in the unfertilized egg or newly fertilized zygote into particular regions of cytoplasm. As early cleavage proceeds to divide up the egg cytoplasm, the maternal localized determinants are inherited by particular blastomeres which are then commited to particular fates. Mechanistically, the determinants may be localized by attachment to some aspect of the cytoskeleton (see, for example, Strome and Wood, 1983). In contrast to induction, where information comes to the cell externally, the determinants of localization are placed within blastomeres. Almost every animal species studied as a developmental system makes use of cytoplasmic localization in the determination of some of its initial embryonic lineages (reviewed in detail in Davidson, 1986). Experimental demonstrations of localization are varied. In some cases, the fate of otherwise naive cells has been changed by the physical transfer of cytoplasm containing the localized determinants, demonstrating the ability of the transfered cytoplasm to confer a specific program of differentiation to the recipient cell (e.g., Illmensee and Mahowald, 1974; Whittaker, 1982). The explantation of cells which have received localized determinants, either to ectopic positions in the embryo or to in vitro culture, with no change in developmental fate, demonstrates that commitment has indeed occurred and that these cells are competent to carry out a given developmental program with no need for (and in spite of) the presence of other cells (e.g., Illmensee and Mahowald, 1974; Okazaki, 1975). The benchmark characteristic of cytoplasmic localization has been the ability of recipient cells to undergo

autonomous differentiation. The requirement for localized determinants can also be shown by the removal of cytoplasm containing the determinants, causing the affected cells to fail to undergo certain programs of differentiation (e.g., Verdonk, 1968).

In embryos which are known to utilize localization, the point of development at which the effect of localization is manifest is almost always in cleavage. Cytoplasmic localization can be considered as an evolutionary adaptation which is utilized to quickly create early embryonic lineages. This adaptation provides two related advantages which allows more rapid development: first, it quickly creates initial polarity in the embryo from which subsequent inductive interactions can be played out, and second, it allows the utilization of maternally synthesized determinants, rather than requiring zygotic products synthesized de novo.

One further process, called probabilistic determination, is an additional means of determination which occurs randomly. For example, any of five Drosophila neurogenic precursors has an equal probability of ultimately becoming a neuron. Once one cell becomes more predisposed to the neural pathway, however, that cell represses by cell-cell interactions the same pathway in its neighbors, which then become neural accessory cells (reviewed in Davidson, 1986).

As mentioned above, all animal species use induction in establishing cell fate to some degree, and many have adopted localization as an adaptation for early commitment. Conceivably, the combination of localization and induction is all that is needed for development to occur. That is, once early cellular commitments are established by localization, or by probabilistic processes, these then by induction cause the creation of further lineages, which interact to establish still further lineages, etc.

Cytoplasmic Localization and Induction in the Sea Urchin Embryo

The sea urchin has been one of the favored experimental subjects of developmental biologists ever since the beginning of the field of cellular embryology. Cultures of 10^7 - 10^8 synchronous embryos can be easily grown in the lab, and embryogenesis proceeds rapidly. The entirety of embryogenesis in so called regular sea urchins is directed towards the construction of a free-swimming, feeding larva, which for the species considered here, *S. purpuratus*, occurs in three days. There are only 1800 cells in the embryo at this stage. Most significantly for analysis of cellular determination, these are divided into a small number of tissues, including ectoderm, divided into oral and aboral ectoderm, a skeleton, a compartmentalized gut, muscle, and migratory mesenchymal cells. The following discussion will focus on two of these, the aboral ectoderm and the skeleton, as model tissues for understanding the processes which occur in early development in the specification and commitment of precursor blastomeres into these tissue types.

The sea urchin egg is endowed with a preformed animal-vegetal axis of symmetry established during oogenesis along which early development proceeds (Schroeder, 1980). After fertilization, a geometrically invariant process of cleavage begins. The first two cleavage planes are parallel to the animal-vegetal axis, and the third orthogonal, resulting in two tiers of four cells each. The fourth cleavage is vertical in the animal pole blastomeres and horizontal and asymmetric in the vegetal pole cells, resulting in eight intermediate-sized mesomeres at the animal pole, four macromeres below, and four small micromeres at the extreme vegetal pole. Further cleavages, and the filling of the interior intercellular space with fluid, ultimately result in a hollow blastula stage embryo, one cell thick, of approximately 200 cells. No migration of cells

occurs during this process of cleavage, which takes about 16 hr in *S. purpuratus*, so that all descendents of any given blastomere remain adjacent.

Early blastomere isolation and culture experiments by Horstadius (1939) were instrumental in providing our present understanding of cellular commitment and the generation of larval tissue types during early sea urchin development. Thus, it was well known that micromeres isolated at the time of their formation at the 16 cell stage and cultured in vitro would proceed to develop and differentiate on schedule, forming spicules often indistinguishable from their in vivo counterparts (Okazaki, 1975). Furthermore, micromeres transplanted to ectopic locations in the cleavage stage embryo still adhere to the skeletogenic pathway (Horstadius, 1939). The micromeres are thus recognized as a lineage determined by cytoplasmic localization. With the observation that a preformed animal-vegetal axis exists in the sea urchin egg, and that the micromeres (and all other cells of the embryo as well) are formed in a particular and invariant position with respect to this axis, a means, if not a mechanism, of localizing materials into the cytoplasm inherited by the micromeres is evident.

The aboral ectoderm is differentially populated by descendents of most of the mesomeres and macromeres, which separate from precursors after the third through fifth cleavage cycles. The oral-aboral axis, the second major axis along which development in the sea urchin occurs, is demonstrable by the eight cell stage (Czihak, 1963, reviewed in Davidson, 1986; Cameron et al., 1987), and may be present earlier. In contrast to the micromeres, the precursors of the aboral ectoderm, and for that matter all of the nonskeletogenic tissues, are not determined at the time of formation. Thus, various combinations of blastomeres show the potential of every cell of the 32 cell stage embryo to create embryonic structures it would normally not contribute to, with the exception of the micromere progeny (Davidson, 1986; Wilt, 1987). This behavior is characteristic

of *regulative embryogenesis*, in that reorganization of early blastomeres results in new assignments of cell fates by a process of cell-cell interactions. While it is not known when this ability to regulate is lost, the observation that programs of gene expression which typify the differentiated state of the aboral ectoderm begin at blastula stage (see below) argues that commitment has probably occurred for this tissue by this point. The period of cleavage, therefore, is of crucial importance in the determination of fates of the blastomeres.

Molecular Markers of Differentiated Sea Urchin Lineages

Genes expressed exclusively in the skeletogenic mesenchyme and the aboral ectoderm have been isolated and characterized. The mesenchyme specific gene, called SM50, encodes a 50 kD spicule matrix protein found during embryogenesis only in the skeleton (Benson et al., 1987; Chapter 1). The aboral ectoderm specific gene, called CyIIIa, encodes a cytoskeletal actin subtype found only in the aboral ectoderm (Cox et al., 1986). Expression of these tissue specific genes follows shortly after commitment of the respective tissue precursors. Thus, expression of SM50 is first detectable at the 120 cell stage, three cleavage cycles after the formation of the micromeres and only two cycles after the definitive segregation of the skeletogenic lineage at fifth cleavage (Benson et al., 1987; Chapter 1). CyIIIa transcripts are found maternally at a low level, but considerable quantities accumulate only after 10 hr of development, towards the end of cleavage (Shott et al., 1984). In both cases, expression of these genes is initiated before any overt morphological manifestation of tissue differentiation is evident. Thus, the onset of expression of these marker genes quickly follows the point of commitment of the precursor cells to their ultimate fates.

A Molecular Interpretation of Localization and Induction

In the sea urchin embryo, because there is no migration of blastomeres during cleavage, progeny of cells are in conjunction throughout the period when induction is causing determination of cell fates. Because cleavage is invariant in the sea urchin embryo, the fate of a cell can be predicted based on its position in the embryo, and a lineage map describing the fate of every cell of the embryo has been drawn (Davidson, 1986). However, there is a distinction between the fate and potential of the early blastomeres. The fate of a cell refers to its ultimate dispensation in the unperturbed embryo. The potential of the cell describes the breadth of development available to the cell given the proper inductive signal, such as by reorganization as described above. In the case of the micromeres, the fate and potential of these cells is the same, as these cells differentiate autonomously from the time of formation. On the other hand, given the appropriate signal, progeny of the other cell types of the cleavage stage embryo can take on wholly new fates, and in fact can become skeletogenic (described in detail in Chapter 3). In order to account for this potential throughout the embryo, no matter what the nature of the determinants which control commitment to the skeletogenic state of differentiation, a system of differential modification must be present which controls the activity of these determinants (as described in Chapter 3). The micromeres presumably inherit the appropriately modified factors through cytoplasmic localization; the other blastomeres receive the unmodified versions, but with the capacity to properly modify them given the appropriate signal. This description of localization and modification is probably a description of the general case. That is, localization probably also occurs in the animal pole blastomeres as well. If these cells are isolated at the eight or 16 cell stage and cultured in vitro, they will differentiate

into *ciliated* ectoderm. This default potential, or fate, can be considered the consequences of localized determinants, with cell-cell interactions required to ultimately produce the refined pattern of oral and aboral ectoderm seen in the pluteus stage larvae. In terms of localization, the difference between the animal pole cells and the micromeres is the inability of the latter to respond to induction. Comparable to the aboral ectoderm precursors of the sea urchin embryo, mesoderm formation in the Xenopus embryo has also been described as a combination of preliminary localization and subsequent induction (Gurdon et al., 1985). Localization therefore does not necessitate the irreversible commitment of recipient blastomeres to particular fates, although historically it has sometimes been narrowly described as such simply for lack of a better way to distinguish this process from induction. Localization is simply one example of the general strategy of the egg, which is to provide the materials needed for early development. These materials include polymerases, tubulins, and ribosomes, and almost certainly also include transcription factors, determinants, and receptors of induction. The maternal legacy therefore provides to the embryo the ability to specify certain lineages, both by localization and by the provision of the machinery needed for inductive interactions. Specification refers to the normal assignment of fate to a cell. Thus, the animal pole blastomeres are specified to become ectoderm, but still express the molecules necessary to receive and interpret signals from inductive interactions which can change this specification. Commitment can be considered as the inability of cells to respond to induction, either through absence of the receptors which mediate induction, loss of the enzymatic machinery necessary to modify determinants in response to induction, or turnover of the factors themselves.

It is well documented that DNA binding proteins regulate gene expression; it is reasonable that such proteins are also determinants of cell fate. The

appearance in sets of blastomeres of DNA binding protein determinants would result in the expression of specific sets of genes, activating the expression of molecules required for further differentiation and repressing the expression of determinants of other fates. In this thesis, I have taken a "bottom up" approach toward the nature of the determinants of the skeletogenic lineage, by analyzing the regulation of the SM50 gene, which is expressed exclusively in this tissue. Even if the proteins which regulate SM50 expression are not the ultimate determinants of the skeletogenic lineage, their activity is certainly a consequence of localization and determination. In Chapters 1 and 2, a complete characterization of the SM50 gene is described, starting with the initial cloning of the gene from cDNA and genomic libraries. Chapter 3 identifies functional domains of the SM50 promotor, as defined by gene transfer experiments, and an analysis of the regulatory proteins which interact with the SM50 promotor is described in Chapter 4. Whether this level of regulatory control is at an intermediate level in the hierarchy of the skeletogenic lineage or at the upper level of the lineage determinants themselves is at this point unknown, but an understanding of these regulatory processes will no doubt be mechanistically relevant to specification and commitment of embryonic blastomeres throughout the animal world.

REFERENCES

Benson, S., Sucov, H., Stephens, L., Davidson, E., and Wilt, F. A Lineage-specific gene encoding a major matrix protein of the sea urchin embryo spicule. I. Authentication of the cloned gene and its developmental expression. *Dev. Biol.* **120,** 499-506 (1987).

Cameron, R. A., Hough-Evans, B. R., Britten, R. J., and Davidson, E. H. Lineage and fate of each blastomere of the eight-cell sea urchin embryo. *Genes and Dev.* **1,** 75-84 (1987).

Cox, K. H., Angerer, L. M., Lee, J. J., Davidson, E. H., and Angerer, R. C. Cell lineage-specific programs of expression of multiple actin genes during sea urchin embryogenesis. *J. Mol. Biol.* **188,** 159-172 (1986).

Czihak, G. Entwicklungsphysiologische Untersuchungen an Echiniden. *Wilhelm Roux's Archiv. Dev. Biol. Entwicklungsmech. Org.* **154,** 272-292 (1963).

Davidson, E. H. *Gene Activity in Early Development,* 3rd ed., Academic Press (1986).

Gurdon, J. B. Mohun, T. J., Fairman, S. and Brennan, S. All components required for the eventual activation of muscle-specific actin genes are localized in the subequatorial region of an uncleaved amphibian egg. *Proc. Natl. Acad. Sci. USA* **82,** 139-143 (1985).

Gurdon, J. B. Embryonic induction-molecular prospects. *Development* **99,** 285-306 (1987).

Hörstadius, S. The mechanics of sea urchin development studied by operative methods. *Biol. Rev. Cambridge Philos. Soc.* **14,** 132 (1939).

Illmensee, K. and Mahowald, A. P. Transplantation of posterior pole plasm in *Drosophila.* Induction of germ cells at the anterior pole of the egg. *Proc. Natl. Acad. Sci. USA* **71,** 1016-1020 (1974).

Okazaki, K. Spicule formation by isolated micromeres of the sea urchin embryo. *Amer. Zool.* **15**, 567-581 (1975).

Schroeder, T. Expressions of the prefertilization polar axis in sea urchin eggs. *Dev. Biol.* **79**, 428-443 (1980).

Shott, R. J., Lee, J. J., Britten, R. J., and Davidson, E. H. Differential Expression of the actin gene family of *Strongylocentrotus purpuratus*. *Dev. Biol.* **101**, 295-306 (1984).

Strome, S. and Wood, W. B. Generation of asymmetry and segregation of germ line granules in early *C. elegans* embryos. *Cell* **35**, 15-23 (1983).

Verdonk, N. H. The effect of removing the polar lobe in centrifuged eggs of *Dentalium*. *J. Embryol. Exp. Morphol.* **19**, 33-51 (1968).

Whittaker, J. R. Muscle lineage cytoplasm can change the developmental expression in epidermal lineage cells of ascidian embryos. *Dev. Biol.* **93**, 463-471 (1982).

Wilt, F. H. Determination and morphogenesis in the sea urchin embryo. *Development* **100**, 559-575 (1987).

CHAPTER 1

A Lineage-Specific Gene Encoding a Major Matrix Protein of the Sea Urchin Embryo

II. Structure of the Gene and Derived Sequence of the Protein

DEVELOPMENTAL BIOLOGY 120, 507-519 (1987)

A Lineage-Specific Gene Encoding a Major Matrix Protein of the Sea Urchin Embryo Spicule

II. Structure of the Gene and Derived Sequence of the Protein

HENRY M. SUCOV,* STEVE BENSON,† JOHN J. ROBINSON,*,[1] ROY J. BRITTEN,*
FRED WILT,† AND ERIC H. DAVIDSON*

*Division of Biology, California Institute of Technology, Pasadena, California 91125, and
†Department of Zoology, University of California, Berkeley, California 94720

Received September 4, 1986; accepted in revised form November 6, 1986

A λgt11 cDNA clone isolated by use of a polyclonal antispicule matrix protein antiserum is shown in the accompanying paper [S. C. Benson, H. M. Sucov, L. Stephens, E. H. Davidson, and F. Wilt (1987) *Dev. Biol.* **120**, 499-506] to encode a prominent 50-kDa spicule matrix protein (SM50). This clone was used to select homologous genomic recombinants, and the structure of the gene was determined. The SM50 gene occurs once per haploid genome. It contains a single intron located within the 35th codon. A unique transcription initiation site 110 nucleotide pairs prior to the translation start signal was mapped by primer extension. The mRNA is 1895 nucleotides in length, excluding the 3′ poly(A) sequence, and contains a single open reading frame 450 codons in length. Though rare in whole embryo RNA the prevalence of the SM50 mRNA is calculated to be about 1% of the total mRNA in skeletogenic mesenchyme cells. The derived peptide sequence indicates a typical *N*-terminal signal peptide, and an *N*-linked glycosylation site near the *C* terminus. About 45% of the length of the protein is included in a domain composed of consecutive approximate repetitions of a 13-amino-acid element, the consensus sequence of which is

$$\text{Trp-Val-Gly-Asp-Asn-Gln-Ala-}\frac{\text{Leu}}{\text{Trp}}\text{-Val-}\frac{\text{Ile}}{\text{Asp}}\text{-Asn-}\frac{\text{Gln}}{\text{Pro}}\text{-}\frac{\text{Val}}{\text{Glu}}$$

The protein also contains an internal domain unusually rich in proline residues and a very basic *C*-terminal region.

INTRODUCTION

In the accompanying paper (Benson *et al.*, 1987) we verify the identity of a cDNA clone selected from an expression library screened with an antibody against sea urchin embryo spicule matrix protein. This clone represents the mRNA coding for a 50-kilodalton (kDa) spicule matrix protein (SM50), one of the major non-mineral components of the embryonic spicule. The *in situ* hybridization results reported by Benson *et al.* (1987) confirm that SM50 mRNA is detectable exclusively in the skeletogenic primary mesenchyme cells. In this paper we describe structural studies carried out on the SM50 gene and an analysis of the derived protein sequence. We have pursued these studies for two reasons. Knowledge of the organization and sequence of the gene is required to proceed with analysis of the mechanism of its lineage-specific regulation in the embryo. However, though it is of obvious importance, regulation of gene activity constitutes only one facet of the molecular processes leading to formation of a differentiated embryo. To proceed to an understanding of the cytotypic characteristics that appear in the various cell types of the embryo the functions of these cells must be described in terms of the specific molecules that they synthesize. One advantage offered by the sea urchin embryo is the opportunity to explore the linkages between specific programs of protein synthesis and embryonic cell function. As a major molecular constituent of a specialized structure, differentially synthesized in primary mesenchyme cells, the spicule matrix protein provides an excellent example. In the following we provide an initial characterization of this functionally dedicated gene product, as deduced from the sequence of the cloned cDNA.

[1] Present address: Department of Biochemistry, Memorial University of Newfoundland, St. John's, Newfoundland, Canada A1B 3X9.

MATERIALS AND METHODS

Construction and screening of the λgt11 cDNA library. Poly(A)$^+$ RNA was isolated from 62-hr early pluteus embryos as described by Shott *et al.* (1984). Blunt-ended double-stranded cDNA was synthesized by standard

0012-1606/87 $3.00

methods and the termini of the molecules were ligated to *Eco*RI linkers. The *Eco*RI-digested ligation mixture was size fractionated on a Bio-Gel A-50m (Bio-Rad) gel filtration column, which was equilibrated and run in 10 m*M* Tris, 1 m*M* EDTA, 0.4 *M* NaCl, pH 7.0. Fractions containing cDNA shorter than 500 nucleotide pairs (ntp) in length were discarded. The cDNA was ligated into the unique *Eco*RI site of λgt11 and the reaction mixture packaged using packaging extracts prepared as described by Hohn and Murray (1977). The library was amplified in *Escherichia coli* strain Y 1088 (Young and Davis, 1983).

Escherichia coli strain LE 392 was infected with recombinant phage and plated at a density of approximately 60,000 phage per 140-mm plate. After induction at 42°C for 15 min the plaques were allowed to develop at 37°C for 6 hr. Dry nitrocellulose filters were then placed on top of the lawns and the plates further incubated at 37°C overnight. The filters were blocked by shaking at room temperature (5 × 2 hr) in phosphate-buffered saline [20 m*M* phosphate, 0.15 *M* NaCl, 1% (w/v) bovine serum albumin, and 0.1% (v/v) Nonidet P-40]. The final blocking reaction was done at 4°C and the filters were then incubated for 12 hr with the purified IgG fraction of the antimatrix protein antiserum (Benson *et al.*, 1986), diluted 1:250 in blocking buffer and cleared of antibacterial antibodies. The filters were then washed (4 × 1 hr) in blocking buffer at 4°C and probed for 2 hr with ^{125}I-protein A diluted 1:1000 in blocking buffer. The filters were finally washed in blocking buffer (5 × 1 hr) at 4°C, air-dried, and set up for autoradiography.

Preparation of uniformly labeled double-strand probes. Single-stranded recombinant M13 DNA was used as template. Two micrograms of DNA was mixed with 10 ng of the 15-mer sequencing primer and hybridized according to standard sequencing techniques. Elongation was in 50 μl with 30 μCi of labeled deoxynucleotide (400 Ci/mmol), 1 nmole of the other unlabeled dNTPs, and 5–10 units of Klenow polymerase; 1 nmole of cold dXTP was added to chase. The reaction was terminated by heating at 68°C for 10 min. Water and restriction buffer were added, and the DNA was double-digested to release the labeled double-stranded insert. The insert was separated on an agarose gel, visualized with ethidium bromide, and isolated onto DE-81 paper (Whatman). The paper was crushed with a glass rod and the probe eluted with two sequential 350-μl washes of 1.5 *M* NaCl, 0.2 *M* NaOH, 50 μg/ml calf thymus DNA, which was neutralized by the addition of 150 μl of 3 *M* NaOAc, pH 5.5.

Genome blots and genomic clones. DNA was isolated from sperm of five individual adult sea urchins as described by Lee *et al.* (1984). Five micrograms of *Eco*RI–*Bam*HI double-digested DNA from each individual was electrophoretically separated, blotted onto nitrocellulose, and probed with an 800-ntp coding region cDNA fragment (the *Eco*RI–*Pst*I fragment of the clone pHS72; see Fig. 1) prepared as described above. Hybridizations were carried out at 65°C in 4× SET, 5× Denhart's solution, 0.1% SDS, 25 m*M* phosphate buffer (pH 6.8), 50 μg/ml calf thymus DNA, with final washes at 65°C in 0.3× SET, 25 m*M* phosphate buffer (pH 6.8), 0.1% SDS, 0.1% Na*PP*$_i$.To reuse, blots were heated to 95°C in 0.1× SSPE [1× = 0.18 *M* NaCl, 10 m*M* $NaPO_4$ (pH 7.7), 1 m*M* EDTA], 0.1% SDS, and allowed to cool slowly to room temperature. Dried, washed blots were exposed overnight before reuse to ensure that no signal remained. The genomic clones utilized for this work were derived from a partial *Mbo*I library (Calzone *et al.*, 1987) constructed in the vector EMBL 3 (Frischauf *et al.*, 1983). These clones were selected by screening with the insert of cDNA clone pHS31 (see Fig. 1), hybridized under the same criterion conditions as utilized for the genome blots.

Primer extension. A double-stranded, 180-ntp *Eco*RI fragment from the 5′ end of the cDNA clone pR524 (see Fig. 1) was labeled as described above. The purified fragment was digested with *Hpa*II and the digest displayed on a 10% nondenaturing acrylamide (19:1 acrylamide:bis) gel and stained with ethidium bromide. A 61-ntp fragment released from the 3′ end of the 180-ntp substrate was utilized as primer for the extension reaction (see Fig. 3 for sequence). The band containing this fragment was excised from the gel and macerated, and the fragment was eluted by diffusion in 0.5 *M* NH_4OAc, 1 m*M* EDTA at 37°C and precipitated with 4 μg carrier tRNA. Approximately 1.5 ng of the primer was mixed with the embryo RNAs designated in text, in 5 μl of 2 m*M* phosphate buffer, pH 6.8. The mixture was heat-denatured at 95°C for 3 min, and hybridized at 43°C for 8 hr in a total volume of 10 μl of 100 m*M* Tris (pH 8.6), 10 m*M* $MgCl_2$, 140 m*M* KCl, 20 m*M* β-mercaptoethanol (β*M*E), 1 m*M* phosphate buffer, pH 6.8. Forty microliters of this buffer, containing in addition 500 m*M* of each deoxynucleotide, 2 μg actinomycin D (Sigma, prepared in 80% EtOH at 0.8 mg/ml), and 30 units RNasin (Promega), was added and incubated with 25 units AMV reverse transcriptase (Life Sciences) at 43°C for 1 hr. The extension products were precipitated and run over a 40-cm 6% acrylamide (19:1 acrylamide:bis) urea denaturing gel.

S_1 mapping. The 700-ntp *Hin*fI fragment shown in Fig. 3 was used. The fragment was prepared from a subclone of the 2.3-kb *Eco*RI–*Sal*I fragment of genomic clone BG305 (Fig. 3) and gel purified, and the ends were filled in by the Klenow polymerase reaction. The fragment was blunt end ligated into the *Sma*I site of MP19, and its orientation determined by sequencing to be complementary to the RNA. A single-stranded, uniformly labeled probe was prepared as described above, except that

the complex was digested with *Bam*HI to terminate the insert at the 3′ end. The sample was then denatured by adding NaOH to 0.25 *M* and separated on a neutral agarose gel; the probe was isolated by binding to DE-81 paper as described above. It was then mixed with sea urchin embryo RNA, and the nucleic acids were precipitated and dissolved in 4 μl formamide. One microliter of 2 *M* NaCl, 5 m*M* EDTA, 0.2 *M* Pipes, pH 6.4 was added, and the mixture was denatured at 85°C for 5 min and hybridized overnight at 37°C. The solution was added to 320 μl of S_1 buffer (0.25 *M* NaCl, 30 m*M* KOAc, pH 4.5, 1 m*M* $ZnSO_4$, 5% glycerol) containing 5 μg calf thymus DNA, 5 μg tRNA, and 125 units of S_1 nuclease (BRL), digested at 16°C for 2 hr, and precipitated, before separation on a 40-cm 8% acrylamide (19:1 acrylamide:bis) urea denaturing gel.

RESULTS

Isolation of an Overlapping Set of λgt11 Clones and Estimation of mRNA Prevalence

The cDNA clone pHS72 utilized for the experiments described in the accompanying paper (Benson *et al.*, 1987) was one of a set of overlapping sequences isolated from a λgt11 expression cDNA library. As described under Materials and Methods this library was constructed from poly(A) RNA of 62-hr prism-stage embryos. At this stage the skeletal rods are lengthening and synthesis of spicule components would be expected to be occurring at a maximal rate. Benson *et al.* (1987) showed that in fact the SM50 mRNA is present at maximum levels after gastrulation. Screening was carried out as described, using a polyclonal antibody raised against spicule matrix protein (see Benson *et al.*, 1986, 1987). Approximately 145,000 primary phage recombinants were initially screened, and 11 distinct, independent reactive isolates were plaque-purified. The inserts of these phage ranged in size from 800 to 1600 ntp, and all cross-hybridized with one that was labeled by nick translation (data not shown). Restriction maps of two representative inserts, pHS31 and pHS72, spanning 1.7 kilobases (kb), are shown in Fig. 1. The maps of all recovered cDNA clones investigated overlap, confirming the hybridization result that these clones represent the same gene. The orientation of these cDNA clones was determined by hybridizing single-stranded probes to RNA gel blots (data not shown) and ultimately by sequencing. Also shown in Fig. 1 is another cDNA clone, designated pR524, that was isolated from the same library in a subsequent round of screening utilizing as probe a genomic DNA fragment. Clone pR524 extends the combined lengths of the overlapping cDNA sequence to 1.85 kb.

The prevalence of the mRNA coding for the SM50 protein should be indicated at least approximately by the

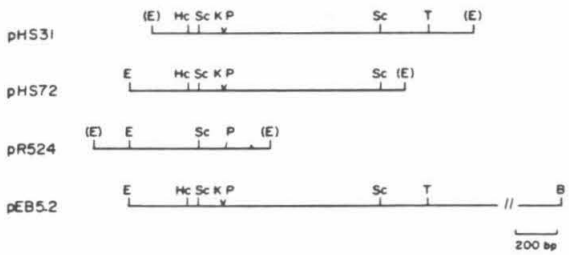


FIG. 1. Alignment of restriction maps of cDNA clones with cloned genomic region of the SM50 gene. pHS31 and pHS72 are cDNA clones isolated from the λgt11 library by antibody screening, pR524 is a cDNA clone isolated from the same library with a genomic DNA probe, and pEB5.2 is a subclone of the 5.2-kb *Eco*RI-*Bam*HI fragment from the genomic clone BG305 shown in Fig. 2b. The probe used to isolate clone pR524 was the 1.1-kb *Eco*RI fragment of BG305 (see Figs. 2 and 3). The clones are oriented so that the 5′ end (with respect to the SM50 mRNA) is at the left. The restriction sites are: E, *Eco*RI; Hc, *Hinc*II; Sc, *Sac*I; K, *Kpn*I; P, *Pst*I; T, *Taq*I; B, *Bam*HI. pR524 was not mapped for *Hinc*II and *Kpn*I sites. (E) designates an *Eco*RI site created during library construction by linker addition.

frequency with which homologous cDNA clones occur in the library. This frequency might be underestimated by the recovery of immunologically detected isolates, however, since there could exist domains of the fusion proteins that are not recognized by the polyclonal antibody, or that are nonreactive because of proximity to the β-galactosidase moiety. In the ideal case, one out of six cloned cDNA inserts should be in the correct orientation and reading frame to produce an immunologically reactive protein. To estimate the prevalence of SM50 coding sequences by nucleic acid homology we rescreened the λgt11 library, using the 750-ntp *Pst*I-*Sac*I fragment of pHS72 (see Fig. 1). As shown below, this fragment lies wholly within the open reading frame designating the SM50 protein. In three independent experiments the average ratio of the number of cDNA clones recovered with the DNA probe to the number recovered with the antibody probe was 6.6, i.e., within 10% of theoretical expectation. Thus the number of immunologically detectable clones is apparently not artifactually depressed. The approximate prevalence of SM50 clones in the 62-hr pluteus cDNA library is therefore about $6.6 \times 11/145{,}000$, or 0.05%. If the SM50 mRNA is cloned with an efficiency not significantly different from the average, about 1/2000 polyadenylated mRNA molecules in the whole pluteus-stage embryo code for this protein. Since various factors might influence the efficiency with which the SM50 message was cloned, this estimate is to be taken as indicating only the approximate prevalence of the mRNA.

At 62 hr the *Strongylocentrotus purpuratus* embryo consists of approximately 1500 cells, and of these about

60 are skeletogenic mesenchyme cells [see Davidson (1986) for review]. The prevalence of the SM50 mRNA in each such cell would thus approach 1% of total poly(A) RNA, or assuming about 10^5 mRNA molecules per cell, $\sim 10^3$ SM50 mRNA molecules per cell. This is about the same as the prevalence estimated for the muscle actin mRNA in pharyngeal muscle precursor cells of late-pluteus-stage embryos (Lee *et al.*, 1986), and several times the per-cell concentrations estimated for other lineage-specific sea urchin embryo mRNAs that have been investigated (e.g., Lynn *et al.*, 1983; Lee *et al.*, 1986).

Organization of the SM50 Gene

Genomic DNA from five adult males was digested with *Eco*RI and *Bam*HI, transfered to nitrocellulose, and hybridized to the 800-ntp *Eco*RI-*Pst*I fragment of cDNA clone pHS72. The result, shown in Fig. 2a, demonstrates that the gene encoding the SM50 protein is single copy. Densitometric analysis indicated that the intensities of the bands of individuals with two alleles (e.g., individuals P3, P5, P6, and P8) were all equal. The intensity of the diploid homozygous band of individual B is twice that of any single band of the other individuals.

A genomic library constructed from the DNA of individual B was screened with the insert of clone pHS31, which overlaps with the probe used in the genome blot experiment described above (Fig. 1). Four overlapping clones were isolated and mapped. These clones, represented in Fig. 2b, include 20.8 kb of continuous genomic sequence. As expected, each contains the 5.2-kb *Eco*RI-*Bam*HI fragment which hybridizes to the cDNA probe used in their isolation, and for which individual B is homozygous. This genomic fragment was subcloned and further mapped. Figure 1 shows that the restriction map of this subclone (designated pEB5.2) is colinear with that of the cDNA clones described above. The nucleotide sequences shown below confirm that the *Eco*RI site by which the insert in pHS72 is terminated is an authentic genomic site rather than a synthetic linker site.

The colinearity of cDNA and genomic sequences does not persist far upstream of the 5′ *Eco*RI site of pHS72. Thus when the 176-ntp *Eco*RI fragment of pR524 that extends to the 5′ side of this site was labeled and hybridized with a digest of the genomic clone BG305 (see Fig. 2), both the 1.1-kb *Eco*RI fragment and the 5′-terminal *Sal*I-*Eco*RI fragment of the insert are found to

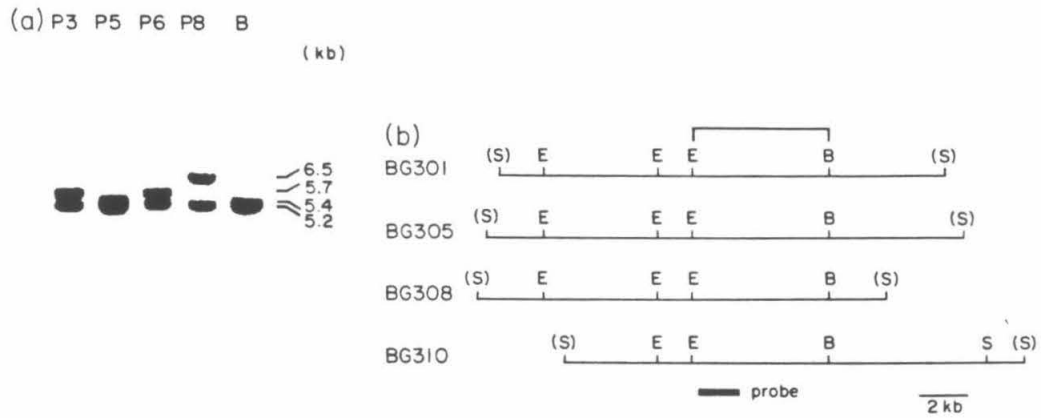


FIG. 2. Genome blot hybridizations and restriction maps of genomic isolates. (a) Genome blot. Five micrograms of sperm DNA from five adult *S. purpuratus* males was double-digested with *Eco*RI and *Bam*HI and blotted onto nitrocellulose. Hybridization was at 65°C in 4× SET hybridization buffer, using the 800-ntp *Eco*RI-*Pst*I fragment of cDNA clone pHS72 as probe. This region is colinear with the genomic sequence (see Fig. 1). The designations of the five individuals are at the top of the figure. (b) Genomic clones of the SM50 gene. DNA from individual B was used to construct a partial *Mbo*I genomic library in the phage vector EMBL3. The genomic clones were isolated by screening this library with the entire insert of pHS31, labeled by the Klenow reaction following digestion with *Eco*RI. These clones were mapped with E, *Eco*RI; B, *Bam*HI; S, *Sal*I. (S) designates a *Sal*I site derived from the EMBL3 polylinker that flanks the insert. The bracket indicates the 5.2-kb *Eco*RI-*Bam*HI fragment displayed in the genome blot of individual B shown in (a).

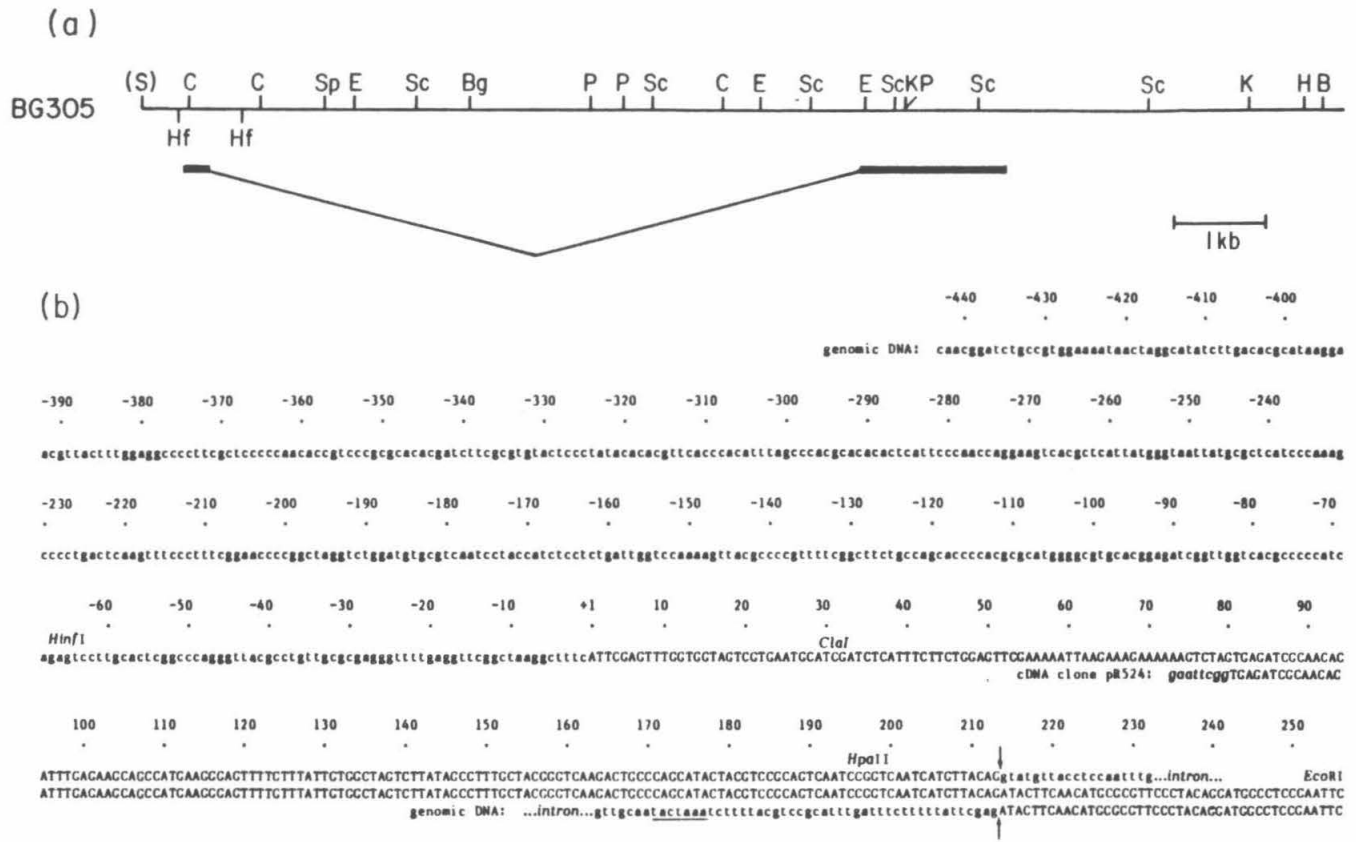


FIG. 3. Sequence of the 5′ region of the gene. (a) Shows 13 kb of cloned BG305 genomic DNA, which encompasses the entire SM50 gene transcription unit. Restriction sites are Hf, *Hin*fI; C, *Cla*I; Bg, *Bgl*II; Sp, *Sph*I; H, *Hin*dIII, and as described in the legends to Figs. 1 and 2. Beneath is shown the schematic organization of the gene. The solid bars represent the two exons, separated by a 7.1-kb intron. (b) The sequence of the genomic DNA is shown extending from 440 nt upstream of the cap site through the entire first exon. Exon sequences are in uppercase letters; 5′ flanking and intron sequences are in lowercase letters. Numbering begins with the transcription initiation site at position +1. The second line of sequence is that of the 5′ *Eco*RI fragment of cDNA clone pR524, which as described in the text overlaps the first and second exons. The linker *Eco*RI site at the 5′ end of this fragment is shown in italics. The genomic sequence at the 3′ end of the intron (lower case) and the sequence at the beginning of the second exon up to the indicated *Eco*RI site are shown on the third line. The splice site is designated by the vertical arrows, and the intron TACTAAA box is underlined.

react (data not shown). Since, as shown in the map of Fig. 2, these fragments are separated by several kilobases of DNA this observation suggested that the 5′ end of the SM50 gene includes a large intron. The sequence comparisons for this region shown in Fig. 3 confirm that the 176-ntp fragment of pR524 spans a splice site. Here it can be seen that the first 133 ntp of this fragment (i.e., excluding the linker *Eco*RI site) lies within the first exon, which terminates at the splice site at position 213, while the remaining 43 ntp belong to the second exon. The sequences at the 5′ and 3′ splice sites conform to consensus splice junction sequences (Mount, 1982), and 40 ntp prior to the 3′ end of the intron is the sequence TACTAAA. As this matches the concensus intron splice signal observed in sea urchin genes (Keller and Noon, 1984), it is probably the splicing intermediate branch point. The two exons are separated by 7.1 kb of intervening sequence. Since the cDNA and genomic sequences are identical, there is no possibility that "microexons" are located within this intron.

Transcriptional Initiation Site of the SM50 Gene

The position indicated for this site was determined by probe protection and primer extension procedures. Experimental data are presented in Fig. 4. A single strand probe consisting of the antisense complement of the 700-ntp *Hin*fI fragment that includes the first exon of the gene (Fig. 3) was utilized for an S_1 nuclease protection measurement. Figure 4a shows the result obtained on hybridizing this probe with total RNA from 6-hr cleavage stage embryos (lane b) or 40-hr gastrula stage embryos (lane c). The RNA gel blots in the accompanying paper (Benson *et al.*, 1987) demonstrate that there is no detectable SM50 message in 6-hr embryos, while at 40 hr the message has reached its maximal level of accumulation. A protected 213-ntp fragment is observed only after reaction with the 40-hr embryo RNA. This fragment, corresponding to the whole first exon, extends from the transcription start site to the splice junction (see Fig. 3). For the primer extension experiment shown in Fig. 4b the antisense strand of the 61-ntp *Hpa*II-*Eco*RI fragment of pR524 (Fig. 3) was labeled and hybridized with tRNA (lane b), or to 62-hr embryo poly(A) RNA (lane c), and incubated with reverse transcriptase. An extension product 255 nt in length is observed, corresponding to the distance between the *Eco*RI site at position 255 of the sequence shown in Fig. 3 and the transcription start site. Both the probe protection and the primer extension experiments yielded products of a single length, indicating a unique initiation site. The initiation site can be identified exactly because the only purine in the indicated vicinity is the adenosine residue at the position designated in Fig. 3 as +1. The experiments shown in Fig. 4 also preclude the possibility that there are additional introns in the 5′ region of the SM50 gene, and thus identify the boundary of the promoter region where transcription factor binding sites and other regulatory sequence elements might be sought.

The Derived Protein Sequence

Genomic and cDNA clones representing the two exons plus the several hundred nucleotides of the 5′ flanking region displayed in Fig. 3 were sequenced as diagrammed in Fig. 5a. Sequencing from the 3′ end of cDNA clone pHS31 (Fig. 1) revealed a long stretch of poly(A) adjacent to the linker-derived *Eco*RI site. The insert of this cDNA clone thus terminates at the 3′ end of the gene sequence. The processed mRNA is 1895 nt long, excluding the poly(A) tail, while the primary transcript is about 9000 nt in length. The mature message contains a single, long open reading frame of 450 codons, beginning with the initiation codon at position 110. Translation of animal mRNAs generally begins at the first AUG sequence encountered, though many exceptions are known (Kozak, 1984). We note that in the case of the SM50 mRNA there is an earlier, out-of-frame AUG codon at position 25, which is followed by a terminationsignal in that phase 12 codons later at position 61. The nucleotide sequence of the mature mRNA and the amino acid sequence of the derived protein are shown in Fig. 5b.

The *N*-terminal region of the derived protein sequence has the characteristics of a signal peptide. Thus immediately following the initiation methionine is a lysine residue, followed by a hydrophobic domain 13 amino acids in length. Signal peptide sequences are typically 15–25 amino acids long and characteristically include a basic residue near the *N* terminus of the protein and a central core of hydrophobic amino acids. They serve to direct the growing polypeptide chain into the endoplasmic reticulum for subsequent post-translational modification. The 50-kDa spicule matrix protein is an *N*-linked glycoprotein (Benson *et al.*, 1986), and it is believed that the spicule matrix proteins are transported to the site of spicule deposition via Golgi-derived vesicles (Gibbins *et al.*, 1969). Thus we expect to find a signal peptide sequence. A calculation to determine the most likely site of signal peptidase cleavage (von Heijne, 1983, 1984) indicated a preferred site after the alanine at residue 15. If this is correct, the mature protein is 435 amino acids long, with a molecular weight of 48.5 kDa. This and several additional features of the protein sequence shown in Fig. 5b are consistent with biochemically defined properties of the 50-kDa spicule matrix protein (Benson *et al.*, 1986). Thus the predicted molecular weight of 48.5 kDa after cleavage corresponds exactly with the size of the deglycosylated protein extracted from sea urchin

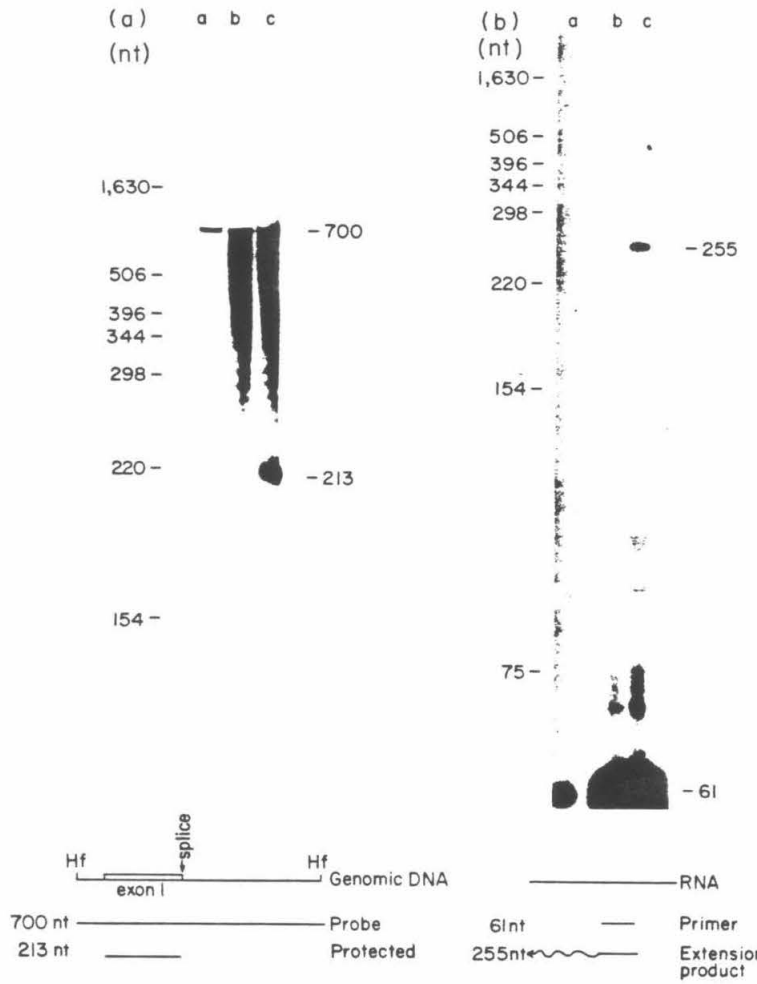


FIG. 4. Mapping the transcription initiation site of the SM50 gene. (a) S_1 nuclease protection experiment. The 700-ntp *Hin*fI fragment from the 2.3-kb *Eco*RI-*Sal*I region of BG305 (see Fig. 3) was subcloned into an M13 vector, and a single-strand labeled probe was prepared as described under Materials and Methods. Approximately 20 ng of this probe (2×10^5 cpm) was mixed with 10 μg tRNA (lane a); 10 μg tRNA plus 20 μg total 6-hr embryo RNA (lane b); or 10 μg tRNA plus 20 μg total 40-hr RNA (lane c). Hybridization and S1 digestion were carried out as described under Materials and Methods. The nuclease-resistant products were precipitated and separated on an 8% acrylamide gel using *Hin*fI-digested pBR322 labeled by the Klenow reaction as markers. (b) Primer extension experiment. The antisense strand of the 61-ntp *Eco*RI-*Hpa*II fragment of the cDNA clone pR524 (see Fig. 3) was labeled and isolated as described under Materials and Methods. This fragment spans the RNA splice site (see text). Approximately 1.5 ng (1×10^5 cpm) was hybridized with 5 μg tRNA (lane b) or 5 μg 62-hr embryo poly(A)$^+$ RNA (lane c), and extended with AMV reverse transcriptase. Products were precipitated and separated on a 6% acrylamide gel using labeled *Hin*fI-digested pBR322 DNA as markers. Lane a shows the primer prior to extension.

spicules, as measured on protein sizing gels. Second, an *N*-linked glycosylation site is present, four amino acids from the *C* terminus. Third, the p*I* of the derived sequence is 5.2 and the authentic SM50 protein is known to be acidic.

Some general features of the protein sequence are shown in Fig. 5c, in which the distribution of hydrophobic and hydrophilic amino acids is presented graphically. Following the presumed signal peptide sequence the protein displays an alternating pattern of hydrophobic

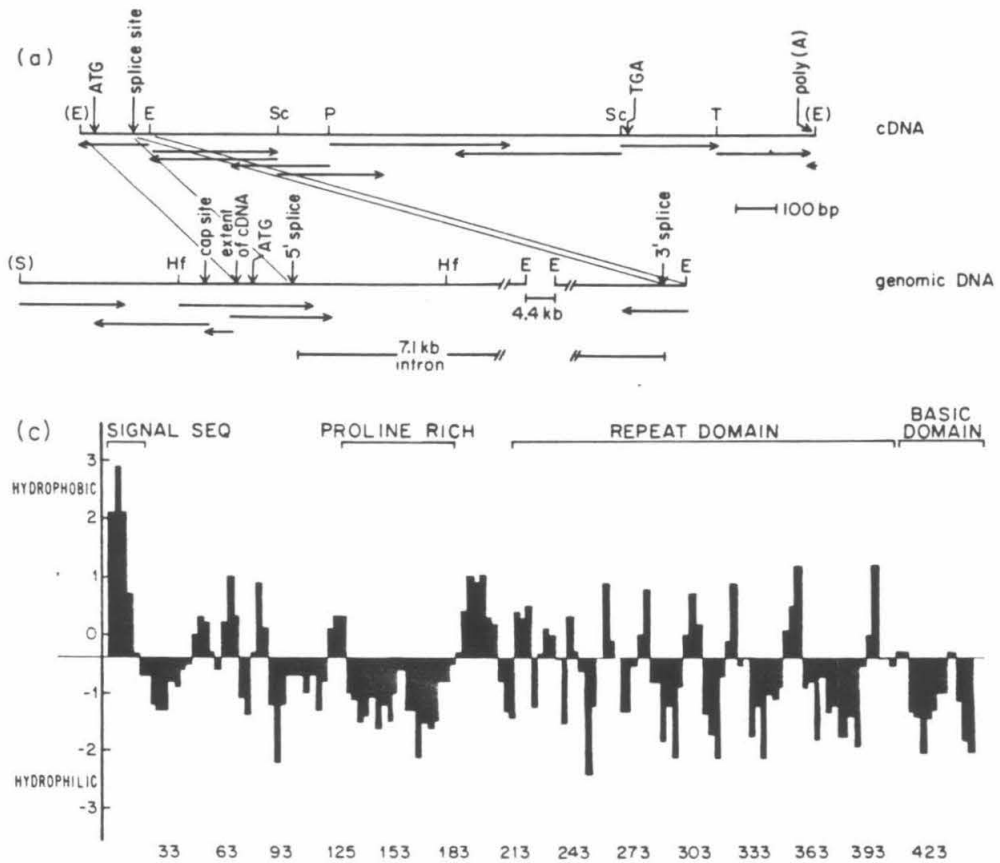


FIG. 5. Derived amino acid sequence of the SM50 protein. (a) Sequencing strategy. A composite cDNA map, compiled from the maps shown in Fig. 1 and representing all but the 5'-most 80 ntp of the transcribed sequence, is shown at the top. Below is shown the sequenced portions of the 2.3-kb *Sal*I-*Eco*RI and the 1.1-kb *Eco*RI genomic fragments, covering the first exon and part of the second exon, respectively (see Fig. 3). (b) Sequence of SM50 mRNA and derived sequence of the protein. The sequence shown begins with the transcription initiation site at position +1. Translation initiates at position 110 and the termination codon is at position 1460. The heavy arrow indicates the proposed signal cleavage site, and the light arrow, the location of the boundary between the two exons. The recognition sequence for *N*-linked glycosylation near the *C* terminus is boxed, and the poly(A) addition signal is underlined. The bracketed region is the internally repetitive domain described in the text. (c) Hydropathy plot and major domains of the protein. The analysis was carried out with the IBI-Pustell program, based on the algorithm of Doolittle.

and hydrophilic regions which is interrupted at amino acid 123 by a hydrophilic, moderately basic domain of unusually high proline content. Between amino acid 125 and amino acid 180 there occur 17 proline residues (31%), while in the remaining 381 amino acids following the signal peptide there occur a total of only 16 proline residues. The most striking feature of the mature protein sequence is a region of 193 amino acids extending from amino acid 214 to amino acid 407, which, except for a 12-amino-acid sequence located in the middle of the region, consists of inexact tandem repeats of a 13-amino-acid sequence. This sequence element is itself composed of alternating clusters of two to four hydrophobic and hydrophilic amino acid residues, and it accounts for the periodic alternation in the hydrophilicity plot shown in Fig. 5c. The similar alternation seen in the *N*-terminal region of the protein is of different origin, as there is no homology evident between this region and the internally

(b)

```
  1 ATTCGAGTTTGGTGGTAGTCGTGAATGCATCGATCTCATTTCTTCTGGAGTTCGAAAAATTAAGAAAGAAAAAAGTCT

 79 AGTGAGATCGCAACACATTTGAGAAGCAGCC ATG AAG GGA GTT TTG TTT ATT GTG GCT AGT CTT ATA
                                  Met Lys Gly Val Leu Phe Ile Val Ala Ser Leu Ile

146 GCC TTT GCT ACG GGT CAA GAC TGC CCA GCA TAC TAC GTC CGC AGT CAA TCC GGT CAA TCA
 13 Ala Phe Ala Thr Gly Gln Asp Cys Pro Ala Tyr Tyr Val Arg Ser Gln Ser Gly Gln Ser
                                                                  EcoRI
206 TGT TAC AGA TAC TTC AAC ATG CGC GTT CCC TAC AGG ATG GCC TCC GAA TTC TGT GAA ATG
 33 Cys Tyr Arg Tyr Phe Asn Met Arg Val Pro Tyr Arg Met Ala Ser Glu Phe Cys Glu Met

266 GTT ACA CCT TGT GGA AAT GGA CCA GCA AAA ATG GGT GCT CTG GCT TCA GTT TCC TCG CCT
 53 Val Thr Pro Cys Gly Asn Gly Pro Ala Lys Met Gly Ala Leu Ala Ser Val Ser Ser Pro

326 CAG GAG AAC ATG GAA ATC TAC CAA TTG GTC GCC GGC TTC TCA CAG GAC AAC CAG ATG GAG
 73 Gln Glu Asn Met Glu Ile Tyr Gln Leu Val Ala Gly Phe Ser Gln Asp Asn Gln Met Glu

386 AAT GAG GTT TGG CTC GGA TGG AAC AGT CAG AGC CCA TTC TTT TGG GAA GAT GGC ACC CCA
 93 Asn Glu Val Trp Leu Gly Trp Asn Ser Gln Ser Pro Phe Phe Trp Glu Asp Gly Thr Pro

446 GCT TAT CCT AAC GGA TTC GCC GCT TTC TCC AGC AGT CCG GCG TCT CCT CCA CGT CCC GGT
113 Ala Tyr Pro Asn Gly Phe Ala Ala Phe Ser Ser Ser Pro Ala Ser Pro Pro Arg Pro Gly
                                  HincII
506 ATG CCC CCA ACT CGT AGC TGG CCC GTC AAC CCT CAG AAC CCC ATG TCA GGA CCA CCA GGA
133 Met Pro Pro Thr Arg Ser Trp Pro Val Asn Pro Gln Asn Pro Met Ser Gly Pro Pro Gly
    SacI
566 AGA GCT CCA GTC ATG AAG CGT CAG AAC CCA CCT GTC CGA CCT GGA CAA GGA GGA CGG CAG
153 Arg Ala Pro Val Met Lys Arg Gln Asn Pro Pro Val Arg Pro Gly Gln Gly Gly Arg Gln

626 ATC CCT CAG GGT GTG GGA CCA CAG TGG GAA GCG GTA GAG GTC ACC GCC ATG CGC GCT TTC
173 Ile Pro Gln Gly Val Gly Pro Gln Trp Glu Ala Val Glu Val Thr Ala Met Arg Ala Phe
        KpnI     PstI
686 GTT TGT CAG GTA CCT GCA GGA GGA ACA TCC CTA TCG GCC AAC AAC CGG GCA TGG GAC AAG
193 Val Cys Glu Val Pro Ala Gly Gly Thr Ser Leu Ser Ala Asn Asn Arg Ala Trp Asp Lys

746 GCG GCT TTG GTA ATC AAC AAC CAG GCA TGG GTG GGC GAC AAC CAG GCT TTG GTA ATC AAC
213 Ala Ala Leu Val Ile Asn Asn Gln Ala Trp Val Gly Asp Asn Gln Ala Leu Val Ile Asn

806 CAG GAA TGG GTC GGC GAC AAC CAG GCT TTG GTA ATC AAC CAG GAA TGG GAG GGC GAC AAC
233 Gln Glu Trp Val Gly Asp Asn Gln Ala Leu Val Ile Asn Gln Glu Trp Glu Gly Asp Asn

866 CAG GCT GGG GTA ATC AAC CCG GTG TGG GTG GGC GAC AAC CAG GCA TGG GTG GAC AAC AAC
253 Gln Ala Gly Val Ile Asn Pro Val Trp Val Gly Asp Asn Gln Ala Trp Val Asp Asn Asn

926 CAG GCT GGG GTA ATC AAC CCG GTG TGG GTG GAC GAC AAC CAG GCA TGG GTG GAC AAC CAG
273 Gln Ala Gly Val Ile Asn Pro Val Trp Val Asp Asp Asn Gln Ala Trp Val Asp Asn Gln

986 GAG TGG GCG GGC GAC AAC CAG GCT TTG GTA ATC AAC CCG GCA TGG TGC ACA ACA ACC AGG
293 Glu Trp Ala Gly Asp Asn Gln Ala Leu Val Ile Asn Pro Ala Trp Trp Thr Thr Thr Arg

1046 CAT GGT GGA CAA CAA CCA GGC TGG GTA ATC AAC CCG GTG TGG GTG GGC GAC AAC CAG GCA
313 His Gly Gly Gln Gln Pro Gly Trp Val Ile Asn Pro Val Trp Val Gly Asp Asn Gln Ala

1106 TGG GTG GAC AAC CAG GAG TGG GCG GGC GAC AAC CAG GTG TGG GTG GAC GAC AAC CAG GCT
333 Trp Val Asp Asn Gln Glu Trp Ala Gly Asp Asn Gln Val Trp Val Asp Asp Asn Gln Ala

1166 TTG GTA ATC AGC CAG GTG TGG GTG GAC GAC AAC CAG GCA TGG GTG GAC AAC AAC CAG GTA
353 Leu Val Ile Ser Gln Val Trp Val Asp Asp Asn Gln Ala Trp Val Asp Asn Asn Gln Val

1226 TGG GTG GAC AAC CAG GAG TGC GCC GGC GAC AAC CAG GTA TGG GAC GGC GAC AAC CAG GCT
373 Trp Val Asp Asn Gln Glu Trp Ala Gly Asp Asn Gln Val Trp Glu Gly Asp Asn Gln Ala

1286 TTG GTA ATC AAC CAG GTG TGG GTG GGC GAC AAC CAG GCA TGG GTG GCC AGC AAC CGA ATA
393 Leu Val Ile Asn Gln Val Trp Val Gly Asp Asn Gln Ala Trp Val Ala Ser Asn Arg Ile

1346 ACC CGA ATA ACC CGA ACC CGA ACA ACC CGA ACA ACC CGA ATA ACC CAA ACC CCA GGT TCA
413 Thr Arg Ile Thr Arg Thr Arg Thr Thr Arg Thr Thr Arg Ile Thr Gln Thr Pro Gly Ser
                                                          SacI
1406 ACA GAC CCC GTA TGC TTC AAG AGG CAC ACC CGT TGG CAT AAC TTG AGC TCA GAC TGA ACT
433 Thr Asp Pro Val Cys Phe Lys Arg Gln Thr Arg Trp His Asn Leu Ser Ser Asp ---

1466 TACTATTTTCTTTGTTGTTATAGACTTGAAAAGTTTTGCCTCGGCGTTTTTTTTTTCTTCTTCCTTCCATACTTTGAGG

1545 AATTACCGCAACCGCAAGTGTTATGTAGAGCCCCCTCAACGCAGAGTTCATTCGGACAAAAGCCTTTCTTGACATTGCTT
                                                    TaqI
1624 CAAAATGTTGTTCTGAAAGAAGGGATGAGACTGATACTTCTACAGCTTCGCTCGAACAATTCGTTTAGTTTCAAGTTTA

1703 ACAGGATATAATATAAACTGGTCTTCCACTTTTCCCACAGTCATATTTTTGTTTTCTCGTTTTAAATTCAATACGTTTA

1782 CTTTTTTTTATAAAAAATAAAAGACAGTGGATGTAAACAAATTTGAAAGATATGTAAAGAATTAAAGTTCGTTTCACTT

1861 TAAATGTGGTAAAATAAAGATGTAATATATCACATAAAAAAAAAAAAAAAAAAAAAAAA
```

FIG. 5—*Continued.*

repetitive domain. In Fig. 6 the repeated sequence elements are shown in an aligned configuration. Many of the divergences from the consensus sequence are substitutions of similar amino acids, e.g., alanine for valine, or vice versa, in positions 2 and 7, and glutamine for asparagine, or vice versa, in positions 5, 11, and 12. Furthermore, many of the deviations are the result of single base changes (e.g., Val [GTG] → Glu [GAG] and Ala [GCG] in position 2). The codon usage within the repeated domain is compared in Table 1 with that elsewhere in the protein. Here it can be seen that a unique pattern of codon choice prevails throughout the repeated domain, with respect to the remainder of the sequence. For example, 65% of valine codons within the domain are GTG, while elsewhere this codon is utilized for valine only 18% of the time; the codon GGC is used for glycine within the domain 83% of the time and only 13% elsewhere; the codon ATC for isoleucine is used in 100% of occurrences within the domain and in only 29% elsewhere; etc. It is reasonable to suppose that the repeated domain evolved by tandem duplications of a precursor sequence coding for an oligopeptide similar to the consensus sequence shown in Fig. 6.

Toward the *C*-terminal end the character of the sequence changes abruptly. The final 41 amino acids constitute an extremely basic region, the p*I* of which is 12.7. This region is unusually rich in arginine (8 occurrences) and threonine (11 occurrences). The glycosylation signal Asn-Leu-Ser occurs at the end of the basic domain, just prior to the translation stop codon.

DISCUSSION

Structural Features of the Gene Coding for the SM50 Protein

Benson *et al.* (1986) demonstrated that the spicule matrix consists of at least 10 different proteins, of which the most prevalent is the 50-kDa species encoded by the gene described in this paper. The genes coding for the other proteins do not appear to be related to the gene coding for SM50 at the nucleic acid level. Thus the genome blot data shown in Fig. 2 indicate that there is only a single copy of the SM50 gene per haploid genome, and no related sequences can be discerned, even in low-criterion hybridizations (not shown). The SM50 gene appears unique, and is not a member of a family of related genes such as, for example, the *Spec* gene family, which codes for a set of Ca-binding proteins expressed in ectoderm cells during embryogenesis (Bruskin *et al.*,

214								Ala	Leu	Val	Ile	Asn	Asn	Gln	Ala
222	Trp	Val	Gly	Asp	Asn	Gln	Ala	Leu	Val	Ile	Asn	Gln	Glu		
235	Trp	Val	Gly	Asp	Asn	Gln	Ala	Leu	Val	Ile	Asn	Gln	Glu		
248	Trp	Glu	Gly	Asp	Asn	Gln	Ala	Gly	Val	Ile	Asn	Pro	Val		
261	Trp	Val	Gly	Asp	Asn	Gln	Ala	Trp	Val	Asp	Asn	Asn	Asn	Gln	
274								Ala	Gly	Val	Ile	Asn	Pro	Val	
281	Trp	Val	Asp	Asp	Asn	Gln	Ala	Trp	Val	Asp	Asn	Gln	Glu		
294	Trp	Ala	Gly	Asp	Asn	Gln	Ala	Leu	Val	Ile	Asn	Pro	Ala		
307	Trp	Trp	Thr	Thr	Thr	Arg	His	Gly	Gly	Gln	Gln	Pro	Gly		
320									Trp	Val	Ile	Asn	Pro	Val	
326	Trp	Val	Gly	Asp	Asn	Gln	Ala	Trp	Val	Asp	Asn	Gln	Glu		
339	Trp	Ala	Gly	Asp	Asn	Gln	Val	Trp	Val	Asp	Asp	Asn	Gln		
352								Ala	Leu	Val	Ile	Ser	Gln	Val	
359	Trp	Val	Asp	Asp	Asn	Gln	Ala	Trp	Val	Asp	Asn	Asn	Gln	Val	
373	Trp	Val	Asp	Asn	Gln	Glu	Trp	Ala	Gly	Asp	Asn	Gln	Val		
386	Trp	Glu	Gly	Asp	Asn	Gln	Ala	Leu	Val	Ile	Asn	Gln	Val		
399	Trp	Val	Gly	Asp	Asn	Gln	Ala	Trp	Val						
consensus:	Trp	Val	Gly	Asp	Asn	Gln	Ala	Leu/Trp	Val	Ile/Asp	Asn	Gln/Pro	Val/Glu		

FIG. 6. The internally repetitive domain of the SM50 protein. The bracketed region in Fig. 5b is aligned so as to demonstrate the repeated nature of the domain. Amino acids that deviate from the consensus shown at the bottom are shaded.

TABLE 1
CODON USAGE WITHIN THE REPETITIVE DOMAIN OF THE SM50 mRNA AND IN THE REMAINDER OF THE PROTEIN

Amino acid	No. codons within domain	Codon usage		Possible No. codons	No. amino acids in rest of protein	Occurrences of this codon	
		No.	%			No.	%
Val	31	20 GTG	65	4	17	3	18
		11 GTA	35			3	18
Asn	29	29 AAC	100	2	14	12	86
Gln	23	23 CAG	100	2	18	10	56
Asp	21	21 GAC	100	2	6	5	83
Ala	18	8 GCT	44	4	22	8	36
		7 GCA	39			4	18
		3 GCG	17			3	14
Gly	12	10 GGC	83	4	24	3	13
		2 GGG	17			0	0
Ile	9	9 ATC	100	3	7	2	29
Glu	7	5 GAG	71	2	10	5	50
		2 GAA	29			5	50
Leu	6	6 TTG	100	6	7	3	43
Pro	4	4 CCG	100	4	29	1	3
Total		160			154		

Note. Codon conservation in the repeat domain. The table shows codon usage in the repetitive domain, excluding the 13 amino acids in the center which show no relatedness to the rest of the region. All of the amino acids, with the exceptions of Trp and Ser, are compared in codon usage with the rest of the translated sequence.

1981; Lynn *et al.*, 1983), or the sea urchin metallothionein gene family (Nemer *et al.*, 1984, 1985).

The relatively high prevalence of the SM50 mRNA probably accounts for the fact that all of the 11 distinct recombinants isolated from the cDNA expression library represent the same transcript species, though determinants for all 10 spicule matrix species are present in the polyclonal antibody used for screening (Benson *et al.*, 1986). We estimated the prevalence of the SM50 transcript in 62-hr spicule-producing mesenchyme cells to be about 1% of the mRNA. An mRNA present at 10% of this level, which would still represent a fair mRNA prevalence for sea urchin embryo cells, might thus easily have escaped detection. Alternatively, it is possible that the SM50 protein is the most antigenic as well as the most prevalent in the spicule matrix, and antibodies against it could be dominant in the polyclonal antiserum used. Thus the screening dilutions we employed might not have permitted detection of clones coding for the minor species.

In internal structure the SM50 gene is not unusual. It contains typical intron processing sequences and the usual poly(A) addition site (Fig. 5b). A feature that appears common in sea urchin genes is the presence of an intron close to the transcription initiation site. For example, introns occur within the untranslated 5′-leader sequences of the *CyIIIa* actin gene (Akhurst *et al.*, 1987), the *CyI* actin gene (K. Katula and E. Davidson, unpublished data), just following the translation start codon in the *SpecI* gene (Hardin *et al.*, 1986), and toward the beginning of the coding region in the *bindin* gene (Gao *et al.*, 1986). The single intron of the SM50 gene occurs within the 35th codon, and begins 215 ntp from the transcription initiation site. However, the region upstream of the transcription initiation site lacks the most commonly occurring promoter sequence features (Fig. 3). Thus the TATA and CAAT sequence elements typically found at −25 and −40 with respect to the initiation site in most eukaryotic genes do not occur in the expected positions. In the SM50 gene the most proximal occurrence of a sequence similar to TATA (TAATTA) is at position −250, and that of the sequence CAAT is at position −175. In Fig. 4 we show evidence that locates the transcription start relatively far downstream from these sequence elements, and that we regard as a convincing demonstration of a unique initiation site. Though genes lacking proximal TATA sequences often initiate at several clustered locations, many exceptions are known. For example vertebrate U-type snRNA genes (Mattaj *et al.*, 1985), the mouse *HPRT* gene (Melton *et al.*, 1984), and the sea urchin *CyIIIa* actin gene (Ackhurst *et al.*, 1987) all initiate at precise sites in the absence of TATA elements. The promoter of the SM50 gene may belong to the class that instead utilizes GC tracts. The promoter

regions of the human *c-Ha-ras* gene (Ishii *et al.*, 1985), the mouse *HPRT* gene (Melton *et al.*, 1984), and the mouse HMG CoA reductase gene (Reynolds *et al.*, 1984), for example, all lack TATA and CAAT sequences, but they contain the Sp1 transcription factor binding sequence GGCGGG (Gidoni *et al.*, 1984). Figure 3 shows that elements consisting of five to seven GC residues occur periodically in the region proximal to the initiation site of the SM50 gene, at about −30, −50, −75, and −100. Direct experimental tests will be required to assess the functional significance of these or other sequence features of the presumed promoter region. One such feature that could be of interest is a heptanucleotide sequence, CGCTCAT, that occurs at position −265 and again at −244, with the closely related variants CGCGCAT at −112 and CACTCAT at −289. It will be interesting to determine whether this sequence occurs in the vicinity of other spicule matrix protein genes as well when these are isolated.

Transcription of the SM50 mRNA

The RNA gel blots included in the accompanying paper (Benson *et al.*, 1987) and others not shown demonstrated that the SM50 transcript is not detectable in the egg or early embryo, and that its prevalence rises sharply during the blastula stage. This gene thus belongs to the "late" class which includes about 10% of the total number of genes active in the sea urchin embryo [$\sim 10^3$ diverse sequences; reviewed by Davidson (1986)]. Densitometric analysis, taking as a reference the prevalence estimated from the frequency of occurrence of SM50 clones, indicates that the steady-state level of the transcript increases at least 100-fold during embryogenesis. This increase occurs while there is only a 2-fold increase in the number of skeletogenic mesenchyme cells, and is most likely regulated at the transcriptional level. From the initial relative rate of SM50 mRNA accumulation (Benson *et al.*, 1987; and unpublished data), we may conclude that the SM50 gene is, by the standards of the sea urchin embryo, a very actively transcribed, cell type-specific locus, and that it produces an mRNA of relatively high stability.

The SM50 protein

Figure 5c indicates that the SM50 protein comprises several dissimilar domains. Forty-five percent of the length of the mature SM50 protein is included in the internally repetitive domain analyzed in Fig. 6. Though its role is uncertain the regularity of this structure implies significance. One possibility is that the repetitive domain provides a calcite binding matrix. A current view of the biomineralization process is that matrix proteins, which are typically acidic, form a lattice-like microenvironment into which ions are introduced and crystallize. The function of the matrix proteins is to direct crystal growth in certain orientations and inhibit growth in others (Weiner, 1984). Though the repetitive domain of the SM50 protein displays no significant homologies with any other proteins included in the National Biomedical Foundation Protein Databank, some similarity exists between this domain and known calcium-binding domains of proteins such as troponin C and parvalbumin. These are 31 amino acids in length and contain a number of acid, amide, or alcohol groups (Asp, Asn, Glu, Gln, Ser, and Thr) at specific Ca^{2+} coordinating locations (Kretsinger, 1980). The consensus sequence of the SM50 protein repeat is much shorter, but it includes two potential coordination sites, in the sequences Asp-Asn-Gln and Asp-Asn-Gln-Glu. Perhaps these differences reflect the different structural requirements for binding crystalline $CaCO_3$ rather than ionic calcium. Alternatively, the repeating structure of this region of the SM50 protein could be utilized to order other matrix proteins that are the immediate calcite-binding elements.

A weak homology was identified between the proline-rich domain of the SM50 protein and a group of human salivary proteins (Wong and Bennick, 1980; Kauffman *et al.*, 1982) that include regions in which approximately one in three amino acid residues is proline. The proline-rich domain of the SM50 protein is identical with these at 21 of 55 amino acids, 13 of which are proline. This domain is not homologous with collagen. The proposed function of the salivary proteins is to regulate calcium levels in saliva by means of a proline-depleted *N* terminus that binds ionic calcium (Wong and Bennick, 1980), and the role of their proline-rich domains *per se* is unclear. Finally, the *C*-terminal basic domain of the SM50 protein is presumably involved in electrostatic interaction with negatively charged regions of other matrix proteins, and its position at one end suggests an anchorage or alignment function.

Calcite occurs widely in the animal kingdom, appearing as a major mineral component of spines and spicules in many invertebrate phyla (Lowenstam, 1981; Kingsley, 1984). Furthermore, calcite matrix proteins of a foraminiferan, a sea urchin, and a mollusk were demonstrated by Weiner (1984) to be remarkably similar in p*I* and amino acid composition, despite the enormous phyletic distance that separates these species. The SM50 protein is the first such protein for which the primary sequence is available. As others become known the essential conserved sequence features of this ancient group of proteins will emerge, and it will be interesting to consider in a comparative light the significance of the various domains that our analysis has revealed in this initial example.

This research was supported by NIH Grants HD-05753 (to E.H.D.) and HD-15043 (to F.W.). H.M.S. was supported by NIH Training Grant GM-07616, and J.J.R. was supported by a fellowship from Procter and Gamble.

REFERENCES

ADDADI, L., and WEINER, S. (1985). Interactions between acidic proteins and crystals: Stereochemical requirements in biomineralization. *Proc. Natl. Acad. Sci. USA* **82**, 4110–4114.

AKHURST, R. J., CALZONE, F. C., LEE, J. J., BRITTEN, R. J., and DAVIDSON, E. H. (1987). Structure and organization of the CyIII actin gene subfamily of the sea urchin, *Strongylocentrotus purpuratus*. *J. Mol. Biol.* **193**, in press.

BENSON, S. C., BENSON, N. C., and WILT, F. (1986). The organic matrix of the skeletal spicule of sea urchin embryos. *J. Cell Biol.* **102**, 1878–1886.

BENSON, S. C., SUCOV, H. M., STEPHENS, L., DAVIDSON, E. H., and WILT, F. (1987). A lineage-specific gene encoding a major matrix protein of the sea urchin embryo spicule. I. Authentication of the cloned gene and its developmental expression. *Dev. Biol.* **120**, 499–506.

BRUSKIN, A. M., BEDARD, P.-A., TYNER, A. L., SHOWMAN, R. M., and KLEIN, W. H. (1982). A family of proteins accumulating in ectoderm of sea urchin embryos specified by two related cDNA clones. *Dev. Biol.* **91**, 317–324.

BRUSKIN, A. M., TYNER, A. L., WELLS, D. E., SHOWMAN, R. M., and KLEIN, W. H. (1981). Accumulation in embryogenesis of five mRNAs enriched in the ectoderm of the sea urchin pluteus. *Dev. Biol.* **87**, 308–318.

CALZONE, F. J., LEE, J. J., LE, N., BRITTEN, R. J., and DAVIDSON, E. H. (1987). A novel family of maternal gene transcripts present in the egg and embryo of *Strongylocentrotus purpuratus*. I. The structure and expression of an interspersed maternal transcript. Submitted for publication.

DAVIDSON, E. H. (1986). "Gene Activity in Early Development," 3rd ed. Academic Press, Orlando, FL.

FRISCHAUF, A., LEHRACH, H., POUSTKA, A., and MURRAY, N. (1983). Lambda replacement vectors carrying polylinker sequences. *J. Mol. Biol.* **170**, 827–842.

GAO, B., KLEIN, L. E., BRITTEN, R. J., and DAVIDSON, E. H. (1986). Sequence of mRNA coding for bindin, a species-specific sea urchin sperm protein required for fertilization. *Proc. Natl. Acad. Sci. USA* **83**, 8634–8638.

GIBBINS, J. R., TILNEY, L. G., and PORTER, K. R. (1969). Microtubules in the formation and development of the primary mesenchyme in *Arbacia punctulata*. I. The distribution of microtubules. *J. Cell Biol.* **41**, 201–226.

GIDONI, D., DYNAN, W. S., and TJIAN, R. (1984). Multiple specific contacts between a mammalian transcription factor and its cognate promoters. *Nature (London)* **312**, 409–413.

HARDIN, H. S., CARPENTER, C. D., HARDIN, P. E., BRUSKIN, A. M., and KLEIN, W. H. (1986). The structure of the SpecI gene encoding a major calcium-binding protein in the embryonic ectoderm of the sea urchin, *Strongylocentrotus purpuratus*. *J. Mol. Biol.* **186**, 243–255.

HOHN, B., and MURRAY, K. (1977). Packaging recombinant DNA molecules into bacteriophage particles *in vitro*. *Proc. Natl. Acad. Sci. USA* **74**, 3259–3263.

ISHII, S., MERLINO, G. T., and PASTAN, I. (1985). Promoter region of the human Harvey *ras* proto-oncogene: Similarity to the EGF receptor proto-oncogene promoter. *Science* **230**, 1378–1381.

KAUFFMAN, D., WONG, R., BENNICK, A., and KELLER, P. (1982). Basic proline-rich proteins from human parotid saliva: Complete covalent structure of protein 1B-9 and partial structure of protein 1B-6, members of a polymorphic pair. *Biochemistry* **21**, 6558–6562.

KELLER, E. B., and NOON, W. A. (1984). Intron splicing: A conserved internal signal in introns of animal pre-mRNAs. *Proc. Natl. Acad. Sci. USA* **81**, 7417–7420.

KINGSLEY, R. J. (1984). Spicule formation in the invertebrates with special reference to the gorgonian *Leptogorgia virgulata*. *Amer. Zool.* **24**, 883–892.

KOZAK, M. (1984). Selection of initiation sites by eucaryotic ribosomes: Effect of inserting AUG triplets upstream from the coding sequence for prepro-insulin. *Nucleic Acids Res.* **12**, 3873–3893.

KRETSINGER, R. H. (1980). Structure and evolution of calcium-modulated proteins. *Crit. Rev. Biochem.* **8**, 119–174.

LEE, J. J., CALZONE, F. C., BRITTEN, R. J., ANGERER, R. C., and DAVIDSON, E. H. (1986). Activation of sea urchin actin genes during embryogenesis. Measurement of transcript accumulation from five different genes in *Strongylocentrotus purpuratus*. *J. Mol. Biol.* **188**, 173–183.

LEE, J. J., SHOTT, R. J., ROSE, S. J., THOMAS, T. L., BRITTEN, R. J., and DAVIDSON, E. H. (1984). Sea urchin actin gene subtypes: Gene number, linkage, and evolution. *J. Mol. Biol.* **172**, 149–176.

LOWENSTAM, H. A. (1981). Minerals formed by organisms. *Science* **211**, 1126–1131.

LYNN, D. A., ANGERER, L. M., BRUSKIN, A. M., KLEIN, W. H., and ANGERER, R. C. (1983). Localization of a family of mRNAs in a single cell type and its precursor in sea urchin embryos. *Proc. Natl. Acad. Sci. USA* **80**, 2656–2660.

MATTAJ, I. W., LIENHARD, S., JIRICNY, J., and DE ROBERTIS, E. M. (1985). An enhancer-like sequence within the *Xenopus* U2 gene promoter facilitates the formation of stable transcription complexes. *Nature (London)* **316**, 163–167.

MELTON, D. W., KONECKI, D. S., BRENNAND, J., and CASKEY, C. T. (1984). Structure, expression and mutation of the hypoxanthine phosphoribosyltransferase gene. *Proc. Natl. Acad. Sci. USA* **81**, 2147–2151.

MOUNT, S. M. (1982). A catalogue of splice junction sequences. *Nucleic Acids Res.* **10**, 459–472.

NEMER, M., TRAVAGLINI, E. C., RONDINELLI, E., and D'ALONZO, J. (1984). Developmental regulation, induction and embryonic tissue specificity of sea urchin metallothionein gene expression. *Dev. Biol.* **102**, 471–482.

NEMER, M., WILKINSON, D. G., and TRAVAGLINI, E. C. (1985). Primary differentiation and ectoderm-specific expression in the animalized sea urchin embryo. *Dev. Biol.* **109**, 418–427.

REYNOLDS, G. A., BASU, S. K., OSBORNE, T. F., CHIN, D. J., GIL, G., BROWN, M. S., GOLDSTEIN, J. L., and LUSKEY, K. L. (1984). HMG CoA reductase: A negatively regulated gene with unusual promoter and 5′ untranslated regions. *Cell* **38**, 275–285.

SHOTT, R. J., LEE, J. J., BRITTEN, R. J., and DAVIDSON, E. H. (1984). Differential expression of the actin gene family of *Strongylocentrotus purpuratus*. *Dev. Biol.* **101**, 295–306.

VON HEIJNE, G. (1983). Patterns of amino acids near signal-sequence cleavage sites. *Eur. J. Biochem.* **133**, 17–21.

VON HEIJNE, G. (1984). How signal sequences maintain cleavage specificity. *J. Mol. Biol.* **173**, 243–251.

WEINER, S. (1984). Organization of organic matrix components in mineralized tissues. *Amer. Zool.* **24**, 945–952.

WONG, R. S. C., and BENNICK, A. (1980). The primary structure of a salivary calcium-binding proline-rich phosphoprotein (protein C), a possible precursor of a related salivary protein A. *J. Biol. Chem.* **255**, 5943–5984.

YOUNG, R. A., and DAVIS, R. W. (1983). Yeast RNA polymerase II genes: Isolation with antibody probes. *Science* **222**, 778–782.

CHAPTER 2

A Regulatory Domain That Directs Lineage-Specific Expression of a Skeletal Matrix Protein Gene in the Sea Urchin Embryo

A regulatory domain that directs lineage-specific expression of a skeletal matrix protein gene in the sea urchin embryo

Henry M. Sucov, Barbara R. Hough-Evans, Roberta R. Franks, Roy J. Britten and Eric H. Davidson

Division of Biology, California Institute of Technology, Pasadena, CA 91125

Running title: Expression of a sea urchin skeletogenic transgene

Key words: Transgenic embryo; spicule; skeletal matrix; ***cis***-regulatory sequences; sea urchin

DNA sequences derived from the 5' region of a gene coding for the 50 kd skeletal matrix protein (SM50) of sea urchin embryo spicules were linked to the CAT reporter gene and injected into unfertilized eggs. CAT mRNA and enzyme were synthesized from these fusion constructs in embryos derived from these eggs, and *in situ* hybridization with a CAT antisense RNA probe demonstrated that expression is confined to skeletogenic mesenchyme cells. A mean of 5.5 of the 32 blastula stage skeletogenic mesenchyme cells displayed CAT mRNA (range 1-15), a result consistent with earlier measurements that indicate that incorporation of the exogenous injected DNA probably occurs in a single blastomere during early cleavage. *In vitro* mutagenesis and deletion experiments showed that CAT enzyme activity in the transgenic embryos is not significantly affected by removal of the SM50 signal peptide from the fusion product, but that the enzymatic activity of this product is enhanced 25-fold by decreasing the number of SM50 amino acids at the N-terminus of the fusion protein from 43 to 4. The *cis*-regulatory sequences that are necessary and sufficient to promote lineage-specific spatial expression in the embryo are located between -440 and +120 with respect to the transcriptional initiation site.

The skeletal elements, or spicules, of the sea urchin embryo are produced by cells of four equivalent lineages, that derive from the 4th cleavage micromeres (for review of the origin and differentiation of the skeletogenic mesenchyme see Davidson 1986, pp. 218-228; pp. 498-501). At 5th cleavage the founder cells of the skeletogenic lineages are separated from their sister cells by an asymmetrically positioned plane of cleavage, the orientation of which is normal to the primordial animal-vegetal axis of the egg. The smaller polar derivatives of this cleavage give rise to the 8 "small micromeres," the ultimate contribution of which remains obscure (however, see Pehrson and Cohen 1986), and the larger derivatives are the four ***clonal precursors*** of the skeletogenic mesenchyme (these are the VA*M*k, VO*M*k, and the two VL*M*k cells of the canonical sea urchin embryo lineage map; Davidson 1986, Fig. 4.5; Cameron et al. 1987). Three cleavage divisions ensue, giving rise to 32 descendants of the skeletogenic founder cells that all display the molecular and structural specializations required for spicule formation. At the swimming blastula stage the presumptive skeletogenic cells are located in a single cell thick disc at the vegetal pole of the embryo, surrounding the 8 small micromeres. In embryos of ***Strongylocentrotus purpuratus***, with which the work described herein was carried out, the skeletogenic precursors ingress into the blastocoel at 12-16 hr postfertilization (16°C). Ingression requires a suite of specific cell functions, including acquisition of motility, changes in cytoskeletal organization, and alteration of surface adhesive properties, and is correlated with numerous changes in protein synthesis pattern, as visualized by two-dimensional gel electrophoresis (Harkey and Whiteley 1983; Pittman and Ernst 1984). Once having assumed their free wandering mesenchymal habit, the cells proceed to explore the interior blastocoel wall, and then as gastrulation begins (~30 hr postfertilization), they coalesce bilaterally on the future oral side of the archenteron, and commence

the generation of the two triradiate spicules. Approximately one further division of these cells occurs during the gastrular period of development, the major phase of skeletogenic activity. During this phase the skeletogenic cells align themselves along the blastocoel wall and form syncytial columns, within which are secreted the elongated skeletal rods. These extend from the initial spicular foci at the oral side of the base of the archenteron toward the region where the mouth will form, and also posteriorly, in the aboral direction, ultimately forcing the ectoderm into the characteristic elongate triangular form of the echinoid pluteus larva. The exact shape of these rods, i.e., the angle at which the oral and aboral elements form, their length, and the brachiations they develop are species-specific characteristics, that depend on the positions taken up by the skeletogenic mesenchyme cells on the inner surface of the blastocoel. The mineral component of the skeletal structures is a 20:1 $CaCO_3$:$MgCO_3$ complex, which is deposited within a protein template secreted by the skeletogenic mesenchyme cells. Benson et al. (1986) solubilized the matrix and found that it is composed of about 10 different proteins, of which a major component is a 50 kd glycoprotein moiety. The mRNA and gene coding for this protein, which is known as SM50 (50 kd spicule matrix protein), have been cloned, sequenced, and characterized (Sucov et al. 1987; Benson et al. 1987).

As expected, the SM50 gene is expressed exclusively in skeletogenic mesenchyme (Benson et al. 1987), though its transcripts first become detectable many hours prior to the onset of skeletogenesis, about the time of ingression. However, at this early stage there are $\leq$10 SM50 mRNA molecules per skeletogenic mesenchyme cell, while in postgastrular stages, when these cells are executing their climax skeletogenic function, the prevalence of SM50 message approaches 200 molecules per cell (Killian and Wilt 1988). The SM50 gene can be regarded as a cardinal molecular marker for the specific and

particular process of differentiation undergone by the skeletogenic mesenchyme lineages. It is a member of a battery of similarly regulated genes, including those genes coding for the other spicule matrix proteins; a gene coding for a mesenchyme cell surface glycoprotein known as msp130, which may be required for normal migration and/or for Ca^{2+} uptake (Carson et al. 1985; Wessel and McClay 1985; Anstrom et al. 1987; Leaf et al. 1987); and genes coding for several unidentified mesenchyme-specific mRNAs and proteins (Harkey and Whiteley 1983; Harkey 1985; Harkey et al. 1988).

The differentiation of skeletogenic mesenchyme is from the beginning an autonomous process. Thus in culture isolated 4th cleavage micromeres display a succession of activities that correspond to the phases of their normal development *in situ*. They divide several times, acquire motility, and display the changes in adhesive properties expected from their behavior in the embryo. They then carry out skeletogenesis *in vitro* with good efficiency, though they tend to produce straight rather than brachiated skeletal rods (Okazaki 1975; Harkey and Whiteley 1980, 1983; McCarthy and Spiegel 1983; Carson et al. 1985). Many changes in synthesis of specific proteins that are correlated with the differentiation of these cells also take place *in vitro* in cultures originated with micromeres (Harkey and Whiteley 1983; Pittman and Ernst 1984). Most significantly for our present purpose, expression of the SM50 gene occurs autonomously in cultures of dissociated micromeres, and it has been shown rigorously by Stephens et al. (1988) that SM50 expression requires no intercellular contact. Thus blastomeres isolated at 2nd cleavage, i.e., prior to micromere formation, and maintained in culture under conditions in which no contact of the mitotic progeny is permitted, begin to produce SM50 mRNA on schedule. SM50 transcripts accumulate in these single cell cultures in an appropriate number of cells, and to within 50% of the normal level of expression

in vivo. It follows that the pattern of expression of this lineage-specific gene depends entirely on factors inherited spatially in the course of the geometrically predetermined sequence of cleavages by which the founder cells are segregated. The lack of requirement for intercellular inductive interaction for differentiation of the skeletogenic mesenchyme is unusual among known sea urchin embryo lineages, most of which require specific interactions in order for their normal fates to be realized (see Discussion for references). The micromeres themselves derive from an element of egg cytoplasm which is defined by its polar location with respect to the primordially organized animal-vegetal axis of the egg. A reasonable interpretation is that spatial regulation of genes such as SM50, that are expressed exclusively in the skeletogenic lineage, occurs by *cis-trans* interactions of localized maternal factors with regulatory regions of that gene, or of other genes which produce factors that have this function (Davidson and Britten 1971). In this paper we report initial steps toward the requisite experimental analysis. We have made use of a gene transfer system developed several years ago (McMahon et al. 1985; Flytzanis et al. 1985; Hough-Evans et al. 1987, 1988) to demonstrate correct spatial regulation of a fusion construct driven by the regulatory domain of the SM50 gene. These experiments show that sequences included within 440 nt of the transcription initiation site suffice to promote active expression that is confined to the skeletogenic mesenchyme cells.

Results

A fusion gene under control of SM50 regulatory sequences is expressed specifically in skeletogenic mesenchyme cells

In Figures 1(a) and (b) are displayed schematically the organization of the SM50 gene and of several constructs used in these experiments, in which the CAT

reporter gene was fused to various 5' regions of the SM50 gene sequence. The body of the SM50 gene consists of two exons, 213 and 1682 nt long respectively, separated by a 7.1 kb intron. Translation begins within the first exon at position +110 with respect to the transcription initiation site. The first 15 of the 34 amino acids encoded by the initial exon apparently specify a signal peptide, and the splice site divides the 35th codon. An initial series of studies was carried out with the construct designated SM50•CAT in Figure 1(b). This construct contains 2.2 kb of 5' flanking sequence from the SM50 gene, plus the entire first exon and intron, fused just within the second exon to the coding region of the vector pSV0CAT (Gorman et al. 1982) so as to maintain a continuous open reading frame. As shown in Figure 1(c) the predicted fusion protein is expected to contain 43 amino acids derived from the SM50 gene (including the 15 amino acid signal peptide), 14 amino acids from the vector sequence preceding the CAT coding region, and then the 231 residues of the CAT enzyme protein.

The SM50•CAT fusion was linearized at the unique *Bam*HI site, and microinjected into unfertilized *S. purpuratus* eggs. The embryos were harvested 48-50 hr after fertilization, at the early pluteus stage, for CAT enzyme assay. At this stage the skeletal rods are rapidly lengthening, and the endogenous SM50 RNA is present at maximum level (Benson et al. 1987; Killian and Wilt 1988). Table 1 demonstrates that CAT enzyme activity in the transgenic embryos was present at 2-3X the background level of the assay, measured in uninjected control embryos. The absolute level of CAT activity measured in these experiments is of course an arbitrary function of the number of embryos pooled for each sample, which was set at a relatively low level (25-40/sample) so as to minimize the number of injected embryos required. However, CAT activity would be expected to be low on a per embryo basis if the construct is expressed only in skeletogenic mesenchyme cells, as these constitute only ~6% of total cells in the pluteus stage embryo.

To determine what cell types in the transgenic embryos express the SM50•CAT fusion gene we carried out *in situ* hybridization using a CAT antisense ^{3}H-RNA probe (Hough-Evans et al. 1987, 1988). For this experiment the embryos were harvested at the early mesenchyme blastula stage, when skeletogenic mesenchyme cells are the sole occupants of the blastocoel, and are hence easy to identify. The embryo at this stage consists of approximately 500 cells, of which about 400 are ectodermal, 60 comprise the vegetal plate from which will later derive secondary mesenchyme, gut, muscle, and the larval rudiment, and 32 are cells belonging to the skeletogenic mesenchyme lineages (Davidson 1986; Fig. 4.5). Complete or nearly complete series of sections of 77 embryos derived from eggs which had been injected with SM50•CAT and which retained good morphology were examined. Of these 61 showed autoradiographic signal. For the purposes of these experiments, a positive signal was one in which five or more grains were clustered over any given cell. This level of signal is easily visualized over background, which on the average was less than 0.4 grains/cell, as observed over uninjected control embryo sections. With the exception of four questionable cases, labeling was clearly confined to skeletogenic mesenchyme cells already in the blastocoel, or to the polar region of the thickened vegetal wall from which the presumptive skeletogenic mesenchyme cells ingress. Representative examples of these *in situ* hybridizations are reproduced in Figure 2. Expression over cells of the polar ectoderm wall was expected, as the endogenous SM50 gene is active prior to ingression of the skeletogenic mesenchyme, as early as the 120 cell stage (Benson et al. 1987; Killian and Wilt 1988), and the embryos were collected while some were in the process of ingression. Of the 61 embryos which contained CAT mRNA, 30 displayed label only in the primary mesenchyme, 10 only in the polar vegetal region of the ectoderm, and 21 in both. The four embryos which showed

some low level labeling over cells that were possibly not of the correct lineage included two in which labeling was marginally above background over a single cell at what appeared to be the lateral side of the vegetal plate; a third in which what appeared to be a single animal pole ectodermal cell was labeled; and one embryo in which three cells were labeled, one each in three adjacent sections, at the lateral side of the vegetal plate and at the animal pole. The significance of these cases is doubtful, since occasional uninjected control embryo sections contained adventitious grain clusters that could have been counted as positive labeling over individual cells, and in addition it is possible to misidentify mesenchyme cells if they have applied themselves closely to the ectoderm wall or have been crushed against it in the course of sectioning. These examples notwithstanding, the results of the *in situ* hybridization experiments demonstrate clearly that the SM50 sequence included in SM50•CAT suffices to promote expression specifically in the skeletogenic mesenchyme lineages.

Mosaic expression of the exogenous fusion gene

DNA microinjected into unfertilized sea urchin egg cytoplasm is rapidly ligated into a concatenated form prior to first cleavage (McMahon et al. 1985). The DNA then enters the nuclear compartment, and in most embryos is stably incorporated into at least one blastomere during the first several cleavages (Hough-Evans et al. 1988). Incorporation appears to occur randomly with respect to cell lineage. Thereafter the exogenous sequences are bequeathed to all the progeny of the blastomere in which the incorporation event occurred, and it replicates at about the same rate as does the DNA of its host cell lineage (Franks et al. 1988; Hough-Evans et al. 1988). The mosaic incorporation pattern is ultimately manifest in a mosaic pattern of expression, the fraction of cells

within a given tissue which express the transgene depending upon the timing and location of the initial integration event.

In the 77 embryos examined by ***in situ*** hybridization in the experiment just described, the median number of ingressed skeletogenic mesenchyme cells clearly identified in the sections was 19. This value ranged from zero to 52, the latter probably representing embryos that had completed ingression, in which many mesenchyme cells had been sectioned and were represented in successive sections. In the following analysis, only ingressed mesenchyme cells were considered, as their progenitors are difficult to distinguish while they are still resident within the vegetal ectodermal wall. Mesenchyme cells within the blastocoel could in most cases easily be distinguished. Eleven embryos that retained five or less ingressed mesenchyme cells were excluded from the analysis. The remaining 66 embryos were individually examined to determine the fraction of ingressed mesenchyme cells that displayed label on hybridization with the CAT antisense ^{3}H-RNA probe. Results are presented in Figure 3(a), where for ease of comparison data for all embryos are normalized to 32 cells. Seventeen of the 66 embryos did not show label in any of their mesenchyme cells, either because they had integrated the exogenous DNA exclusively into nonskeletogenic lineages or had failed to incorporate it at all. Generally only a few mesenchyme cells were labeled in each embryo, although in some a significant fraction displayed autoradiographic grains. For labeled embryos the mean fraction of labeled mesenchyme cells was 5.5/32, or 17.2%. These numbers could be underestimated, as we do not know whether all the cells that expressed the transgene displayed significant autoradiographic label at the exposure used. Reference to the lineage from which the skeletogenic mesenchyme derives shows that incorporation of the DNA into a single ancestral blastomere of this lineage from the 4-cell to the 32-cell stage would result in labeling of 25% of the

mesenchyme cells, while integration into a single blastomere of this lineage at the 64-cell stage would result in 12.5% of the mesenchyme cells labeled. In the inset in Figure 3(a) the same data are classed in such a way that each bar represents embryos in which the DNA could have entered the skeletogenic mesenchyme lineage at a given cleavage, e.g., for the class labeled 6.3%, or 2/32 cells labeled, the implication would be that the DNA was incorporated at 7th cleavage; for the class labeled 12.5%, or 4/32, at 6th cleavage; and for the class labeled 25%, at any time between 2nd and 5th cleavage. However a further restriction is imposed by the high fraction of embryos displaying some label in mesenchyme cells, i.e., 49/66. A single random incorporation event anywhere in the embryo after the horizontal 3rd cleavage would result in no more than 50% embryos with any labeled mesenchyme cells, since only the lower 4 cells of the 8 cell embryo contribute to the skeletogenic mesenchyme. In addition, three embryos fall in the class labeled 50%, probably indicating incorporation at the 2-cell stage. Thus it is unlikely that the large 25% class in the inset in Figure 3(a) is due to incorporation as late as 5th cleavage. We may conclude that incorporation occurred in about 2/3 of the embryos (i.e., those falling in the classes labeled 12.5% and 25%) sometime between 2nd cleavage and 6th cleavage, with a few earlier and some later incorporations as well.

A similar analysis was carried out with embryos derived from eggs injected with a slightly different construct which displays the same spatial expression pattern as SM50•CAT (*viz.* the *Bgl*II-*Bam*HI fragment of 2.2PL•CAT; see Fig. 1 and below). Thirty-five of 37 embryos in this experiment displayed six or more ingressed mesenchyme cells, and the mean number of mesenchyme cells per embryo was 16. The distribution of labeled mesenchyme cells in this sample is shown in Figure 3(b). This distribution is obviously almost the same as that of Figure 3(a) (the mean of Fig. 3(b) is 6.4/32 cells *vs.* 5.5/32 cells for Fig. 3(a)). A

detailed analysis of incorporation frequency of DNA injected into *S. purpuratus* egg cytoplasm was recently carried out by Hough-Evans et al. (1988), in which a different construct was used, and the fraction of all embryo cells bearing the exogenous DNA was determined directly by DNA *in situ* hybridization. This study indicated that in about 3/4ths of the embryos the exogenous DNA could have been incorporated in a single 2nd, 3rd, or 4th cleavage blastomere. While incorporation into the skeletogenic mesenchyme lineage certainly occurs during these same stages, the largest class of labeled embryos in this study is that in which incorporation apparently occurred at 6th cleavage. Thus micromeres and their descendants may be slightly more refractory to incorporation than is the remainder of the embryo, or because they individually include several-fold less cytoplasmic volume than do the meso- or macromeres (Ernst et al. 1980) they may simply have a smaller chance of inheriting the exogenous DNA concatenate.

Deletion of SM50 amino acid sequence from the fusion gene product

As shown in Figure 1(c) the fusion product of the SM50•CAT construct contains 43 amino acids derived from the N-terminus of the SM50 protein, including the putative signal peptide. As described in the following section we measured the amount of CAT mRNA generated in embryos bearing the SM50•CAT fusion, and comparison with earlier studies carried out with a different CAT fusion (Flytzanis et al. 1987) indicated that the level of enzyme activity measured in the present experiments was over 10-fold lower than would be expected from the level of transcript observed. A possible explanation was that the signal peptide might lead to secretion into the endoplasmic reticulum, and/or modification and loss of the CAT enzyme activity. The sequence coding for the signal peptide was

thus deleted by oligonucleotide-directed mutagenesis, as indicated in Figure 1(d) and Materials and methods. The fusion protein coded by the deletion SM50Δ•CAT, now includes only 23 SM50 amino acids. However, as shown in Table 1 SM50Δ•CAT produces only about 2-fold more CAT enzyme activity than does the parental construct. Thus there is no evidence that the intracellular trafficking presumably mediated by the signal peptide significantly affects extractable CAT enzyme activity.

The SM50Δ•CAT construct contains a new *Sal*I restriction site at the point of the deletion, as shown in Figure 1(d). This site was fused in frame to the *Sal*I site immediately preceding the CAT coding sequence, to produce a third construct, called 2.2•CAT (Fig. 1(b)). This includes the same upstream SM50 gene sequences as does SM50•CAT, but codes for only four N-terminal SM50 amino acid residues. The remainder of the first exon, the large intron, and the fragment of the second exon included in the other constructs so far considered have been removed. For comparison, recall that the SM50•CAT fusion product contains 43 SM50 amino acids, or 28, if the 15 amino acid signal peptide is processed off *in vivo*. In addition the translation products of SM50•CAT, SM50Δ•CAT, and 2.2•CAT all include the same initial 13 amino acids that result from translation of vector sequences, as shown in Figure 1(c). A three-way comparison of the constructs indicates that 2.2•CAT generates 25-fold greater CAT activity per average embryo than does SM50•CAT (Table 1, experiment 1). In other experiments, 2.2•CAT produced 11-fold and 13-fold greater activity than did SM50•CAT (Table 1, experiments 2 and 3). As shown below, the SM50 regulatory sequences required for proper spatial regulation are retained in 2.2•CAT, and therefore the much higher level of CAT activity displayed by this construct is not a result of ectopic expression of the transgene in additional cell types of the embryo.

On the evidence thus far, this enhanced CAT enzyme activity could result either from removal of an intragenic site of transcriptional repression located within the intron or exon sequences deleted with the *SalI* fragment in the construction of 2.2•CAT, or from removal of an N-terminal protein sequence that in some way interferes with CAT enzyme function or accumulation. We show in the following that the latter is the correct explanation.

Probe excess titration of CAT mRNA in transgenic embryos

The amount of CAT mRNA generated in early pluteus stage embryos bearing SM50•CAT or 2.2•CAT was measured by single-strand probe excess titration, in order to assay directly the transcriptional productivity of these constructs. Two experiments were carried out in which unfertilized eggs were injected with SM50•CAT and 2.2•CAT, and samples were taken at 50 hr postfertilization for measurement of CAT enzyme activity, CAT DNA content, and CAT mRNA content. The CAT assays from these samples are those already discussed (experiments 2 and 3, Table 1). A measurement of CAT DNA content in the embryos of experiment 2 is shown in Figure 4(a), obtained by slot blot hybridization of DNA extracted from transgenic embryos, to a probe consisting of the sense strand of the CAT mRNA sequence. In this experiment the average embryo in the sample injected with SM50•CAT contained 8×10^4 molecules of the exogenous construct, and the corresponding average value for the 2.2•CAT sample was 1.8×10^4 exogenous DNA molecules per embryo. The difference may have been due to inadvertent introduction of more SM50•CAT DNA during injection. Previous studies have shown that for a given batch of eggs the amount of DNA retained and replicated during development depends mainly on the amount injected and not at all on the sequence, providing that the injected molecules are >1-2 kb in length (McMahon et al. 1985; Flytzanis et al. 1987; Livant et al. 1988).

The samples of transgenic embryos to be used for CAT mRNA measurement were collected together with a 13-fold excess of carrier embryos grown in parallel from uninjected eggs, and total RNA was extracted (see Materials and methods). To control for recovery and integrity of the extracted RNA, the quantity of CyIIIa cytoskeletal actin mRNA was determined, as the number of transcripts of this mRNA species had been measured by titration in *S. purpuratus* embryos in an earlier study (Lee et al. 1986). The CyIIIa titrations carried out on the SM50•CAT sample, the 2.2•CAT sample, and control uninjected embryos are shown in Figure 4(b), and the parallel CAT mRNA titrations in Figure 4(c). As shown in Figure 4(b), the SM50•CAT and 2.2•CAT samples contained almost equivalent concentrations of CyIIIa message, corresponding to 4.4×10^4 and 4.1×10^4 CyIIIa transcripts per embryo respectively, while the uninjected embryo control indicated 2.7×10^4 transcripts per embryo. Lee et al. (1986) found 5.8×10^4 CyIIIa transcripts per embryo at 36 hr and 8.8×10^4 at 65 hr. Interpolating, we would expect from these measurements about 7×10^4 CyIIIa transcripts per embryo at 50 hr. We assume the measurements of Lee et al. (1986) to be the more accurate since they were carried out on relatively large rather than microscale samples of embryos, and thus conclude that we recovered about 60% of the correct number of CyIIIa transcripts. However, since this factor would cause an underestimate in the actual amount of CAT mRNA in the experimental samples of less than a factor of two, and since the two experimental samples are in this respect indistinguishable, we have not normalized the CAT mRNA measurements shown in Figure 4(c) to a level that would assume 100% CyIIIa mRNA recovery. A direct conversion of the CAT mRNA titration shown to transcript prevalence (Lee et al., 1986) yields a value of 1.8×10^4 CAT mRNA molecules per embryo in the SM50•CAT sample, and 6.4×10^3 CAT mRNA molecules per embryo in the

2.2•CAT *sample*. As expected there is no CAT mRNA in the uninjected control, though this sample is undegraded, since it displays ample CyIIIa mRNA (Fig. 4(b)). The 2.8-fold greater amount of CAT mRNA in the SM50•CAT sample may be accounted for by the 4-fold greater average content of CAT DNA retained in the SM50•CAT embryos in this experiment, and certainly does not explain the 12-fold lower recovery of CAT enzyme activity. In any case, assuming that the two message species are translated with equal efficiency since they share the same 5' untranslated region and translation initation codon, then either the fusion protein derived from SM50•CAT is 34-fold less active than the enzyme derived from 2.2•CAT (i.e., 2.8×12, the average ratio of CAT enzyme activity per embryo in the 2.2•CAT sample to that in the SM50•CAT sample), or it is turned over in the cell 34 times more rapidly, or a combination of both. The extraneous N-terminal SM50 polypeptide sequence could be responsible for either phenomenon. However, it may seem a less than probable explanation that a portion of the *normal* SM50 protein sequence would result in an increase in turnover rate in cells that also express this protein endogenously. The more likely alternative is that the larger N-terminal SM50 sequence causes steric hindrance of the enzyme, resulting in >10X lower acetylation activity compared to the fusion protein produced by 2.2•CAT, which includes but four SM50 amino acids.

Thus a conclusion from these CAT mRNA measurements is that quantitatively significant negative transcriptional regulatory sites are unlikely to be present in the large intron of the SM50 gene. We have seen that if anything the SM50•CAT gene is more active than the 2.2•CAT gene, though the several-fold difference in transcript concentration is probably due simply to the subsaturation level of exogenous genes in this particular sample of 2.2•CAT embryos. An ancillary, though indirect argument to this effect is that in experiment 1 of

Table 1, for which there are neither CAT DNA nor CAT mRNA measurements, the ratio of CAT enzyme activity per embryo between the SM50•CAT and the 2.2•CAT constructs, i.e., 25X, is close to the apparent ratio of CAT enzyme activity per transcript, i.e., 34X. This is what would be expected were the exogenous DNA saturating in experiment 1, as is usually the case in sea urchin embryos raised from eggs injected in the cytoplasm by this method (Flytzanis et al. 1987; Livant et al. 1988; Lai et al. 1988), and were the two constructs exactly equal in transcriptional activity. In any case, the minimum qualitative conclusion is that these experiments provide no evidence for important regulatory sequences located in the intron or in the exon sequences deleted in 2.2•CAT. As we now show, the necessary and sufficient sequences can be identified in the flanking upstream region of the SM50 gene.

Subdivision of the upstream regulatory domain

The 2.2 kb *Sal*I fragment of 2.2•CAT was subcloned into a slightly different CAT vector known as pUC•PL•CAT (Bond-Matthews and Davidson 1988), to generate 2.2PL•CAT (Fig. 1(b)). This construct is identical to 2.2•CAT in the SM50 upstream and coding sequences it includes, and these SM50 sequences are fused to the parental CAT vector at the same ***Sal*I** restriction site. The translation product of 2.2PL•CAT includes an additional three amino acids derived from the slightly longer vector sequences preceding the CAT coding region. The rationale for this construct is the presence in 2.2PL•CAT of polylinker restriction sites at the 5' end of the SM50 sequence, permitting the entire insert of 2.2PL•CAT to be excised as a *Sph*I-*Bam*HI fragment (Fig. 1(b)). After gel purification the insert could then be injected without appended plasmid vector sequences. Several additional restriction fragments were likewise derived

from 2.2PL•CAT that include progressively less 5' SM50 promotor sequence, all of which terminate with the identical CAT-SV40 vector sequence (see Fig. 1(b)). Injection of fragments bordered at the 5' end by the *Sph*I site (at position -2200), the *Eco*RV site (at -1100), or the *Bgl*II site (at -440) resulted in a mild difference in CAT activity that was not statistically significant (Table 2). The *Bgl*II-*Bam*HI fragment was separately subcloned and again gel purified to preclude the possibility of contamination, again with no difference observed in CAT activity compared to the *Sph*I-*Bam*HI fragment (Table 2, experiment 2). Thus, the sequences upstream of the *Bgl*II site of the SM50 promotor are dispensable, as assayed by CAT activity at the early pluteus stage. Further deletions do result in significant decrease in CAT activity, confirming that important regulatory sequences required for SM50 function are found within the *Bgl*II site (data not shown).

In order to determine whether spatial regulation is maintained after deletion of the SM50 promotor to the *Bgl*II site, embryos injected with the *Bgl*II-*Bam*HI fragment were collected at mesenchyme blastula stage and prepared for *in situ* hybridization. Representative examples of sections are shown in Figure 5. Expression of this transgene is again limited to ingressed mesenchyme cells and to the polar region of the vegetal ectoderm wall, i.e., the identical pattern observed with SM50•CAT. Complete sections of 13 embryos were recovered, and partial series (at least three sections) of an additional 24 embryos were examined as well. Fourteen displayed no labeled cells. Eight embryos showed label in ingressed mesenchyme cells only, seven showed label over polar regions of the vegetal ectoderm wall only, and eight showed label over both. As seen in a few sections from embryos expressing SM50•CAT, several sections displayed label over regions of the embryo that were possibly not mesenchyme or vegetal mesenchyme precursors. In one case, two cells at the animal pole of the embryo

were labeled, and in a second case, a single cell at the lateral side of the vegetal plate was labeled. However, the vast majority of embryos clearly demonstrated the correct expression pattern. As noted above (Fig. 3(b)), the observed frequency of labeled cells in the mesenchyme population of these embryos was similar to that observed in embryos injected with SM50•CAT.

Embryos bearing the *Bgl*II-*Bam*HI fragment were also raised to early pluteus stage, and then collected for ***in situ*** hybridization. At this stage, the skeletogenic mesenchyme cells have for the most part distributed themselves along the length of the growing spicules, although small clusters of these cells persist at the ends of the spicule rods (Gustafson and Wolpert, 1963). The ectoderm and intestinal tract are close to fully differentiated, and migratory derivatives of the non-skeletogenic secondary mesenchyme are apparent within the blastocoel. In sectioned embryos that displayed CAT mRNA the label was again seen over mesenchyme cells, and was not present over gut or ectoderm cells. One example is shown in Figure 5(d-e). However, it is not possible to determine whether the labeled cells surround a spicule ***in vivo***, since the mineral elements of the skeleton are dissolved during fixation, and therefore it remains unknown whether the labeled cells were skeletogenic or secondary mesenchyme cells. It can be excluded, nonetheless, that widespread ectopic expression of the transgene occurs in highly differentiated embryos. The unambiguous results obtained with mesenchyme blastula stage embryos (Fig. 5(a-c)) provide the major conclusion from this initial dissection of the SM50 regulatory domain. This is that the SM50 sequences which lie between the *Bgl*II site at position -440 and the *Sal*I site at position +120 are both sufficient and necessary for the correct spatial activation of this gene.

Discussion

Regulated expression of the SM50•CAT fusion in transgenic embryos

This study demonstrates that the lineage-specific expression of a gene required for mesenchymal skeletogenesis in the embryo is mediated by *cis*-regulatory sequences within a few hundred nucleotides of the initiation site of the gene. There could be additional quantitative regulatory sites that we did not assay either upstream of -2200, or within the body of the gene that was replaced by the CAT reporter gene sequence in the fusion we studied. We have no way of directly comparing the transcriptional activity of the CAT fusions with that of the native SM50 gene. This is because there are multiple copies of the gene incorporated, due to the initial concatenation of the injected molecules, and the subsequent replication of the concatenates along with the host cell genomes (McMahon et al. 1985; Flytzanis et al. 1985; Hough-Evans et al. 1988; Franks et al. 1988). However, the amount of expression obtained from the SM50•CAT fusions is consistent with expectation were the number of exogenous genes saturating with respect to endogenous regulatory factors, according to results obtained in earlier studies with a CAT construct under the control of the CyIIIa actin gene, which is expressed exclusively in aboral ectoderm (CyIIIa•CAT; Flytzanis et al., 1982) CyIIIa•CAT fusion. The measurements shown in Figure 4(a) indicate about 8×10^4 molecules of the exogenous construct per embryo for the SM50•CAT sample, or $\sim10^5$ in those embryos that show SM50•CAT expression. The measurements of Figure 3 indicate that on the average these are located in 5.5/32 or 17.2% of the mesenchyme cells. Accepting this as an estimate for the whole embryo, in the ~1000 cell early pluteus there would be about 500 fusion genes per cell, *vs.* the two copies of the

native *SM50 gene* per cell (Sucov et al. 1987). For comparison, Livant et al. (1988) showed that *half*-saturation of the expression of a fusion construct under control of the CyIIIa regulatory sequence occurs at ~50-200 fusion genes per cell depending on the batch of eggs. The amount of CAT mRNA we measured in the present work is about 9-fold greater than the amount of endogenous SM50 mRNA per cell. Thus there would be about 1760 CAT mRNA molecules per active cell (1.8×10^4 CAT mRNA transcripts per embryo, divided by 60 skeletogenic mesenchyme cells × 0.17), compared with about 200 SM50 mRNA molecules per skeletogenic mesenchyme cell (Killian and Wilt 1988). The equivalent ratio for the CyIIIa•CAT fusion indicates that at saturation with exogenous genes about 20-fold more CAT mRNA is produced per aboral ectoderm cell bearing the CyIIIa•CAT fusion than endogenous CyIIIa mRNA (Flytzanis et al. 1987; Hough-Evans et al. 1988). Of course to demonstrate that the incorporated SM50 fusions do in fact titrate out the available factor(s) that interact with them would require direct *in vivo* saturation and competition measurements, as have been carried out with the CyIIIa gene (Flytzanis et al. 1987; Livant et al. 1988; R. Franks, unpublished data). We here rely, at least in part, on the quantitative similarity displayed by these two systems with respect to these measurements of reporter gene activity. In summary, the level of activity observed in embryos bearing SM50•CAT constructs is clearly not an inconsequential trickle, and could represent the maximum level of expression permitted by the concentrations of the available transcription factors.

The main focus of this work has been the spatial rather than the quantitative regulation of the SM50 gene, as this displays so striking a lineage specificity. The preservation of this lineage specificity in the expression of the SM50•CAT fusions implies that *trans*-acting factors that recognize the necessary regulatory elements included in the fusion constructs must be localized, or

activated, *exclusively* in the skeletogenic mesenchyme cell lineages, and that the causal basis of SM50 expression in skeletogenic mesenchyme cells is the interaction of the factors with the ***cis***-regulatory regions of this gene. We showed earlier that the identical CAT reporter construct is activated in a wholly exclusive set of embryonic lineages, the aboral ectoderm, when placed under the control of *cis*-regulatory sequences of the CyIIIa cytoskeletal actin gene, which is normally expressed only in aboral ectoderm (Flytzanis et al. 1987; Hough-Evans et al. 1987, 1988). The comparison provides a strong argument that it is simply the spatial distribution of active forms of the respective regulatory factors amongst the founder cells of the early embryo that determines the initial pattern of differential gene expression.

Autonomous and regulative differentiation of skeletogenic mesenchyme

The simple paradigm just derived fits the normal differentiation of the skeletogenic mesenchyme lineage from its 5th cleavage founder cells, since this lineage differentiates autonomously *in vitro* (see Introduction for references), and since experimental test ***excludes*** any requirement for (inductive) intracellular interactions in SM50 expression by cells of this lineage (Stephens et al. 1988). Thus it is difficult to avoid the conclusion that localized maternal cytoplasmic factors that directly or indirectly ***cause*** skeletogenic gene expression are segregated to the skeletogenic mesenchyme founder cells in the course of the geometrically programmed cleavage divisions by which they arise. How then to interpret the many observations, classical and modern, that demonstrate a regulative potential for skeletogenesis in lineages of the sea urchin embryo that never in normal development contribute to the skeletogenic mesenchyme? Among these observations are the following: (*i*) The tier of

blastomeres *lying immediately* above the micromeres at the 64-cell stage, i.e., the veg_2 tier of Hörstadius (1939) (in the current lineage map [*op cit*] these are the VAM1*L* and VLM2*L* cells that normally give rise to the definitive vegetal plate, archenteron, and non-skeletogenic mesenchyme) will produce small spicules if cultured in isolation (Hörstadius 1935); one cell of this tier will generate extensive skeletal rods if combined with a mid-cleavage animal half embryo, while two or more cells of this tier combined with the animal half embryo will generate an almost normal skeletal structure (reviewed by Hörstadius 1939, 1973, Figs. 15, 16, and 22). (*ii*) Whole embryos from which the micromeres have been removed regulate to give rise to pluteus larvae that contain normal skeletal elements (early references summarized by Hörstadius 1973, p. 60; Langelan and Whiteley 1985); the skeletogenic mesenchyme in these experiments also arises from mesenchymal derivations of the veg_2 tier. (*iii*) The next tier of blastomeres lying above veg_2, i.e., veg_1 of Hörstadius (1939) (the VAM1*U*, VAM2*U*, VOM1*U*, VOM2*U*, and the two VLM1*U* and VLM2*U* sister cells of veg_2 blastomeres in the lineage map) may also produce small skeletal elements if combined with an animal half embryo (Hörstadius 1973, p. 49). It is now known that in normal development some of these cells give rise to the apical region of the aboral ectoderm (i.e., VAM1*U* and VAM2*U*), and others to the anal plate of the aboral ectoderm (the VLM1*U* and VLM2*U* cells) and to the supra-anal ectoderm (VOM1*U* and VOM2U) (Cameron et al. 1987). (*iv*) Isolated animal half embryos which consist exclusively of cells that in normal development are ancestral to oral and aboral ectoderm, and to their neurogenic derivatives, can be induced to give rise to complete pluteus larvae that include normal formed skeletal structures, by treatment with LiCl (von Ubisch 1925, 1929; Hörstadius 1973, Fig. 34). (*v*) Removal of ingressed skeletogenic mesenchyme cells from within the blastocoel induces a quantitative, compensatory conversion of non-

skeletogenic secondary mesenchyme cells to the function of a skeletogenic lineage (Fukashi 1962; Ettensohn and McClay 1988). An apparent paradox arises from this evidence of widespread nonlocalized *potential* to produce skeletogenic cells, in considering results such as those presented here, that by logic generate the conclusion that normal specification of skeletogenic function is caused by the localization of regulatory factors, which result in lineage-specific, differential gene expression.

The following simple hypothesis expresses in terms of *trans*-regulatory factors both the autonomous differentiation of the skeletogenic mesenchyme from its founder cells in normal development, and the striking regulative ability of other cells in the early embryo to activate skeletogenic genes. We imagine that active forms of the necessary factors are sequestered in the polar region of the egg cytoplasm that is inherited by the skeletogenic founder cells, or that the capacity to activate these factors is sequestered there. Alternatively, if regulation is for this case negative, the active forms of skeletogenic gene repressors would be sequestered in normal embryos to non-skeletogenic lineages. However, *inactive forms of the same regulatory factors* would be present or be synthesized elsewhere, perhaps globally, in the early embryo. "Inactive" here could mean unable to translocate from cytoplasm to nucleus, unable to bind to specific DNA target sequences, or unable to affect transcription by interaction with other factors once bound, but in any case the state of inactivity must be one that can be relieved by a modification that occurs as a response to a ligand-specific *inductive* stimulation of a cell surface signalling system. The sea urchin embryo must employ such systems, since inductive interactions are essential for lineage specification along both the animal-vegetal and oral-aboral axes (see discussions of Hörstadius 1973, pp. 47-67; Davidson 1986, pp. 501-505; Wilt 1987). Thus the regulative ability of

various cells of the embryo to give rise to skeletogenic lineages in ectopic blastomere combinations would reflect the activation of the skeletogenic *trans*-regulators in consequence of an abnormal stimulation (or repression) of signalling systems that are in undisturbed embryos utilized for intercellular inductive processes. The effect of LiCl may be easily explicable in these terms as well, as this agent affects enzymatic metabolism of inositol phosphates generated in response to ligand-specific receptor interactions (e.g. see Berridge and Irvine 1984; Thomas et al. 1984; Mitchell 1986; Sherman et al. 1986; Hansen et al. 1986), and may have direct effects on certain G-proteins involved in receptor coupling as well (Drummond 1988; Avissar et al. 1988). In summary, the interpretation we offer states in essence that both the localized capacity for autonomous skeletogenic differentiation and the more widespread capacity for regulative skeletogenic differentiation might be explained by the spatial distribution with respect to lineage founder cells of modified and of unmodified *trans*-regulators in the cleavage stage embryo, and the ability of inductive cell signalling systems to effect their modification. A difficulty might be that the normal intercellular inductions occurring in undisturbed embryos might then be expected to produce skeleton in, say, veg_2 cells. However, there is as yet no evidence on the diversity or specificity of different inductive interactions, or on the differences that might distinguish the responses to ectopic *vs.* normal inductive signal responses. Furthermore, a secondary phenomenon particular to skeletogenesis probably also is important in interpreting the particular results of blastomere recombination experiments. This is that once embarked on their course of differentiation skeletogenic mesenchyme cells, whatever their origin, evidently emit signals that repress further mesenchyme cell differentiation, thus ensuring that the proper number of these cells will not be exceeded (Ettensohn and McClay 1988). In any case, the mysterious phenomenon of regulative

development, which has puzzled observers ever since its discovery in the initial phase of experimental embryology a century ago (reviews in Wilson 1925, Chapters 13 and 14; Davidson 1986, Chapter 6), may be explained generally in a similar way. That is, the regulative capacity of given portions of an early embryo could simply reflect the distribution of *trans*-regulators that can be converted to active forms through the operation of cell signalling systems, or treatments that affect these systems.

Materials and methods

Construction of SM50•CAT

A genomic clone was isolated from the same sea urchin EMBL3 genomic library as those previously described (Sucov et al. 1987), and selected for its longer 5' end. The *Sal*I-*Kpn*I fragment containing the 5' portion of the SM50 gene was subcloned into pUC18, restricted at the *Kpn*I site, and digested with Bal-31. *Sal*I linkers were added, the DNA restricted with *Sal*I, and the fragments containing the SM50 gene sequences were cloned into pUC18. One clone terminated at the 3' end just within the SM50 second exon. Its insert was excised with *Sal*I and cloned into a version of pSV0CAT (Gorman et al. 1982) modified so as to contain a *Sal*I site at the junction of the CAT coding region (Katula et al. 1987). The relevant sequence of this clone is shown in Figure 1(c).

Construction of SM50Δ•CAT

A region of SM50•CAT containing the first exon was subcloned into Bluescript (Stratagene), and single strand DNA was isolated according to recommended procedures. The 37-mer used for mutagenesis (Fig. 1(d)) was synthesized at the

Caltech Microchemical Facility, and phosphorylated with polynucleotide kinase. An equimolar amount of template and oligo were mixed, denatured at 65°C for 5 min, and rehybridized at 37°C for 10 min. Extension with Klenow polymerase and deoxynucleotides, and recircularization by ligation were as recommended (Stragene). The DNA was used to transform JM101. Following chloramphenicol amplification as described in the Bluescript manual, colonies on replicate filters were screened with the 37-mer, labeled with polynucleotide kinase and $^{32}\gamma$-ATP. Hybridization was at 65°C in 4X SET, 5X Denhardts solution, 0.1% SDS, 25 mM phosphate buffer (pH 6.8), and 50 µg/ml denatured calf thymus DNA; the final wash was at 65°C in 0.3X SET, 0.1% SDS, 25 mM phosphate buffer, and 0.1% $NaPP_i$. Approximately 1% of the recombinant colonies had been successfully mutagenized.

Construction of 2.2•CAT and 2.2PL•CAT

The 2.2 kb *Sal*I fragment from SM50Δ•CAT was subcloned into pSV0CAT, modified to contain a *Sal*I site by Flytzanis et al. (1987), to generate 2.2•CAT. The same fragment was inserted into the CAT vector pUC•PL•CAT (Bond-Matthews and Davidson 1988), modified to create a SalI site by fill-in of the BglII site and attachment of SalI linkers to generate 2.2PL•CAT.

Embryos

S. purpuratus eggs were microinjected as described (McMahon et al. 1985) with either *Bam*HI-linearized plasmid DNA or gel-purified fragments. Approximately 1500 molecules of CAT-containing DNA were introduced per egg, with a 4-fold molar excess of *Pst*I restricted carrier sea urchin genomic DNA partially digested to an average length of 10 kb. The inclusion of carrier DNA augments expression (R. Franks, unpubl. data). For CAT assays, embryos were collected at

prism stage and assayed by standard procedures, as described by Franks et al. (1988). For *in situ* analysis, a method to simplify the collection of the relatively small number of injected embryos was developed. Thirty µl of molten 1.5% agarose in sea water was allowed to harden in the bottom of a siliconized 200 µl yellow pipet tip. Embryos were transferred into approximately 100 µl of sea water overlying the agarose cushion, then spun in a Beckman microfuge with an Eppendorf tube as a holder at setting 3 for 2 min. As much of the overlying sea water as possible was aspirated with a drawn-out Pasteur pipette under a dissecting microscope. Seven µl of molten agarose in sea water were used to resuspend the loose embryo pellet. The yellow tip was held in ice-cold sea water for 20 sec to harden the agarose, then cut with a razor. The chunk containing the embryos was pushed into glutaraldehyde fix buffer and further processed and paraffin embedded essentially as described by Hough-Evans et al. (1987). All sections were cut at 5 microns. Hybridization was as described by Hough-Evans et al. (1987).

Probe excess RNA titration

Embryos were harvested 50 hr after injection, at the early pluteus stage of development. Aliquots were removed for analysis of CAT activity (quantitated in Table 1 as experiments 2 and 3) and of DNA content, the latter measured by a procedure described by Franks et al. (1988). Briefly, DNA was isolated from embryo pellets by proteinase K digestion and organic extraction, then denatured and affixed to nitrocellulose using a slot blot apparatus. The filter was hybridized with a single strand SP6-derived RNA probe containing only CAT sequences at 42°C in 50% formamide, 5X SSC, 1X Denhardts solution, 20 mM phosphate buffer, and 50 µg/ml calf thymus DNA; the final wash was at 60°C in 1X SSC, 0.2% SDS. For isolation of RNA, 1455 experimental embryos were

mixed with a 13-fold excess of uninjected embryos, pelleted, and frozen dry. Eighty percent of the experimental embryos are derived from experiment 2 of Table 1, the remainder from experiment 3. Embryo pellets were resuspended in a urea sarcosine buffer and digested with proteinase K as described by Lee et al. (1986), followed by organic extraction and ethanol precipitation. Nucleic acids were twice digested for 1 hr with 4 units of RNase-free DNase (RQ1, Promega) with the addition of 40 units RNasin (Promega) and 5 mM DTT, extracted, and precipitated. RNA recovery was 40-50 µg. Probe excess titration was exactly as described by Lee et al. (1986), using SP6-transcribed RNA probes labeled with 800 Ci/mmol ^{32}P-rUTP. The 131 nt CyIIIa probe was derived from the *Sal*I-*Bal*I fragment of pSV0CAT and contains only CAT sequences. Data reduction was as described by Lee et al. (1986).

Acknowledgements

This research was supported by NIH Grant HD-05753. H.M.S. was supported by NIH Training Grant GM-07616.

References

Anstrom, J. A., J. E. Chin, D. S. Leaf, A. L. Parks, and R. A. Raff. 1987. Localization and expression of msp130, a primary mesenchyme lineage-specific cell surface protein of the sea urchin embryo. *Development* **101:** 255-265.

Avissar, S., G. Schreiber, A. Danon, and R. H. Belmaker. 1988. Lithium inhibits adrenergic and cholinergic increases in GTP binding in rat cortex. *Nature* **331:** 440-442.

Benson, S. C., N. C. Benson, and F. Wilt. 1986. The organic matrix of the skeletal spicule of sea urchin embryos. *J. Cell Biol.* **102:** 1878-1886.

Benson, S. C., H. Sucov, L. Stephens, E. Davidson, and F. Wilt. 1987. A lineage-specific gene encoding a major matrix protein of the sea urchin embryo spicule. I. Authentication of the cloned gene and its developmental expression. *Dev. Biol.* **120:** 499-506.

Berridge, M. J. and R. F. Irvine. 1984. Inositol triphosphate, a novel second messenger in cellular signal transduction. *Nature* **312:** 315-321.

Bond-Matthews, B. and N. Davidson. 1988. Transcription from each of the *Drosophila* act 5c leader exons is driven by a separate functional promoter. *Gene* **62:** 289-300.

Cameron, R. A., B. R. Hough-Evans, R. J. Britten, and E. H. Davidson. 1987. Lineage and fate of each blastomere of the eight-cell sea urchin embryo. *Genes & Dev.* **1:** 75-85.

Carson, D. D., M. C. Farach, D. S. Earles, G. L. Decker, and W. J. Lennarz. 1985. A monoclonal antibody inhibits calcium accumulation and skeleton formation in cultured embryonic cells of the sea urchin. *Cell* **41:** 639-648.

Davidson, E. H. 1986. *Gene Activity in Early Development*, 3rd ed. Academic Press, Orlando, Florida.

Davidson, E. H. and R. J. Britten. 1971. Note on the control of gene expression during development. *J. Theor. Biol.* **32:** 123-130.

Drummond, A. H. 1988. Lithium affects G-protein receptor coupling. *Nature* **331:** 388.

Ernst, S. G., B. R. Hough-Evans, R. J. Britten, and E. H. Davidson. 1980. Limited complexity of the RNA in micromeres of sixteen-cell sea urchin embryos. *Dev. Biol.* **79:** 119-127.

Ettensohn, C. A. and D. R. McClay. 1988. Cell lineage conversion in the sea urchin embryo. *Dev. Biol.* **125:** 396-409.

Flytzanis, C. N., A. P. McMahon, B. R. Hough-Evans, K. S. Katula, R. J. Britten, and E. H. Davidson. 1985. Persistence and integration of cloned DNA in postembryonic sea urchins. *Dev. Biol.* **108:** 431-442.

Flytzanis, C. N., R. J. Britten, and E. H. Davidson. 1987. Ontogenic activation of a fusion gene introduced into sea urchin eggs. *Proc. Natl. Acad. Sci. USA* **84:** 151-155.

Franks, R. R., B. R. Hough-Evans, R. J. Britten, and E. H. Davidson. 1988. Direct introduction of cloned DNA into the sea urchin zygote nucleus, and fate of injected DNA. *Development* **102:** 287-299.

Fukushi, T. 1962. The fates of isolated blastoderm cells of sea urchin blastulae and gastrulae inserted into the blastocoel. *Bull. Mar. Biol. Station of Asamushi.* **11:** 21-30.

Gorman, C., G. Merlino, M. Willingham, I. Pastan, and B. Howard. 1982. The Rous sarcoma virus long terminal repeat is a strong promoter when introduced into a variety of eukaryotic cells by DNA-mediated transfection. *Proc. Natl. Acad. Sci.* **79:** 6777-6781.

Gustafson, T. and L. Wolpert. 1963. The cellular basis of morphogenesis and sea urchin development. *Int. Rev. Cytol.* **15:**139-214.

Hansen, C. A., S. Mah, and J. R. Williamson. 1986. Formation and metabolism of inositol 1,3,4,5-tetrakisphosphate in liver. *J. Biol. Chem.* **261:** 8100-8103.

Harkey, M. A. and A. H. Whiteley. 1980. Isolation, culture, and differentiation of echinoid primary mesenchyme cells. *Wilhelm Roux's Arch.* **189:** 111-122.

Harkey, M. and A. H. Whiteley. 1983. The program of protein synthesis during the development of the micromere-primary mesenchyme cell line in the sea urchin embryo. *Dev. Biol.* **100:** 12-28.

Harkey, M. A. 1985. Aspects of gene expression in the sea urchin micromere-primary mesenchyme cell line. In *The Cellular and Molecular Biology of Invertebrate Development* (ed. R. H. Sawyer and R. M. Showman), The Bell W. Baruch Library in Marine Science No. 15, pp. 311-327. Univ. South Carolina Press, Columbia, South Carolina.

Harkey, M. A., H. R. Whiteley, and A. H. Whiteley. 1988. Coordinate accumulation of primary mesenchyme-specific transcripts during skeletogenesis in the sea urchin embryo. Submitted for publication.

Hörstadius, S. 1935. Über die Determination im Verlaufe der Eiachse bei Seeigelen. *Pubbl. Stn. Zool. Napoli* **14:** 251-479.

Hörstadius, S. 1939. The mechanics of sea urchin development, studied by operative methods. *Biol. Rev. Cambridge Phil. Soc.* **14:** 132-179.

Hörstadius, S. 1973. *Experimental Embryology of Echinoderms.* Clarendon Press, Oxford.

Hough-Evans, B. R., R. R. Franks, R. A. Cameron, R. J. Britten, and E. H. Davidson. 1987. Correct cell type-specific expression of a fusion gene injected into sea urchin eggs. *Dev. Biol.* **121:** 576-579.

Hough-Evans, B. R., R. J. Britten, and E. H. Davidson. 1988. Mosaic incorporation of an exogenous fusion gene expressed exclusively in aboral ectoderm cells of the sea urchin embryo. *Dev. Biol.*, in press.

Katula, K. S., B. R. Hough-Evans, R. J. Britten, and E. H. Davidson. 1987. Ontogenic expression of a Cyl actin fusion gene injected into sea urchin eggs. *Development* **101:** 437-447.

Killian, C. E. and F. H. Wilt. 1988. The accumulation and translation of a spicule matrix protein mRNA during sea urchin embryo development. *Dev. Biol.*, in press.

Lai, Z., R. Maxson, and G. Childs. 1988. Both basal and ontogenic promoter elements affect the timing and level of expresion of a sea urchin H1 gene during early embryogenesis. *Genes & Dev.* **2:** 173-183.

Langelan, R. E. and A. H. Whiteley. 1985. Unequal cleavage and the differentiation of echinoid primary mesenchyme. *Dev. Biol.* **109:** 464-475.

Leaf, D. S., J. A. Showman, J. E. Chin, M. A. Harkey, R. M. Showman, and R. A. Raff. 1987. Antibodies to a fusion protein identify a cDNA clone encoding msp130, a primary mesenchyme-specific cell surface protein of the sea urchin embryo. *Dev. Biol.* **121:** 29-40.

Lee, J. J., F. J. Calzone, R. J. Britten, R. C. Angerer, and E. H. Davidson. 1986. Activation of sea urchin actin genes during embryogenesis. Measurement of transcript accumulation from five different genes in *Strongylocentrotus purpuratus*. *J. Mol. Biol.* **188:** 173-183.

Livant, D., A. Cutting, R. J. Britten, and E. H. Davidson. 1988. Titration of the activity of a fusion gene in intact sea urchin embryos. Submitted for publication.

McCarthy, R. A. and M. Spiegel. 1983. Protein composition of the hyaline layer of sea urchin embryos and reaggregating cells. *Cell Differentiation.* **13:** 93-102.

McMahon, A. P., C. N. Flytzanis, B. R. Hough-Evans, K. S. Katula, R. J. Britten, and E. H. Davidson. 1985. Introduction of cloned DNA into sea urchin egg

cytoplasm: Replication and persistence during embryogenesis. *Dev. Biol.* **108:** 420-430.

Mitchell, B. 1986. Profusion and confusion. *Nature* **319:** 176-177.

Okazaki, K. 1975. Spicule formation by isolated micromeres of the sea urchin embryo. *Amer. Zool.* **15:** 567-581.

Pehrson, J. R. and L. H. Cohen. 1986. The fate of the small micromeres in sea urchin development. *Dev. Biol.* **113:** 522-526.

Pittman, D. and S. G. Ernst. 1984. Developmental time, cell lineage and the environment regulate the newly synthesized proteins in sea urchin embryos. *Dev. Biol.* **106:** 236-242.

Sherman, W. R., B. G. Gish, M. P. Honchar, and L. Y. Munsell. 1986. Effects of lithium on phosphoinositide metabolism *in vivo*. *Fed. Proc.* **45:** 2639-2646.

Stephens, L., T. Kitajima, and F. Wilt. 1988. The role of cell interactions in tissue-specific gene expression of sea urchin embroyos. Submitted for publication.

Sucov, H. M., S. Benson, J. J. Robinson, R. J. Britten, F. Wilt, and E. H. Davidson. 1987. A lineage-specific gene encoding a major matrix protein of the sea urchin embryo specule. II. Structure of the gene and derived sequence of the protein. *Dev. Biol.* **120:** 507-519.

Thomas, A. P., J. Alexander, and J. R. Williamson. 1984. Relationship between inositol polyphosphate production and the increase of cytosolic free Ca^{2+} induced by vasopressin in isolated hepatocytes. *J. Biol. Chem.* **259:** 5574-5584.

von Ubisch, L. 1925a. Entwicklungsphysiologische Studien an Seeigelkeimen. I. *Z. wiss. Zool.* **124:** 361-381.

von Ubisch, L. 1925b. Entwicklungsphysiologische Studien an Seeigelkeimen. II. *Z. wiss. Zool.* **124:** 457-468.

von *Ubisch, L.* 1925c. Entwicklungsphysiologische Studien an Seeigelkeimen. III. *Z. wiss. Zool.* **124:** 469-486.

von Ubisch, L. 1925d. Über die Entodermisierung ektodermaler Bezirke des Echinoiden-Keimes und die Reversion dieses Vorganges. *Verh. phys.-med. Ges. Würzburg* **50:** 13-19.

von Ubisch, L. 1929. Über die Determination der larvalen Organe under die Imaginalanlage bei Seeigelen. *Wilhelm Roux's Arch.* **117:** 81-122.

Wessel, G. M. and D. R. McClay. 1985. Sequential expression of germ-layer specific molecules in the sea urchin embryo. *Dev. Biol.* **111:** 451-463.

Wilson, E. H. 1925. *The Cell in Development and Heredity*, 3rd ed. Macmillan, New York.

Wilt, F. H. 1987. Determination and morphogenesis in the sea urchin embryo. *Development* **100:** 559-575.

Table 1: CAT enzyme activity in transgenic *S. purpuratus* embryos bearing various SM50•CAT fusion constructs

	Experiment 1		Experiment 2		Experiment 3	
construct	% acetylation	units CAT/embryo	% acetylation	units CAT/embryo	% acetylation	units CAT/embryo
SM50•CAT	0.13	3.6×10^{-7}	0.24	3.0×10^{-7}	0.14	1.3×10^{-7}
SM50Δ•CAT	0.24	1.0×10^{-6}				
2.2•CAT	1.20	9.5×10^{-6}	1.81	3.3×10^{-6}	0.98	1.7×10^{-6}
uninjected controls	0.07	—	0.08	—	0.08	—

In experiment 1 each value represents the average of two samples (SM50•CAT, 2.2•CAT, uninjected controls) or three samples (SM50Δ•CAT), each sample containing 25-45 injected embryos. In experiments 2 and 3 each value represents one sample of 46 injected embryos.

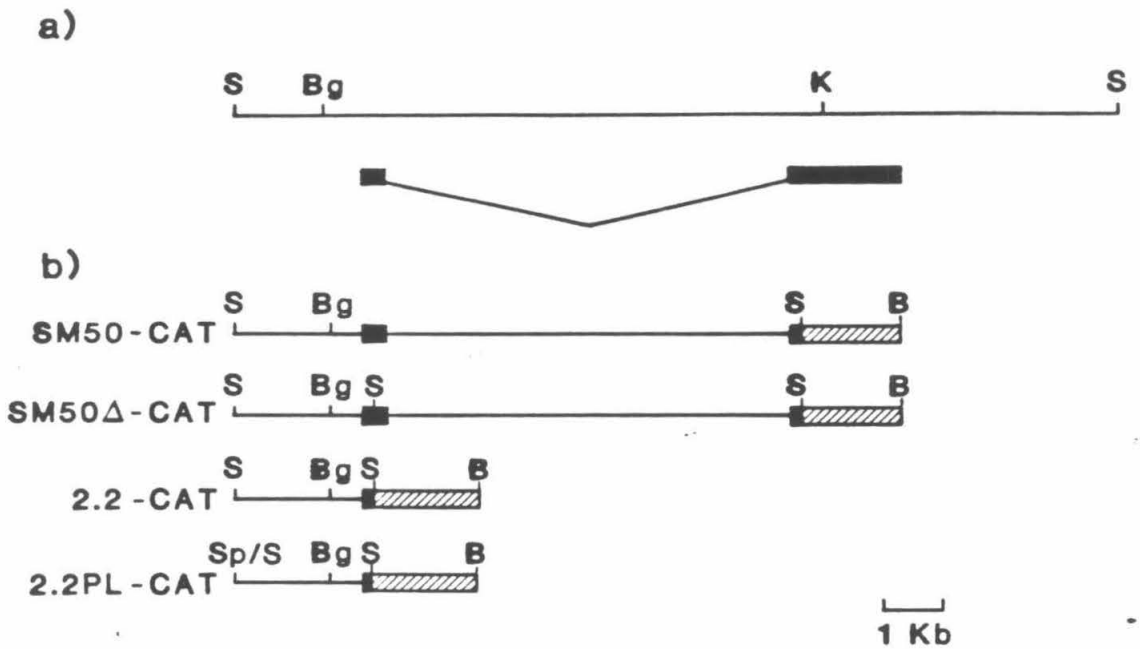
Table 2: *CAT enzyme* activity in transgenic embryos bearing various 5' SM50•CAT deletions[a]

Fragment of 2.2PL•CAT injected[b]	CAT enzyme units/embryo (av) expt. 1	expt. 2
*Sph*I - *Bam*HI	1.6×10^{-5}	1.1×10^{-5}
*Eco*RV - *Bam*HI	1.8×10^{-5}	
*Bgl*II - *Bam*HI	1.9×10^{-5}	1.4×10^{-5}

[a]Each value represents the average of 2 or 3 samples, except for the *Eco*RV - *Bam*HI fragment (1 sample). Each sample contained 45-60 injected embryos.

[b]See Figure 1(b).

Figure 1. *Structure of SM50* gene and of CAT gene fusions. (a) Organization of the SM50 gene. Exons of the SM50 gene are depicted as solid rectangles. The genomic clone from which the CAT constructs were derived is shown above (see Materials and methods). Restriction sites: Bg, *Bgl*II; K, *Kpn*I; R, *Eco*RV; S, *Sal*I; (b) SM50•CAT constructs. Regions derived from exons of the SM50 gene are depicted as filled rectangles; flanking and intron sequences as thin lines. The CAT coding region and SV40 polyadenylation region derived from pSV0CAT are depicted as a striped box. The vector and nonessential SV40 sequences appended to the 5' end of each construct are not shown. B, *Bam*HI; Sp, *Sph*I. The scale in (a) and (b) is the same. (c) N-terminal sequence of the expected fusion protein, and nucleotide sequence of the junction region of SM50•CAT. The protein coding region derived from the first exon of the SM50 gene is shown at the top. Below is shown the junction between sequences of the second SM50 exon and the CAT vector, which were ligated together at the indicated *Sal*I site. The boxed ATG codon (codon 58) is the original translation initiation codon of the bacterial CAT protein. The first 43 amino acids of the open reading frame of the construct derive from SM50, and residues 44-57 are from translation of vector sequences preceding the beginning of the CAT coding region. Numbering begins with the SM50 initiation codon. (d) Oligonucleotide directed mutagenesis, and deletion of the sequence coding for the signal peptide. The coding strand of the first exon is shown on the top line of sequence. Below is the 37 nucleotide oligomer used for mutagenesis. The location of the introduced *Sal*I restriction site, and the translation product of the resultant open reading frame are shown.



c)

5'UT...GCAGCCATCAAGGCAGTTTTGTTTATTGTGGCTAGTCTTATAGCCTTTGCTACGGGTCAAGACTGCCCCAGCATACTACGTCGGCAGTCAATCGGGTCAATCATGTTACAGgt...intron

MetLysGlyValLeuPheIleValAlaSerLeuIleAlaPheAlaThrGlyGlnAspCysProAlaTyrTyrValArgSerGlnSerGlyGlnSerCysTyrAr

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34

SM50 exon I

SalI

intron...agATACTTCAACATCCGCCTTCCCTCGTCGACCAGCTTGGCGACATTTTCAGGAGCTAAGGAAGCTAAAATG...CAT coding region

gTyrPheAsnMetArgValProTrpSerThrSerLeuAlaArgPheSerGlyAlaLysGluAlaLysMet

35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58

SM50 exon II

pSVO+CAT vector

d)

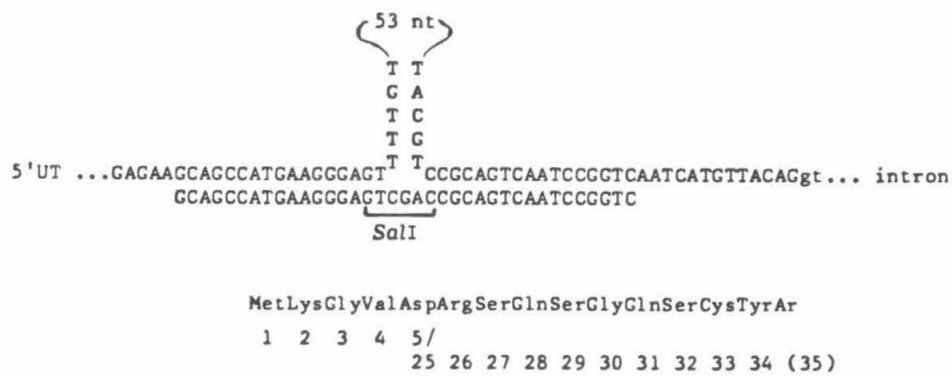


Figure 2. *In situ* hybridizations of embryos expressing SM50•CAT. Sections of six different embryos at mesenchyme blastula stage are shown. The sections are oriented so that the vegetal pole is at bottom. The embryo in panel (b) was distorted during sectioning; its vegetal pole is at bottom right. The number of labeled mesenchyme cells, and the number of mesenchyme cells present in each panel is: (a) 5 of 6, (b) 4 of 13, (c) 3 of 5, (d) 5 of 9, (e) 3 of 14, and (f) 1 of 9. The single labeled cell in the section shown in panel (f) was the only mesenchyme cell of 43 labeled in this embryo. Labeling at the vegetal pole is seen (slightly out of focus) in the embryo section of panel (c).

a
b
c
d
e
f

Figure 3. Mosaic expression of SM50•CAT constructs in mesenchyme cells labeled by *in situ* hybridization. The fraction of labeled mesenchyme cells in 66 embryos grown from eggs injected with SM50•CAT (a) or of 35 embryos grown from eggs injected with 2.2•PL•CAT (b) is shown. The fraction of labeled cells for each embryo was normalized to 32 cells, rounded to the nearest 0.2. Of the labeled embryos, the mean normalized number of labeled cells is (a) 5.5, and (b) 6.4 (arrows). Below the horizontal axis of panel (a) is shown the boundaries that were used to group the embryos into mosaicism classes as shown in the inset. The model used to group the embryos assumes that all of the injected CAT DNA is stably integrated into a single cell during early cleavage (see Hough-Evans et al. 1988) and detectably expressed in all the progeny which become primary mesenchyme. Thus, for an integration at the 2-cell stage, 50% of the mesenchyme cells are expected to be labeled as each blastomere contributes equally to the mesenchyme population. Integration events at the 4 cell through 32 cell stage would all result in 25% of the mesenchyme cells labeled in those embryos which show label. Later integration events would result in geometrically decreasing numbers of labeled cells. This behavior can be described by the function 2^{-n}, where n=1 for the 2 cell stage, n=2 for the 4 through 32 cell stage, and n=3, 4, 5 ... for the 64, 120, and 240 cell stages. The range used to group the experimental data of panel (a) for the inset was $2^{-n(\pm 0.5)}$.

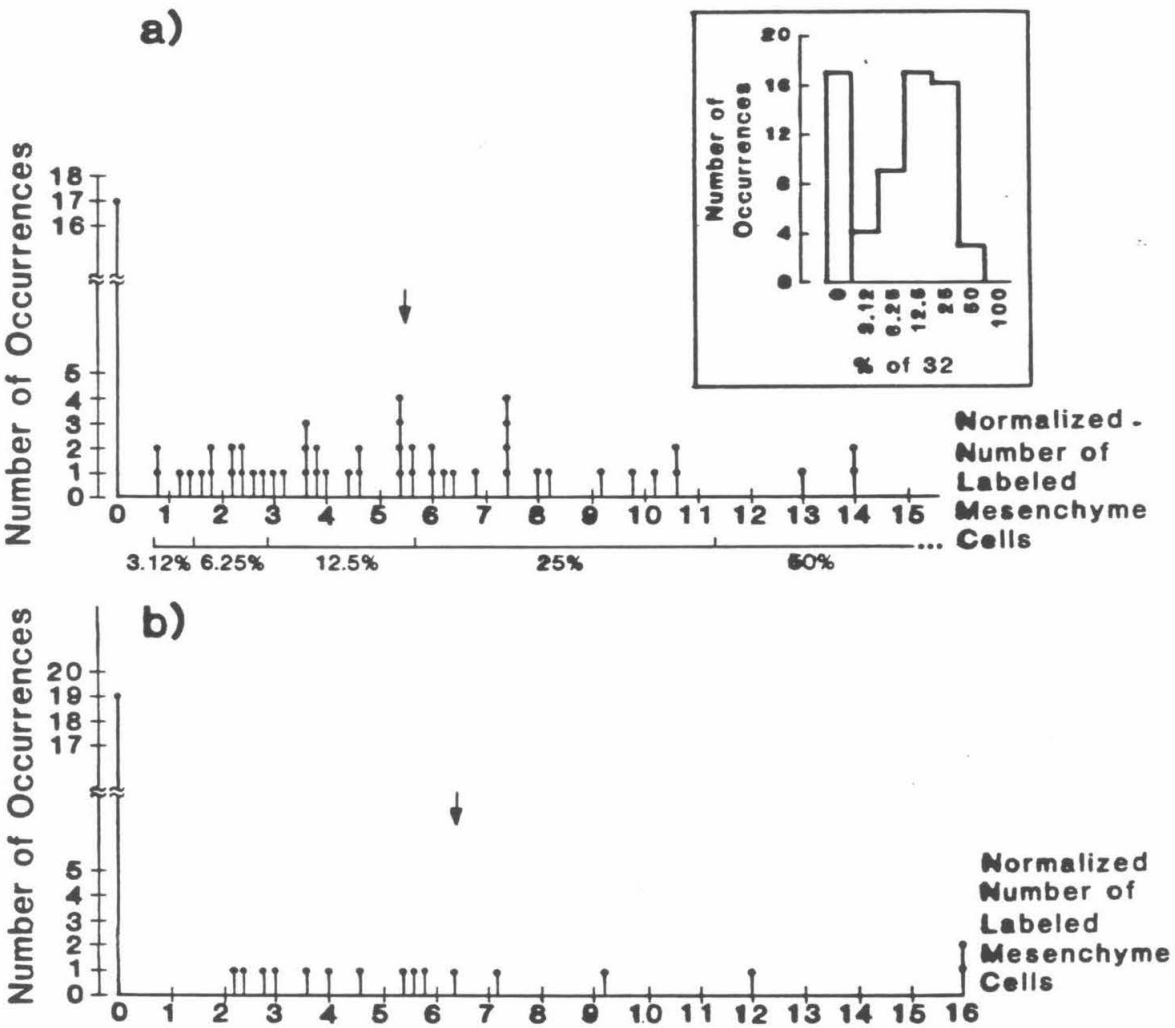

a)
Number of Occurrences
18
17
16
6
4
3
2
1
0
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15
Normalized - Number of Labeled Mesenchyme Cells
3.12% 6.25% 12.5% 25% 50%
Number of Occurrences
20
16
12
8
4
0
0 3.12 6.25 12.5 25 50 100
% of 32
b)
Number of Occurrences
20
19
18
17
5
4
3
2
1
0
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16
Normalized Number of Labeled Mesenchyme Cells

Figure 4. *Titration of fusion gene transcripts.* (a) Quantitation of CAT DNA in experimental embryos. DNA was extracted from 612 carrier embryos plus 30 experimental embryos grown 50 hr (early pluteus) from eggs injected with SM50•CAT or 2.2•CAT. The yield of DNA, determined by DAPI fluorescence measurement, was 1.5 μg (approximately complete) for each experimental sample. 1.25 μg of DNA from the SM50•CAT sample (slot 1) or the 2.2•CAT sample (slot 2), or a known amount of linearized CAT DNA mixed with 1.25 μg of carrier embryo DNA (slots 4-9) were denatured, affixed to nitrocellulose, and hybridized with a probe specific for CAT sequences. Standards in slots 4-9 are 0, 0.5, 1, 2, 5 and 10 x 10^6 molecules of CAT DNA respectively. Slot 3 contains no DNA. Conversion of hybridized cpm, measured by scintillation counting of excised bands, to molecules of CAT DNA indicates that the SM50•CAT sample contained 2 x 10^6 molecules of CAT DNA, and the 2.2•CAT sample 0.4 x 10^6 molecules CAT DNA. (b) Titration of the CyIIIa message. Increasing amounts of RNA extracted from 18800 50 hr carrier embryos plus 1455 50 hr experimental embryos expressing SM50•CAT (open circles) or 2.2•CAT (open triangles), or from control (uninjected) embryos (filled circles) were hybridized in solution to an excess of ^{32}P-labeled antisense CyIIIa RNA probe. Unhybridized RNA was digested with RNase A and T1, and TCA precipitable cpm measured and plotted vs. input mass of sea urchin RNA. The slope of the line is a function of message prevalence and is independent of RNA recovery between the samples. Lines were fitted to data points by linear least squares regression. (c) Titration of the SM50•CAT message. Symbols and method are as in (b), except that RNA samples were hybridized to an antisense CAT RNA probe.

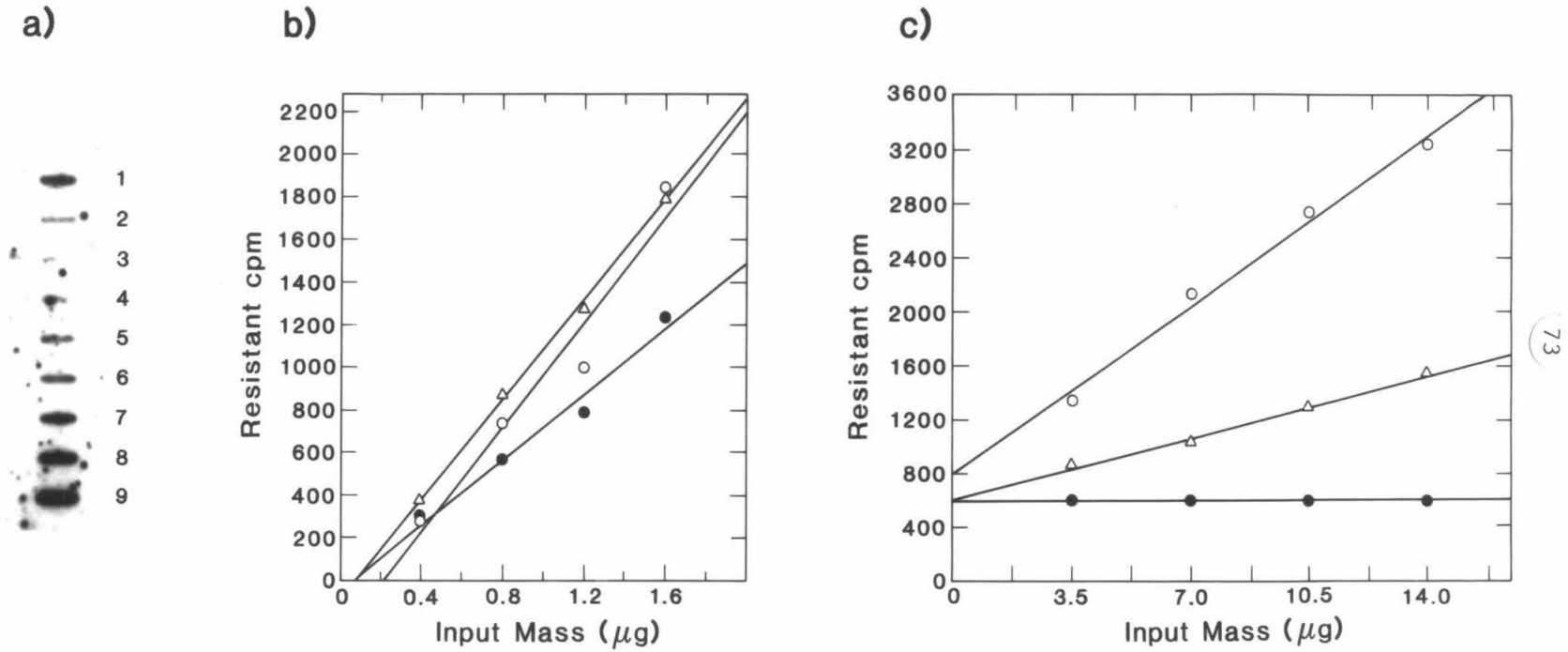
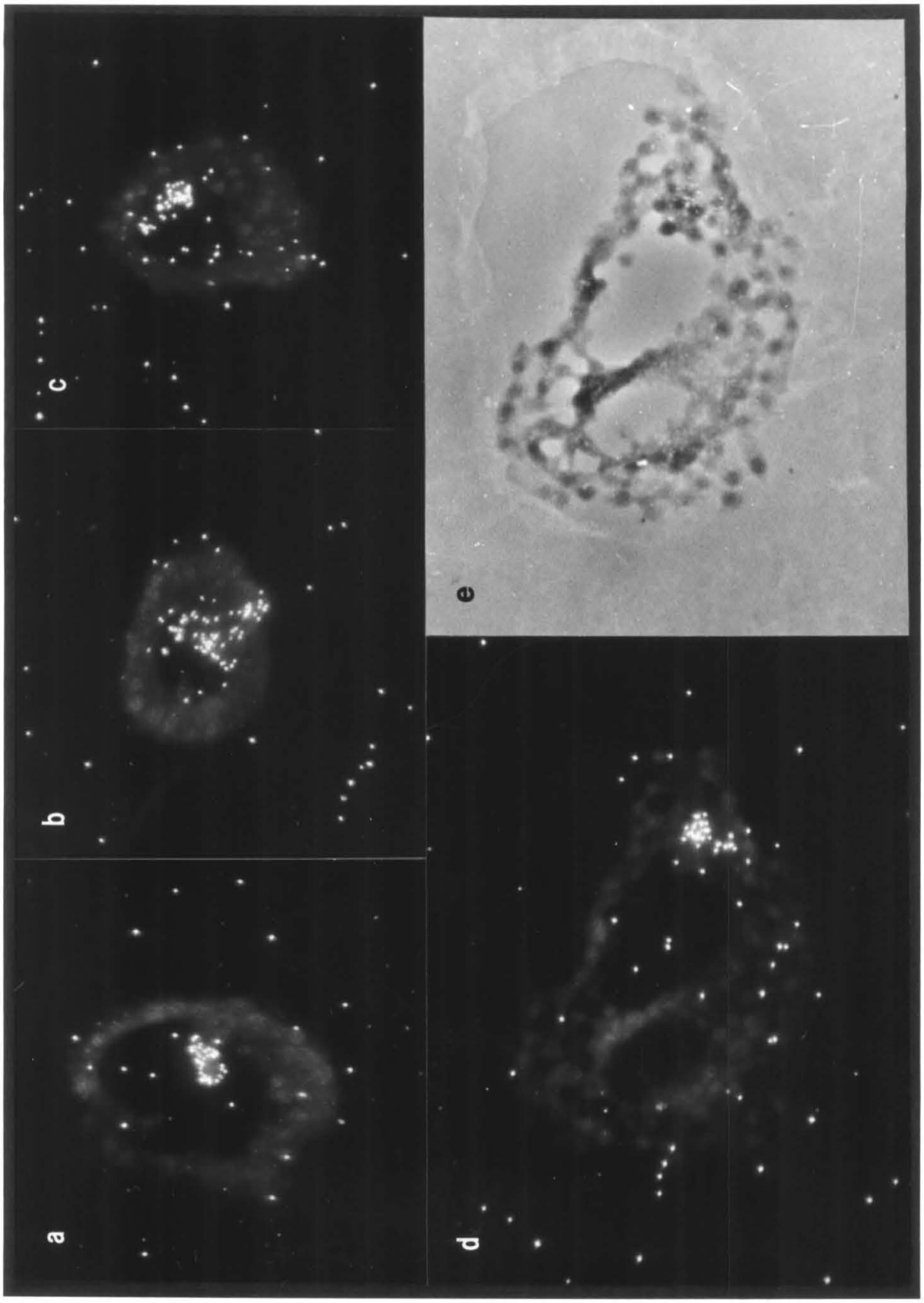

a)
1
2
3
4
5
6
7
8
9
b)
Resistant cpm
2200
2000
1800
1600
1400
1200
1000
800
600
400
200
0
0
0.4
0.8
1.2
1.6
Input Mass (μg)
c)
Resistant cpm
3600
3200
2800
2400
2000
1600
1200
800
400
0
0
3.5
7.0
10.5
14.0
Input Mass (μg)

Figure 5. *In situ* hybridization of embryos expressing the *Bgl*II-*Bam*HI fragment of 2.2PL•CAT. (a)-(c) represent embryos collected at mesenchyme blastula stage; (d)-(e) represent an embryo at early pluteus stage (60 hr of development), seen in darkfield (d) and phase contrast (e). Mesenchyme blastula stage embryos are oriented with the vegetal pole at the bottom. In (b) labeling can be seen in cells still within the vegetal wall, that would shortly have ingressed into the blastocoel. The number of labeled mesenchyme cells, and the number of mesenchyme cells present is: (a) 2 of 4, (b) 3 of 5, and (c) 3 of 7. The pluteus stage embryo (which is slightly distorted from sectioning) is oriented with the vegetal pole down and the future site of the mouth at the top left. The cluster of mesenchyme cells labeled in this section are located at the future apex of the pluteus.


a
b
c
d
e

CHAPTER 3

Regulatory Proteins Which Interact with the Promoter of a Skeletal Matrix Protein Gene (SM50) in the Sea Urchin Embryo

Regulatory Proteins Which Interact with the Promoter of a Skeletal Matrix Protein Gene (SM50) in the Sea Urchin Embryo

Henry M. Sucov, Frank J. Calzone, Kellie L. Whittaker,
Roy J. Britten, and Eric H. Davidson

Division of Biology
California Institute of Technology
Pasadena, CA 91125

ABSTRACT

Nuclear proteins isolated from sea urchin embryos at the mesenchyme blastula stage have been analyzed for binding activity to the promoter of the SM50 gene, a gene expressed exclusively in embryonic skeletogenic cells. Five binding sites have been identified and mapped to varying degrees of resolution. Two of the five sites are homologous and are potential enhancer elements. Two sites are also present in the promoter of a gene expressed in a different tissue, the aboral ectoderm specific cytoskeletal actin gene CyIIIa. One of these is a CCAAT element, the other is a putative repressor element. The fifth binding site overlaps the binding site of the putative repressor and may function as a positive regulator by preventing binding of the repressor.

INTRODUCTION

The central process of developmental biology is the differentiation of tissues from undifferentiated, uncommitted embryonic blastomeres. The manifestation of tissue differentiation is the expression of characteristic sets of genes. One approach to understanding the mechanisms of development is to understand the processes which regulate the expression of these tissue specific genes.

It is by now commonly recognized that transcription by RNA polymerase II in eukaryotic cells is mediated by the presence of DNA binding proteins arranged on gene promoter regions. The regulation of gene expression is therefore a consequence of how these proteins are regulated - how their presence is controlled, how their DNA binding properties are manifest, and how they regulate transcription.

The sea urchin embryo is an ideal system for a biochemical analysis of the nature of DNA binding proteins in developmental gene regulation. The complete lineage of the early embryo is known (Davidson, 1986), cloned DNA probes representing various genes expressed in specific tissues of most of the embryo are available (Angerer and Davidson, 1984), and quantities of synchronous embryos approaching the numbers available with tissue culture can be easily grown in the laboratory to any stage of early development. This paper concerns itself with one of the more accessible lineages of the sea urchin embryo, that which gives rise to the embryonic skeleton, or spicule.

Early cleavage in the sea urchin proceeds along a preformed animal-vegetal axis established during oogenesis (Schroeder, 1980). The first two cleavages are vertical, the third horizontal, and the fourth vertical in the animal pole blastomeres and horizontal and asymmetric in the vegetal pole blastomeres, resulting in three tiers of different sized cells - eight mesomeres, four

macromeres, and four micromeres. The micromeres form invariantly at the extreme vegetal pole of the embryo. These cells again divide, separating into four small and four large micromeres, the latter which are irreversibly and exclusively committed to the skeletogenic pathway. The progeny of these cells, known as the primary mesenchyme, become the first cells to invade the blastocoel, whereupon they take up positions along the future oral side of the blastocoel and begin the deposition of the spicule.

The commitment of the micromere progeny to the skeletogenic lineage has been described in terms of cytoplasmic localization, a process whereby maternally derived determinants synthesized during oogenesis and stored along the animal-vegetal axis of the egg at the vegetal pole are inherited exclusively by the micromeres, causing their commitment to the skeletogenic pathway. Thus, micromeres isolated at the time of their formation at the 16 cell stage and cultured in vitro will proceed to develop and differentiate on schedule, forming spicules often indistinguishable from their in vivo counterparts (Okazaki, 1975). In order to experimentally approach the nature of these processes of cytoplasmic localization and commitment in the sea urchin micromeres, we have cloned and characterized a gene expressed in the skeletogenic progeny of these cells. This gene, called SM50, encodes one of the major structural proteins of the spicule, and is expressed in embryogenesis exclusively in the skeletogenic mesenchyme (Benson et al., 1987; Sucov et al., 1987). We have previously shown by gene transfer studies that a 562 bp SM50 promoter fragment is necessary and sufficient to direct expression of a reporter gene at high levels specifically in the mesenchyme in transgenic embryos (Sucov et al., 1988). In this report, we analyze the regulatory proteins which interact with the SM50 promoter.

Our understanding of the processes which occur at the SM50 promoter have been considerably augmented by a parallel analysis of the proteins which interact

with the promoter of a second lineage specific gene, the CyIIIa cytoskeletal actin gene (Calzone et al., 1988), which is expressed exclusively in the aboral ectoderm (Cox et al., 1986). For the purposes of this paper, it is sufficient to point out that the CyIIIa and SM50 genes are expressed in nonoverlapping sets of tissues, which do not share a common precursor past the 8 cell stage (i.e., before the formation of the micromeres).

RESULTS

The sequence of a 562 bp BglII-SalI fragment, which contains the functional SM50 promoter as defined by gene transfer studies, is shown in Figure 1. The fragment contains 438 bp of upstream promoter sequence and the first 120 bp of the SM50 first exon, including the SM50 translation initiation codon at position +110. In the CAT constructs used in gene transfer experiments this fragment was fused to the CAT coding region at the introduced SalI site.

Proteins which Interact with the SM50 Promoter

Nuclear proteins were isolated from 27 hr mesenchyme blastula stage whole embryos, and assayed for the ability to bind to specific sequences in the SM50 promoter by the gel retardation assay. At this stage the SM50 gene is active and transcripts are accumulating to near maximum levels (Killian and Wilt, 1988). Five different protein binding sites have been identified and mapped in the SM50 promoter fragment, as indicated in Figure 1. For factors A, P3A, and CCAAT, the location of binding sites have been resolved to restriction fragments which are able to bind proteins; for factor B, the boundaries of the protein binding site have been further delimited by DNaseI footprinting. The binding site mapping experiments are described in detail below.

It is of course possible that additional protein binding sites exist which have not been detected, either because the prevalence of the protein at blastula stage is not sufficiently high to permit detection, or because the conditions employed are not optimal for binding or detection. In fact, there is preliminary evidence that an additional binding site does exist proximal to the transcription initiation site (see below). The binding of some factors is also strongly influenced by the

choice of nonspecific DNA competitor used in the gel shift reaction mixture. For example, factor B binds in the presence of poly (dI)/d(C), but barely at all in the presence of poly (dI-C)/(dI-C) (data not shown). Essentially the entirety of the 560 bp BglII-SalI fragment has been examined by gel shift with poly d(I-C)/d(I-C), poly d(I)/d(C), and poly d(A-T)/d(A-T), both as the whole fragment and as restriction subfragments. The detection of any additional binding sites may therefore require more refined probes or procedures.

In most cases, factor binding results in several bands in the gel shift experiments described below. This is a common observation, and is generally taken to indicate multiple forms of the same factor (i.e., modifications) or different proteins which recognize the same sequence. If the template contains separate and distinct binding sites of two proteins, these can usually be identified by restriction mapping experiments as described below.

Using the BglII-AvaII fragment as probe, two distinct binding sites for a single protein, called factor A, have been detected. The two binding sites have been mapped by progressively digesting asymmetrically labeled probes with restriction enzymes to the 30 bp HaeIII-HhaI fragment and the 25 bp HinfI-HpaII fragment, as shown in Figure 2a-c. That these two binding sites represent the same factor, rather than individual factors, is evidenced by the strikingly identical pattern of bands in the gel shift experiments shown in Figure 2, and the ability of one competitor to compete comparably for both factors (Fig. 2d). An 11 of 15 bp sequence homology (CCCCTTCGCTCCCCC and CCCTTTCGGAACCCC; underlined in Fig. 1) is found within the regions to which the factor A binding sites map and may represent the sequence which is recognized by the factor. The gel shift band pattern seen with factor A is rather complex, consisting of at least five complexes. Probes which contain both factor A binding sites also show two additional bands of faster mobility (e.g., the second

lane of Fig. 2a), which represent occupancy of factor A at both sites. While binding of two proteins to a template usually results in complexes of slower mobility, the faster migration seen here may indicate some sort of bimolecular interaction which makes the DNA more compact and more electrophoretically mobile.

The binding site of factor B has been similarly mapped by restriction gel shifts (Fig. 3a-b), and further defined by DNaseI footprinting to lie between positions -135 and -99 (data not shown). Binding of factor B to this sequence is considerably augmented by the usage of poly d(I)/d(C) as the nonspecific competitor DNA in gel shift experiments, rather than poly d(I-C)/d(I-C), perhaps because of the unusual sequence of alternating purines and pyrimidines found in the region of the binding site. Binding of factor B is reduced considerably (see legend to Fig. 3c) by a deletion of 10 bp (bases -113 to -104; see also Fig. 1), which coincidently creates a SmaI restriction site (CCCGGG) at the point of deletion. 36 bp is an unusually large region to be protected by a single protein in DNaseI footprints, and may therefore be the binding site of two adjacent proteins. However, we have so far been unable to resolve two distinct DNA binding activities at this sites.

The CCAAT motif found in inverted orientation at position -155 shares a larger 13 of 14 sequence homology with the CCAAT element of the CyIIIa actin gene known as P4. Both genes share the conserved core sequence TGATTGG with the CCAAT box of a histone gene from a related sea urchin species (Barberis et al., 1987). The presence of a CCAAT-binding protein has been demonstrated in embryonic extracts by DNaseI footprinting of the actin and histone genes (Calzone et al., 1988; Barberis et al., 1987). That the SM50 element is in fact a CCAAT box is evidenced by protection of the core G residues of the SM50 CCAAT sequence in a methylation interference assay (data

not shown), and by the ability of unlabeled SM50 fragments to compete effectively for the binding of the CCAAT factor to a labeled CyIIIa probe (Fig. 4a).

The SM50 promoter fragment also reacts with another CyIIIa binding factor known as P3A (P. Thiebaud and E. Davidson, unpubl. obs.). The binding site of P3A on the CyIIIa promoter has been mapped by DNaseI footprinting to a specific 20 bp sequence (F.C., unpubl. obs.). The region of SM50 responsible for competition with P3A was determined by using a variety of SM50 fragments as competitors for binding of nuclear proteins to labeled CyIIIa P3 DNA (Fig. 4b). The binding of a separate protein (P3B), which recognizes an octamer sequence also on the P3 probe, is equally diminished by all the competitor fragments (due to the relatively large mass of competitor DNA present), serving as a useful internal control. Two of the competitor fragments (lanes 5 and 7) are derived from the SM50 construct described above which deletes 10 bp from the factor B binding site, creating a SmaI restriction site; the others represent the original SM50 template. As shown in Figure 4b-c, the region of SM50 responsible for competition with P3A directly overlies the factor B binding site. Thus, the BglII-AvaII and SmaI-SalI fragments do not compete (lanes 4 and 7), while the HpaII-BstNI fragment does (lane 8), indicating that the responsible region of SM50 lies between the AvaII site (position -155) and the introduced SmaI site at position -104. Interestingly, the BglII-SmaI fragment competes weakly (lane 5), as if the binding site were located quite near but upstream of position -114. There is a 7 of 8 bp homology between the sequence GCGCAAAC of the P3A binding site of CyIIIa and the sequence GCGCACAC of SM50 (at position -123 to -116), which is probably responsible for binding. P3A is not factor B, as P3 DNA and P3A oligomer cannot compete for factor B binding to SM50 under conditions where SM50 DNA competes very efficiently (data not shown). Furthermore, the

affinity of P3A protein for CyIIIa and SM50 sequences is approximately equal (Thiebaud and Davidson, unpubl. obs.), ruling out the alternative that P3A binds cooperatively with a second protein as part of a more stable factor B complex, as P3 DNA would still be expected to compete for factor B binding. It seems likely that P3A and factor B bind to the same or overlapping sequences, with factor B having either the higher prevalence or affinity in the conditions employed in the gel shift reactions, thus occluding the visualization of the P3A complex. We have not observed additional protein binding sites in this region of SM50, yet have not systematically examined this DNA for P3A binding either.

In addition to the four proteins described above, there is also preliminary evidence that one additional factor binds to the SM50 promoter, in the region between -50 and -20 (dashed box in Fig. 1). The evidence comes from weak DNaseI protections in footprinting reactions at this region, and from occasional complexes seen in gel shift reactions which cannot be attributed to other binding proteins (data not shown). It might be predicted that such a site should exist, simply because most eukaryotic genes have factors important for transcription in the -80 to -20 region (i.e., the TATA-binding proteins).

In Vivo Functional Analysis of Promoter Binding Sites

A series of Bal-31 5' deletions were constructed in the SM50 promoter as a means of identifying important cis-acting promoter elements. All constructs contain the identical sequences at the extreme 5' end, i.e., a HindIII polylinker restriction site, are fused to the CAT coding region at the SalI site, and terminate at the end of the CAT/SV40 domain with a BamHI site. Gel purified DNA fragments derived from these constructs by HindIII-BamHI digest were microinjected into sea urchin eggs, and embryos isolated 2 days later (50 hr) for

analysis of CAT activity. As shown in Figure 5a, deletion of sequences from the BglII site at -438 through position -253 coincided with a gradual increase in CAT activity. The small magnitude of the effect suggests that the explanation may be trivial, i.e., some aberration of the eggs of this experiment which progressively affected the retention of these samples upon microinjection (the amount of CAT DNA per sample has not been measured). Further deletion of the SM50 promoter to -196, eliminating the second binding site of factor A, results in a dramatic 48-fold decrease in CAT expression (relative to the -253 construct). These experiments demonstrate that while deletion of the upstream factor A binding site results in little effect on expression, deletion of both A binding sites has severe consequences. Factor A is therefore a required positively acting protein, and at minimum, one copy of the factor A binding site is required for SM50 expression. It should be noted that the tissue specificity of these deletions, in particular the -253 construct, have not be determined.

A further deletion, to -140, results in a 5-fold rise in expression. This construct eliminates the CCAAT box homology, and may affect the ability of P3A to bind the SM50 promoter as well (see Discussion). It is unlikely that this construct affects the ability of factor B to bind.

The effect of the 10 bp deletion in the factor B binding site was also examined in vivo. As described above, this deletion reduces the affinity of factor B for its cognate sequence. A side by side comparison of the CAT activity generated by the BglII-BamHI fragments of either the unaltered CAT construct, or the deletion construct, shows no difference between the two (Fig. 5b). It is possible that the effect on binding seen in vitro with this deletion construct is not mimicked in vivo, if for example factor B were present in sufficient concentration to saturate this site even at the lower affinity. If binding in vivo is affected by the deletion, the minimum conclusion is that

binding of protein at site B is not a necessary condition for expression of the SM50-CAT transgene at the 50 hr pluteus stage.

Developmental Appearance of the DNA Binding Activities

Protein extracts prepared from nuclei of 7 hr cleavage stage embryos (64 cell stage), 24 hr mesenchyme blastula stage embryos, and 40 hr midgastrula stage embryos, as well as extracts prepared from egg cytoplasm (purified on a heparin agarose column to select for the class of proteins with DNA affinity characteristics) were tested for the presence of DNA binding activity. The SM50 gene is not active during oogenesis or in the egg, as evidenced by the inability to detect transcripts in the egg RNA population; nor is it active at the 16 cell stage at the time of the formation of the micromeres (Killian and Wilt, 1988). Transcripts are first seen at the 120 cell stage, at a level of only several messages per expressing cell (Killian and Wilt, 1988), and accumulate thereafter. The SM50 gene therefore becomes transcriptionally active at some time right around the 64 cell stage.

Probes containing the SM50 factor A and factor B binding sites were examined with these extracts, as shown in Figure 6a-b. It should be emphasized that no quantitation of binding activity has yet been measured for these factors, and that the results presented here are qualitative only. A binding activity for some of the faster migrating complexes of factor A is weakly seen in the egg extract, and from 7 hr on the mature pattern is evident. A complex of factor B of different mobility is seen weakly in the egg extract, while the mature pattern is present at 7 hr and 24 hr, and several additional complexes seen at gastrula stage as well. The complex seen with the egg extract may be a modification of the mature binding activity, new binding activities specific for the same

sequences, or represent wholly new sequence specificities. The probe containing the factor B binding site also showed bands of faster mobility which are not specific, as evidenced by the inability of unlabeled DNA to compete for these bands.

The two binding activities which were identified from the CyIIIa gene, P3A and the CCAAT-binding factor, have previously been analyzed in a similar manner (Calzone et al., 1988), and have a qualitatively similar pattern. Their developmental appearance is summarized in Figure 6c. Briefly, for both probes, complexes of different mobility and of very low prevalence are weakly detectable in the egg cytoplasm. In 7 hr nuclear extracts, complexes of identical mobility and comparable prevalence to those seen with 24 hr extracts are detected which persist through later development. It is assumed that the ability of the P3A and CCAAT binding proteins to interact with the SM50 promoter at these developmental stages is not different than their ability to interact with the CyIIIa elements. Thus, factor A may be maternal, and all four factors which bind to the SM50 promoter are present in mature form at the developmental time of SM50 activation.

DISCUSSION

This is a preliminary report of an ongoing investigation, and clearly there are several issues which need to be addressed in the near future. One is the determination of any additional binding sites which may be in the SM50 promoter. There is preliminary evidence of a binding site close to the transcription initiation site. It has also not been unambiguously determined that there are not protein binding sites in some of the more upstream regions of the promoter, such as between the two factor A elements, despite the apparent absence of complexes seen in the restriction mapping studies. It has been observed that longer probes with more than one binding site can occasionally obscure the presence of second sites (F.C., unpubl. obs.), and thus probes specific for these domains must be prepared. Finally, the exact extent of the binding of factor P3A must be determined. Nonetheless, despite the preliminary nature of these results, the general organization of the SM50 promoter is apparent.

Positive and negative acting elements and factors

It is possible to determine at least a putative function for most of the binding factors described here. The CCAAT factor is likely to be a positive acting factor, simply because it has been so determined in other promoters where it has been studied, including the two sea urchin genes CyIIIa (R. Franks and E. Davidson, unpubl. obs.) and the testes-specific H2b histone gene (Barberis et al., 1987). Factor A is most likely to also be a positively acting factor, as the 5' deletion construct which removes the A binding site results in a dramatic decrease in CAT expression. Whether the upstream A site can effectively

substitute for the more proximal site A is unknown, but at least as far as this preliminary investigation shows, the upstream A site is dispensible. The robust amount of CAT activity seen with the -253 construct can be converted to approximately 10^4 messages per expressing cell, based on measurements (described in Sucov et al., 1988) relating CAT activity and steady state transcript levels. On the other hand, the -197 construct which deletes both A binding sites is calculated to generate only 200 transcripts per expressing cell. It was observed (Sucov et al., 1988) that injected transgenes, present in several hundred copies per cell, generate a steady state message level approximately 20-fold greater than the endogenous genes. It follows that if the endogenous SM50 genes were to carry the deletion of the two factor A binding sites, the amount of message attained might therefore be 10 or fewer transcripts per cell, or approximately what is seen in late cleavage stage embryos (Killian and Wilt, 1988). One possibility, then, is that the presence of factor A serves to increase the level of SM50 expression above some basal level, much as an enhancer might function. Alternatively, the combination of the various factors may determine expression level.

The most interesting factor described here is the P3A protein, which has been tentatively identified as a repressor in the CyIIIa gene (Franks and Davidson, unpubl. obs.). When a CyIIIa promoter-CAT transgene is introduced into sea urchin eggs with an excess of P3 competitor DNA, expression is seen in a variety of tissues in addition to the normal site of expression, the aboral ectoderm. The most likely explanation for this deranged expression pattern is that P3A is a repressor, and that spatially inappropriate expression occurs when the factor is titrated off the CyIIIa gene by in vivo competition. This argument would suggest that P3A is not present in the aboral ectoderm, or that P3A is prevented from binding in the aboral ectoderm (i.e., by specific modification) but not in other tissues.

By *extension*, P3A may be a repressor of SM50 as well. One curious observation, which has not yet been followed up, is the increase in expression observed in the deletion mutant -140 compared with the -196 construct. The approximately 5-fold increase is about what might be expected were this construct to remove a repressor, with the basal level of expression seen in the -196 construct now in effect over a majority of tissues in the embryo. The P3A binding site in SM50 might extend from the 8 bp homology with CyIIIa around position -120 up through -140, and would consequently be compromised by the -140 deletion.

If P3A is a repressor of both SM50 and CyIIIa, any model of gene regulation must account for the exclusive expression of SM50 in the skeletogenic mesenchyme, and of CyIIIa in the aboral ectoderm. One straightforward model for the regulation of both genes might be that P3A and factor A are not active in the aboral ectoderm, and both active in the mesenchyme. CyIIIa is expressed in the aboral ectoderm by virtue of the absence of repression, and SM50 inactive because of the absence of the activator protein factor A. In the skeletogenic mesenchyme, CyIIIa would be repressed by P3A. In order to express SM50, some means of specifically preventing P3A binding to the SM50 gene must be present. One possibility suggested by Figure 1 is that factor B interferes with the ability of P3A to bind the SM50 promoter by acting as a gene specific displacement protein. The high level of expression seen with the factor B deletion construct (Fig. 5b) is explained if factor A is unique to the mesenchyme and P3A is not present at saturating amounts (it is known that P3A can be titrated by in vivo competition [Franks and Davidson, unpubl. obs.]). Alternatively, P3A may be specifically modified in the mesenchyme so that it no longer binds to the SM50 promoter, but continues to recognize the CyIIIa promoter.

Transcription factors and lineage determinants

The experiments described in this chapter are primarily designed to elucidate the regulatory mechanisms governing the expression of the SM50 gene in the skeletogenic mesenchyme. Toward this end, the various factors which bind to the promoter, their developmental profile, and their putative regulatory functions have been described. The distinctive aspect of SM50 expression, of course, is its strictly regulated spatial expression in the micromere progeny, a lineage determined through cytoplasmic localization. It is ultimately the goal of this work to understand the nature of lineage determination and the mechanistic aspects of cytoplasmic localization which result in determination of the micromeres. It is widely held that the factors which cause lineage determination are likely to be DNA binding proteins, which cause (either directly or indirectly) the expression of particular sets of genes and the repression of others, ultimately resulting in a characteristic tissue phenotype. Accordingly, it is plausible (even if simplistic) that one or more of the factors described in this paper are the "determinants" of the mesenchyme lineage. The contrary perspective would be that the factors described herein are regulators of SM50 expression, but the actual determinants of the skeletogenic lineage are different molecules which are ultimately responsible for the appearance of these factors. It can therefore be asked what characteristics are expected of actual lineage determinants subject to cytoplasmic localization, as opposed to regulatory proteins whose appearance is regulated by the localized determinants. For example, it might be expected that localized determinants must be maternal in origin. The temporal appearance of the binding activities (Fig. 6) demonstrates that the mature activities are present at the 64 cell stage (7 hr), just prior to the onset of SM50 transcription at the 120 cell stage (Killian and Wilt, 1988), but except for factor

A, not detectable in the egg cytoplasmic extracts. However, the absence of maternal activity does not rule out a maternal aspect for these proteins, as they may be present as RNA species and localized or selectively translated in the micromeres in early cleavage.

A model was proposed in Sucov et al. (1988) to account for the observation that every cell of the cleavage-stage embryo can become skeletogenic. The model proposes that determinants of the skeletogenic lineage are found in every cell of the early embryo, but modified (and inactive) in all cells except the micromeres. Modification may involve the ability of the proteins to bind DNA, or to activate (or repress) transcription. Preliminary experiments to test this model have been undertaken in regards to the tissue distribution of the factors described in this paper. Nuclear extracts prepared from ectoderm and mesenchyme/endoderm "bags" were assayed by gel shift for the presence of the binding activities. Factors A and B (data not shown) and the CyIIIa binding proteins P3A and CCAAT (F.C., unpubl. obs.) are all found in both the ectoderm and mesenchyme/endoderm fractions. These results are somewhat equivocable, as the purity of these fractions were not directly determined. A nuclear protein preparation made from adult coelomocytes would be a valuable comparison, as these cells are easily prepared in large quantities and do not express the SM50 gene or the CyIIIa gene. It would also be possible, although more difficult, to isolate nuclear proteins from isolated micromeres and isolated primary mesenchyme stage bags. If the appearance of any of the SM50 binding proteins are subject to regulation by skeletogenic lineage determinants, their distribution is expected to be regionally confined. On the other hand, if these proteins are the lineage determinants themselves, they may be expected to be found throughout the embryo.

MATERIALS AND METHODS

Protein Extracts. Nuclei were isolated from freeze-thawed sea urchin embryos at migratory mesenchyme blastula stage (27 hr of development), and nuclear protein isolated by 0.4M $(NH_4)_2SO_4$ extraction, exactly as described by Calzone et al. (1988). The yield of protein from $2x10^7$ embryos, as determined by absorbance at 260 and 280 nm, was 1.9 mg, and was stored at -70 at a concentration of 2.5 mg/ml. Extracts used for developmental time points were isolated in an identical fashion and provided by F.C. (Calzone et al., 1988). The egg extract was prepared by heparin agarose chromatography of egg cytoplasm by F.C. The ectoderm and mesenchyme/endoderm "bag" fractions were prepared by F.C. as described by Harkey and Whiteley (1980).

Gel shift assays. DNA restriction fragments were gel isolated and labeled by incorporation of nucleotides (specific activity of 3000 Ci/mmol) by incubation with Klenow polymerase. For restriction mapping studies, fragments labeled at one end only were isolated and further restricted. In these reactions, then, additional SM50 DNA is present but is unlabeled. Gel shift reactions contained approximately 0.1-1.0 ng of labeled DNA (20,000 - 50,000 CPM), 6 ug of nonspecific DNA, up to 100 ng of unlabeled specific competitor DNA in competition reactions, and 1-5 ug of nuclear protein, in a reaction volume of 10 µl containing 2 µl of 5x buffer (5x buffer is 50 mM Hepes pH=7.9, 2.5 mM DTT, 75 ug/ml yeast RNA, 25 mM MgCl2, and 330 mM KCl), 2 µl of 10% glycerol/0.1% NP40, and 2 µl of protein plus buffer. Reactions were incubated on ice for 10 min, loaded on a 5% polyacrylamide gel with 0.5x TBE running buffer and run at 200 V, then dried and exposed for autoradiography. For quantitation, bands were excised from the dried gels and scintillation counted. The nonspecific DNAs

used for each gel shift were: factor A, P3A, and CCAAT factor, poly d(A-T)/d(A-T); factor B, poly d(I)/d(C).

Construction of 5' Bal-31 deletions. A construct containing the 2200 bp region upstream of the SM50 transcription site fused to the CAT coding region (2.2PL-CAT; Sucov et al., 1988), was restricted with BglII and then treated with Bal-31 for varying extents. After organic extraction, the DNA was digested with XhoI, which cuts in the polylinker region 5' of the SM50 sequences. The ends were made blunt with Klenow, then religated under conditions to promote recircularization. The extent of deletion was determined by sequencing.

Construction of B2SΔ10-CAT. Plasmid B2S-CAT, a subclone of the BglII-BamHI fragment of SM50-CAT (Sucov et al., 1988) was restricted at the unique BstXI site found at -114 of the SM50 promoter, and briefly treated with Bal-31 nuclease. The ends were made blunt with Klenow and deoxynucleotides, then religated under conditions to promote circularization. The extent of deletion was determined by sequencing.

Injections and CAT assays. Gel purified SM50-CAT fragments were isolated from either BglII-BamHI restricted B2S-CAT or B2SΔ10-CAT, or from HindIII-BamHI restriction digests of the Bal-31 5' deletion series. In the latter case, the HindIII site derives from vector polylinker sequences. DNA was microinjected into sea urchin eggs as described in Sucov et al. (1988), fertilized in situ, and harvested at the 50 hr early pluteus stage. CAT assays followed standard procedures, as described in Sucov et al. (1988).

REFERENCES

Angerer, R. C. and Davidson, E. H. Molecular indices of cell lineage specification in sea urchin embryos. *Science* **226,** 1153-1160 (1984).

Barberis, A. Superti-Furga, G. and Busslinger, M. Mutually exclusive interactions of the CCAAT-binding factor and of a displacement protein with overlapping sequences of a histone gene promoter. *Cell* **50,** 347-359 (1987).

Benson, S., Sucov, H., Stephens, L., Davidson, E., and Wilt, F. A Lineage-specific gene encoding a major matrix protein of the sea urchin embryo spicule. I. Authentication of the cloned gene and its developmental expression. *Dev. Biol.* **120,** 499-506 (1987).

Calzone, F. J., Thézé, N., Thiebaud, P., Hill, R., Britten, R. J. and Davidson, E. H. Developmental appearance of factors that bind specifically to *cis*-regulatory sequences of a gene expressed in the sea urchin embryo. Submitted for publication (1988).

Cox, K. H., Angerer, L. M., Lee, J. J., Davidson, E. H., and Angerer, R. C. Cell lineage-specific programs of expression of multiple actin genes during sea urchin embryogenesis. *J. Mol. Biol.* **188,** 159-172 (1986).

Davidson, E. H. *Gene Activity in Early Development*, 3rd Ed., Academic Press (1986).

Harkey, M. A. and Whiteley, A. H. Isolation and differentiation of echinoid primary mesenchyme cells. *Wilhelm Roux's Arch.* **189,** 111-122 (1980).

Killian, C. E. and Wilt, F. H. The accumulation and translation of a spicule matrix protein mRNA during sea urchin embryo development. Submitted for publication (1988).

Okazaki, K. Spicule formation by isolated micromeres of the sea urchin embryo. *Amer. Zool.* **15**, 567-581 (1975).

Schroeder, T. Expressions of the prefertilization polar axis in sea urchin eggs. *Dev. Biol.* **79**, 428-443 (1980).

Sucov, H. M., Benson, S., Robinson, J. J., Britten, R. J., Wilt, F. and Davidson, E. H. A lineage-specific gene encoding a major matrix protein of the sea urchin embryo spicule. II. Structure of the gene and derived sequence of the protein. *Devel. Biol.* **120**, 507-519 (1987).

Sucov, H. M., Hough-Evans, B. R., Franks, R. R., Britten, R. J. and Davidson, E. H. A regulatory domain that directs lineage-specific expression of a skeletal matrix protein gene in the sea urchin embryo. Submitted for publication (1988).

Figure 1. Sequence of the SM50 promoter, and location of protein binding sites. The sequence shown extends from a genomic BglII site at -438 to an introduced SalI site at +120. The unique transcription initiation site is at +1; the SM50 translation initiation codon is at +110. Protein binding sites are indicated as boxes. The two binding sites of factor A are resolved to 25 and 30 bp restriction fragments; a shared 11 of 15 bp sequence is underlined. The binding site of factor B has been defined by DNaseI footprinting; the location of the 10 bp deletion (and the created SmaI restriction sequence CCCGGG) in the construct B2SΔ10-CAT is indicated. For the CCAAT factor, the 14 bp homology with the element in the CyIIIa gene is boxed. The binding site of P3A is only resolved to the 45 bp AvaII-SmaI fragment. There is preliminary evidence for a binding site between -50 and -20, as indicated by the dotted box. Arrowheads show the 5' sites of Bal-31 deletion series constructs.

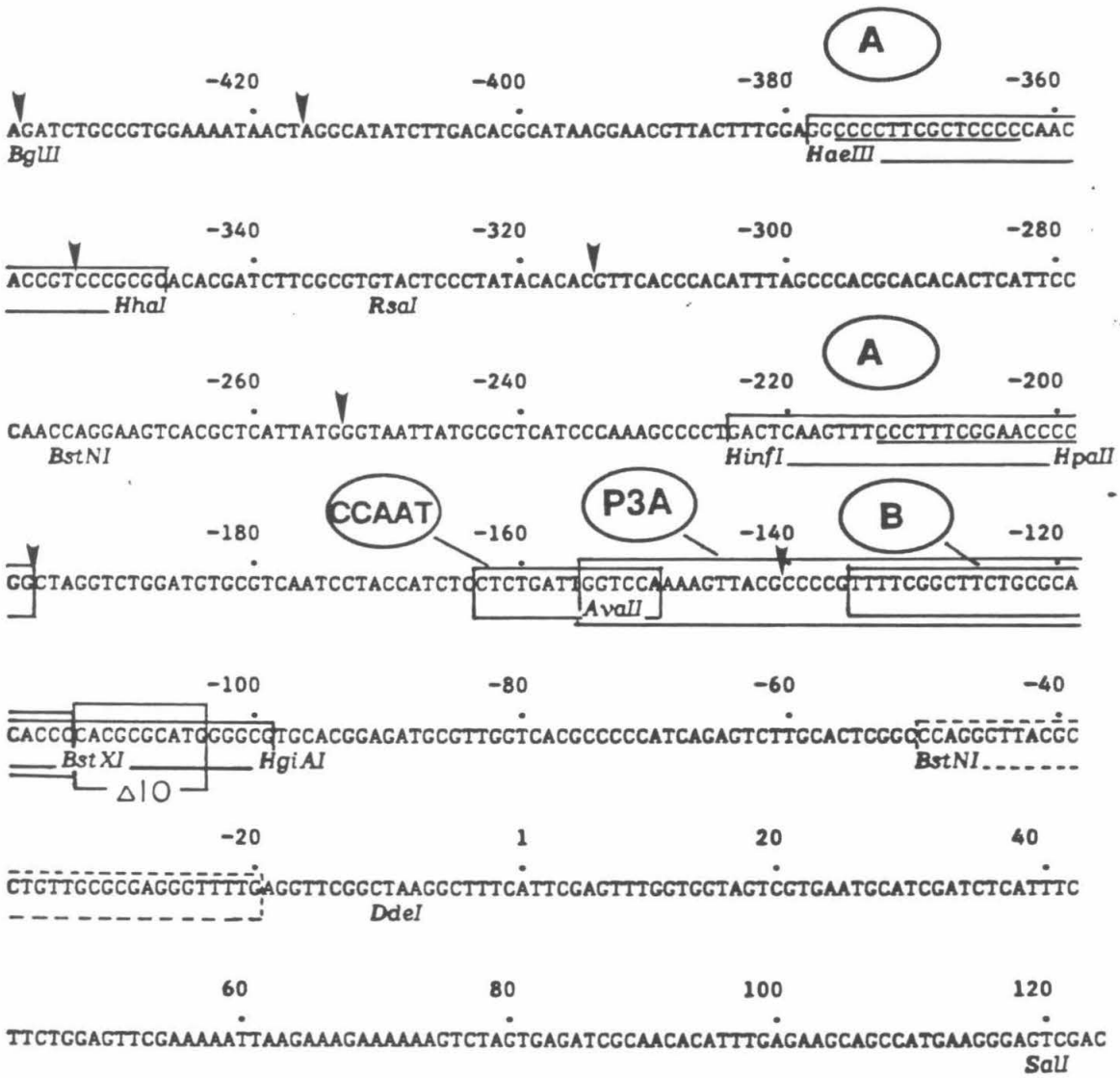

A
-420
-400
-380
-360
AGATCTGCCGTGGAAAATAACTAGGCATATCTTGACACGCATAAGGAACGTTACTTTGGAGGCCCCTTCGCTCCCCCAAC
BglII
HaeIII
-340
-320
-300
-280
ACCGTCCCGCGCACACGATCTTCGCGTGTACTCCCTATACACACGTTCACCCACATTTAGCCCACGCACACACTCATTCC
HhaI
RsaI
A
-260
-240
-220
-200
CAACCAGGAAGTCACGCTCATTATGGGTAATTATGCGCTCATCCCAAAGCCCCTGACTCAAGTTTCCCTTTCGGAACCCC
BstNI
HinfI
HpaII
CCAAT
P3A
B
-180
-160
-140
-120
GGCTAGGTCTGGATGTGCGTCAATCCTACCATCTCCTCTGATTGGTCCAAAAAGTTACGCCCCGTTTTCGGCTTCTGCGCA
AvaII
-100
-80
-60
-40
CACCCCACGCGCATGGGGCGTGCACGGAGATGCGTTGGTCACGCCCCCATCAGAGTCTTGCACTCGGCCCAGGGTTACGC
BstXI
HgiAI
BstNI
Δ10
-20
1
20
40
CTGTTGCGCGAGGGTTTTGAGGTTCGGCTAAGGCTTTCATTCGAGTTTGGTGGTAGTCGTGAATGCATCGATCTCATTTC
DdeI
60
80
100
120
TTCTGGAGTTCGAAAAATTAAGAAAGAAAAAAGTCTAGTGAGATCGCAACACATTTGAGAAGCAGCCATGAAGGGAGTCGAC
SalI

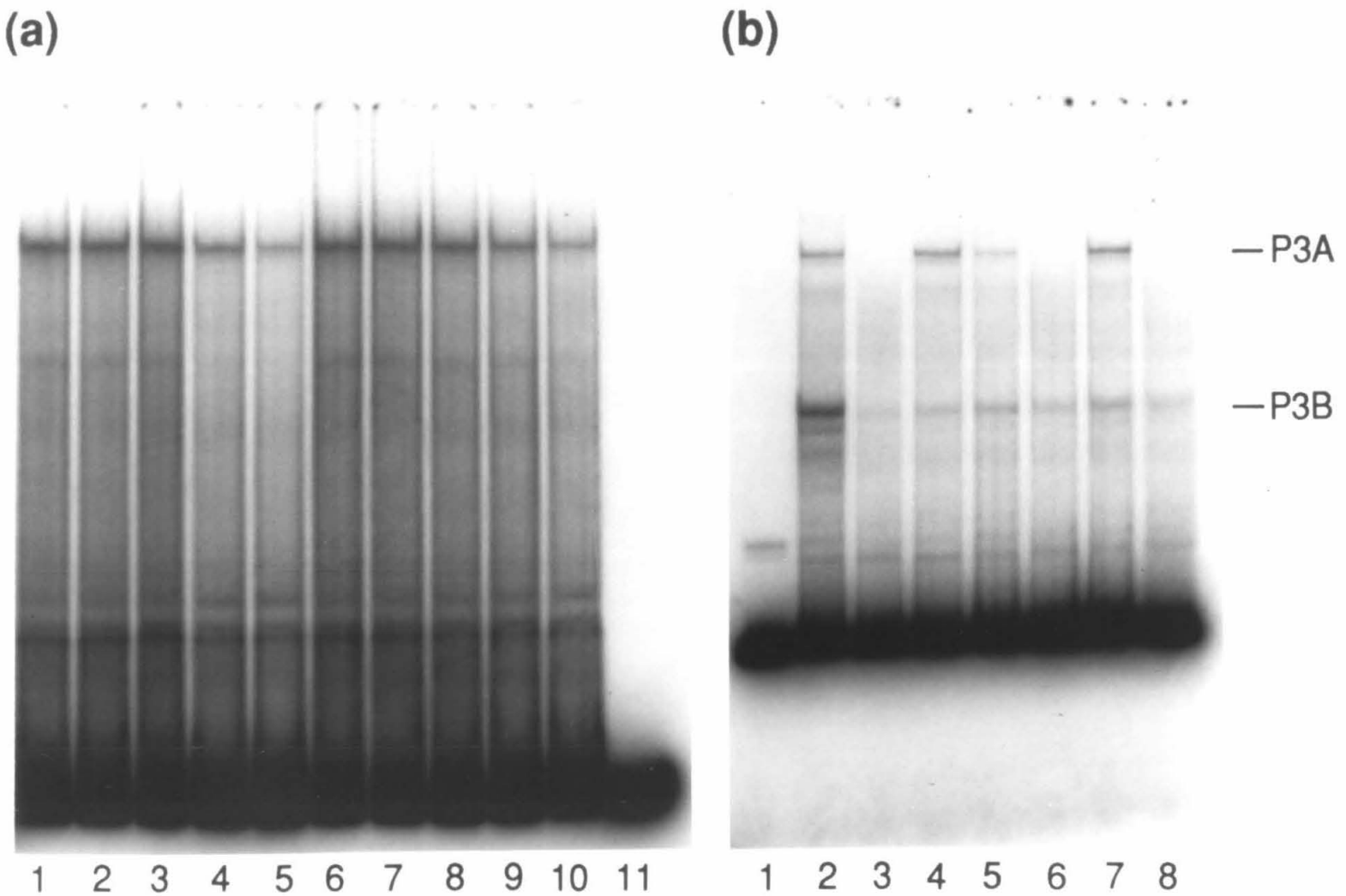
Figure 2. *Binding of* factor A. (a-c) Gel shift mapping of the factor A binding sites. The BglII-AvaII fragment of the SM50 promoter was labeled selectively at the BglII site (a and b) or at the AvaII site (c), then digested with the indicated restriction enzymes. Each lane contains approximately 30,000 CPM of probe, either without (-) or with (+) the addition of 2.5 ug of blastula stage nuclear protein to the reaction. The third lane of (a) also contains approximately a 50-fold molar excess of unlabeled BglII-AvaII fragment as a specific competitor. (d) Competition experiment for binding of factor A to the two factor A binding sites. The labeled BglII-RsaI restriction fragment, containing the upstream site (lanes 1-4), or the BstNI-AvaII restriction fragment containing the downstream site (lanes 5-8), were mixed with nuclear protein (lanes 2-4 and 6-8) and either no competitor (lanes 2 and 6), a 3-fold molar excess of unlabeled BglII-RsaI fragment (lanes 3 and 7), or a 7-fold molar excess of competitor (lanes 4 and 8). Lanes 1 and 5 show the probes without the addition of protein or competitor.

(a)
Bgl II (-438)
Rsa I (-331)
Bst NI (-275)
Hinf I (-224)
Hpa II (-200)
Ava II (-155)
Ava II Hpa II Hinf I Bst NI Rsa I
(b)
Bgl II (-438)
Hae III (-378)
Hhe I (-350)
Rsa I (-331)
Rsa I Hha I Hae III
(c)
Bgl II (-438)
Hae III (-378)
Rsa I (-331)
Bst NI (-275)
Hinf I (-224)
Hpa II (-200)
Ava II (-155)
Bgl II Hae III Rsa I Bst NI Hinf I Hpa II
(d)
1 2 3 4 5 6 7 8

Figure 3. Binding of factor B. (a-b) Gel shift mapping experiment. SM50 promoter fragments were labeled at either the SalI site (a) or the AvaII site (b) and further restricted as indicated. 2.5 ug of nuclear protein was added to lanes marked (+); no protein was added to lanes marked (-). (c) Effect of the 10 bp deletion on binding of factor B. The AvaII-SalI fragment of either the unaltered SM50 promoter (lanes 1-2) or the deletion mutant carrying a 10 bp deletion as described in the text (lanes 3-4) were mixed with either 2.5 ug (lanes 1 and 3) or 5 ug (lanes 2 and 4) of nuclear protein. The ratio of the affinity of factor B for the wild type and deletion probe is to a first approximation the ratio of the amount of complex formed at a given level of protein input, as the amount of DNA is constant between the reactions. This ratio is 40 for 2.5 ug protein, and 28 for 5 ug protein, for an average of approximately 35-fold difference.

(a)
Bgl II (-438)
Hpa II (-200)
Ava II (-155)
Bst XI (-144)
Hgi AI (-99)
Bst NI (-50)
Dde I (-11)
Sal I (+120)
Bgl II Hpa II Ava II Bst XI Hgi AI Bst NI Dde I
(b)
Ava II (-155)
Bst NI (-50)
Dde I (-11)
Sal I (+120)
Sal I Dde I Bst NI
(c)
1 2 3 4

Figure 4. Gel shift competition of SM50 promoter fragments for CyIIIa binding factors. (a) Competition for the CCAAT-binding factor between the CyIIIa gene element and SM50 fragments. A 53 bp CyIIIa probe (containing no additional protein binding sites) was mixed with protein and varying amounts of unlabeled B2SΔ10-CAT BglII-SmaI fragment (lanes 2-5), or the HpaII-BstNI fragment of SM50 (lanes 7-10) prior to electrophoretic separation. The amount of SM50 DNA added increases from no SM50 DNA added (lanes 1 and 6) to approximately 10x the molar amount of the CyIIIa DNA (lanes 5 and 10). (b) Gel shift competition for the P3A factor. The binding of protein to a probe derived from the CyIIIa promoter containing the binding sites of two distinct proteins, P3A and P3B (an octamer binding protein), competed with fragments of the SM50 promoter. Lane 1 is the probe without the addition of protein; protein was added to lanes 2-8. Lane 2 contains no SM50 competitor; the remaining lanes contain approximately a 50-fold molar excess of the following unlabeled SM50 promoter restriction fragments: the entire BglII-SalI fragment (lane 3); the BglII-AvaII fragment (lane 4); the BglII-SmaI fragment of the deletion clone B2SΔ10-CAT (lane 5); the HpaII-SalI fragment (lane 6); the SmaI-SalI fragment of B2SΔ10-CAT (lane 7); the HpaII-BstNI fragment (lane 8). All fragments were derived from the wild type template except for lanes 5 and 7, which utilized the deletion construct B2Sdelta10. The intensity of the P3B bands are diminished in all cases compared to lane 2, because of the addition of the large mass of competitor DNA. The signal of the P3B band therefore serves as an internal standard for this semispecific effect. (c) Diagrammatic representation of the P3A gel shift competition. At top is the 560 bp BglII-SalI SM50 promoter, and below are the fragments used as competitors for P3A, numbered according to the order of lanes in panel (b). The bracket shows the region of SM50 which competes for P3A binding.


(a)
(b)
P3A
P3B
1 2 3 4 5 6 7 8 9 10 11
1 2 3 4 5 6 7 8


(c)
Bgl II (-438)
Hpa II (-200)
Ava II (-155)
[Sma I (-114/-103)]
Bst NI (-50)
Sal I (+120)
lane
competition for P3A
3 +
4 -
5 weak
6 +
7 -
8 +

Figure 5. In vivo expression of SM50 deletion constructs. (a) 5' deletion series. Embryos were grown 50 hr from eggs injected with HindIII-BamHI fragments isolated from the 5' deletion series constructs and assayed for CAT activity. The SM50 5' endpoint of each fragment is indicated underneath. Each sample contains 30-40 transgenic embryos. The amount of CAT activity for each construct, expressed in terms of units CAT x 10^{-5}/embryo, is: 1.4 (-437), 2.9 (-416), 4.3 and 3.9 (-353), 5.6 and 5.7 (-314), 5.7 (-253), 0.12 (-196), and 0.57 (-140). The last lane contains uninjected control embryos. (b) Effect on SM50 expression of a 10 bp deletion in the factor B binding site. CAT assays of two samples of embryos expressing the unaltered construct B2S-CAT (1-2), and three samples of embryos expressing the deletion construct B2SΔ10-CAT (3-5), all collected 50 hr after fertilization, are shown. Each sample contains 50 transgenic embryos. The amount of CAT activity in units CAT x 10^{-5}/embryo is: 4.2 and 5.0 (B2S-CAT), and 4.0, 4.6, and 6.4 (B2SΔ10-CAT).

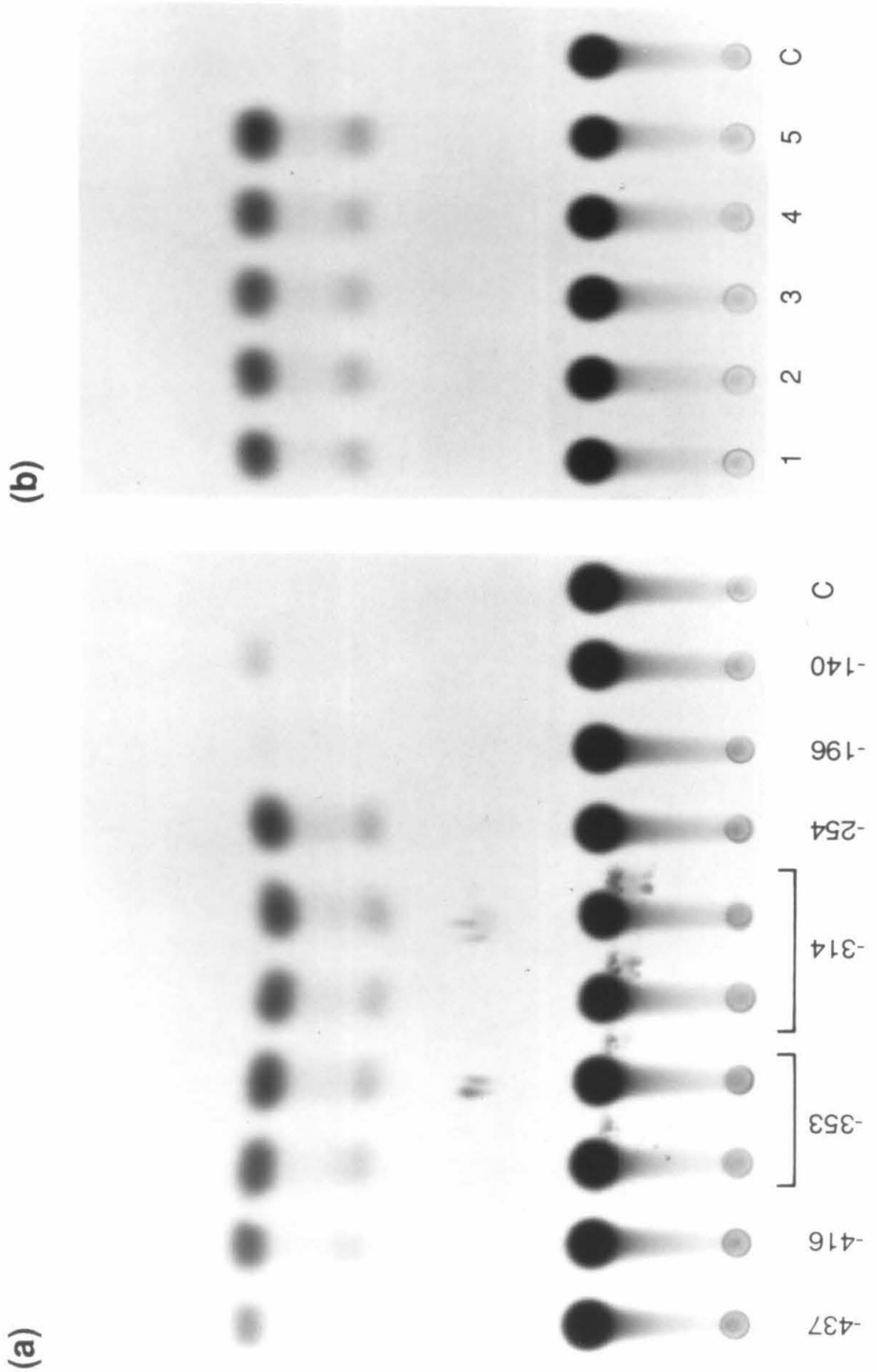
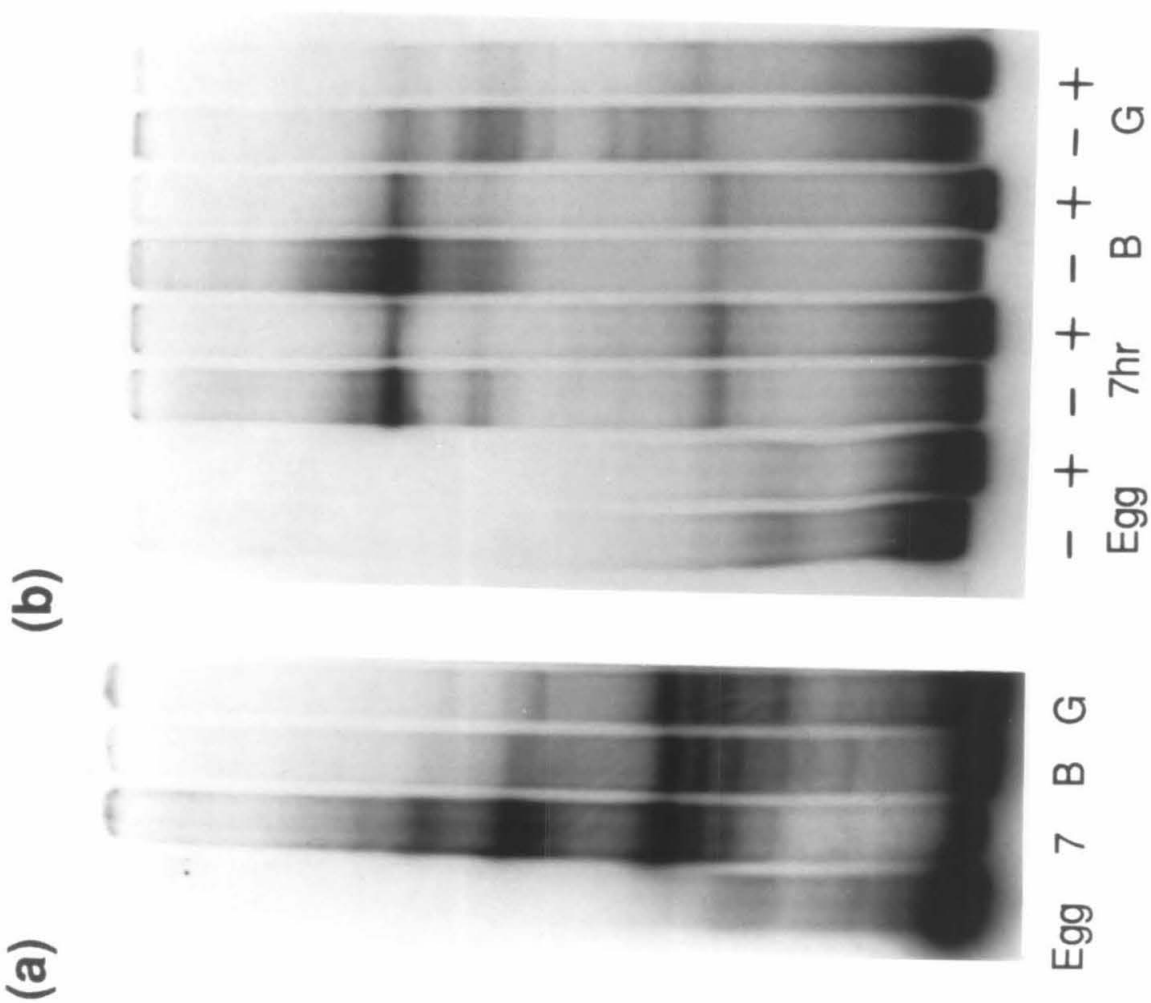
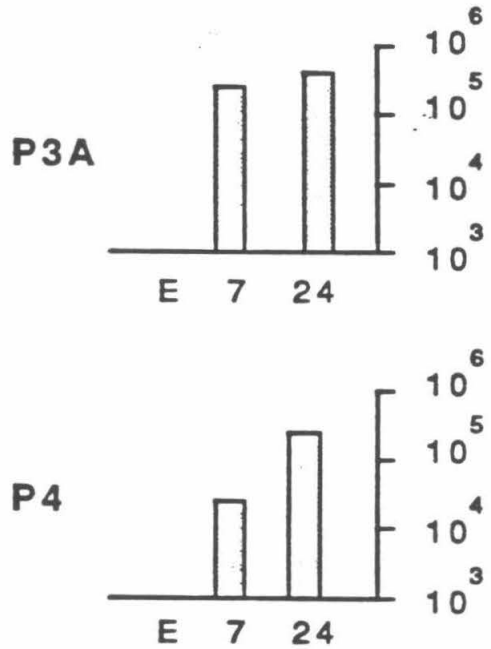

(a)
C
-140
-196
-254
-314
-353
-416
-437
(b)
C
5
4
3
2
1

Figure 6. Developmental time course of appearance of factors which bind the SM50 promoter. (a) Developmental appearance of factor A. The BglII-AvaII probe (which contains both of the A binding sites) was mixed with egg cytoplasmic protein (E), and nuclear proteins from embryos of: 7 hr (7), 24 hr blastula (B), and 40 hr gastrula (G). (b) Developmental appearance of factor B. The AvaII-SalI probe was mixed with developmentally staged proteins as in panel (a), either without (-) or with (+) a 40-fold excess of unlabeled competitor fragment present so as to indicate sequence specific vs. nonspecific complexes. Only those complexes which are reduced in intensity by competition are real. (c) Developmental appearance of the CyIIIa factors P3A and the CCAAT-binding factor (P4). Data are redrawn from Table 2 of Calzone et al. (1988), and indicate the prevalence of these factors in egg cytoplasm, 7 hr nuclei, and 24 hr nuclei. The log scale at right is in molecules/embryo. Not shown for both factors are weak complexes of different mobility which are present in the egg at prevalences less than ten-fold that seen for the mature complexes at 24 hr.


(a)
(b)
Egg 7 B G
– + – + – + – +
Egg 7hr B G

(C)



APPENDIX

A Lineage-Specific Gene Encoding a Major Matrix Protein of the Sea Urchin Embryo

I. Authentication of the Cloned Gene and Its Developmental Expression

DEVELOPMENTAL BIOLOGY 120, 499–506 (1987)

A Lineage-Specific Gene Encoding a Major Matrix Protein of the Sea Urchin Embryo Spicule

I. Authentication of the Cloned Gene and Its Developmental Expression[1]

STEVE BENSON,*† HENRY SUCOV,‡ LAURIE STEPHENS,* ERIC DAVIDSON,‡ AND FRED WILT*

*Department of Zoology, University of California, Berkeley, California 94720; †Department of Biology, California State University, Hayward, Hayward, California 94542; and ‡Division of Biology, California Institute of Technology, Pasadena, California 91125

Received September 4, 1986; accepted in revised form November 6, 1986

The developing sea urchin embryo forms endoskeletal $CaCO_3$ containing spicules which are elaborated by the primary mesenchyme cells, descendants of the micromeres, beginning at gastrulation. In this and the accompanying paper [H. M. Sucov, S. Benson, J. J. Robinson, R. J. Britten, F. Wilt, and E. H. Davidson (1987) *Dev. Biol.* 120, 507–519] the isolation and characterization of a gene that encodes a 50-kDa spicule matrix glycoprotein that we call SM50 are described. A cloned cDNA isolated from a λgt11 library was used in hybrid-selected translation and hybrid arrest of translation experiments to verify that the cDNA encodes a spicule matrix protein. The cognate RNA transcript encodes a 50-kDa protein which is precipitated by polyclonal antisera against spicule matrix proteins and is present only in polyadenylated RNA at stages known to be making a spicule. The cloned cDNA sequence described in the accompanying paper was used to follow the time of expression of the cognate gene by RNA blotting analysis. The 2.2-kb mRNA is first detected at late cleavage stages and rapidly accumulates as the primary mesenchyme forms, reaching an apparent maximum concentration in the late gastrula and pluteus stages. The cDNA was also used to identify the cells that contain the transcripts by hybridization *in situ*. Hybridization to cellular transcripts is first detected in primary mesenchyme cells as they enter the blastocoel, and transcripts are confined to these cells during spicule formation and subsequent development.

INTRODUCTION

The sea urchin embryo forms calcareous endoskeletal spicules during gastrulation, which increase in size and complexity as the pluteus larva matures. The embryology of these skeletal elements has been the subject of numerous investigations (Okazaki, 1975a). At the fourth cell division a quartet of small micromeres is formed at the vegetal pole, and these micromeres subsequently give rise to the primary mesenchyme cells that invade the blastocoel just prior to the archenteron formation. The primary mesenchyme cells migrate along the basal lamina of the blastocoel and begin to fuse into a syncytial array (Okazaki, 1975b). Within the syncytium an intracellular vacuole forms. $CaCO_3$ is deposited within the vacuole in association with the matrix, thereby forming the hard skeletal rod of the spicule (Gibbins *et al.*, 1969; Benson *et al.*, 1983). Spicules are first observed in the latter half of gastrulation and increase in size and complexity as the larva matures (reviewed by Horstadius, 1973; Okazaki, 1975a). Okazaki (1975b) has shown that micromeres may be isolated and, when cultured, give rise to spicules autonomously.

The spicule has recently been purified and the proteins that constitute the organic matrix of the spicule have been characterized (Benson *et al.*, 1986). The matrix is an assemblage of at least 10 acidic glycoproteins, four of which have apparent molecular weights of 47, 50, 57, and 64 kDa and comprise over 70% of the mass. A polyclonal antibody which reacts specifically with the matrix proteins was raised to follow the accumulation and localization of these glycoproteins during embryogenesis. The antigens are found only in primary mesenchyme cells and increase in concentration during overt spicule formation.

The antibody afforded an opportunity to screen an expression vector cDNA library to determine if cDNAs and genomic DNA sequences that encode the spicule matrix proteins could be identified. Such cloned DNAs would be valuable tools for the analysis of tissue-specific gene expression in a well-understood cell lineage of developing sea urchin embryos.

We have isolated and characterized one such cDNA that encodes a predominant spicule matrix glycoprotein. In this report we describe the experiments that show that this cDNA encodes a spicule protein, and we de-

[1] Supported by grants from the National Institutes of Health (HD 15043 to F.H.W; HD 05753 to E.H.D.) and the University of California (to F.H.W.).

0012-1606/87 $3.00

scribe the time course of expression of the gene and its cellular localization. In the accompanying paper (Sucov *et al.*, 1987) we describe the structure of the gene, the sequence of the cloned cDNA, and the protein it encodes as deduced from the cDNA sequence.

METHODS

Embryos and RNA preparation. Embryos of *Strongylocentrotus purpuratus* were fertilized and cultured by standard methods (Hinegardner, 1967) in Millipore-filtered seawater containing 10 μg/ml of gentamycin sulfate at a concentration of 3.6×10^3 embryos/ml at 15°C. RNA was prepared by washing the embryos twice in cold 1.5 M glucose followed by extraction with guanidinium chloride by using the method of Childs *et al.* (1979). Poly(A)$^+$ RNA was prepared from total embryo RNA on oligo(dT)-cellulose (Collaborative Research, grade T-3) by using the method of Aviv and Leder (1972). All RNA samples were translated in a nuclease-treated reticulocyte lysate to monitor RNA integrity (Pelham and Jackson, 1976).

Clones and probes. The cDNA clone (pHS72) used in all these experiments contains a 1.3-kb insert and was isolated as described in the accompanying paper (Sucov *et al.*, 1987). The eukaryotic insert, bounded by *Eco*RI linkers, was excised from its PUC8 vector and subcloned into the transcription vector PS65 (Melton *et al.*, 1984) in both orientations. Plasmid was grown in *Escherichia coli* strain LE 392 and isolated after amplification with chloramphenicol using the alkaline sodium dodecyl sulfate method of Birnboim and Doly (1979), followed by banding in ethidium bromide–CsCl gradients by using standard methods. Single-stranded ^{32}P-labeled RNA transcripts were synthesized by SP6 polymerase (Promega Biotech and Boehringer-Mannheim) (Melton *et al.*, 1984) after linearization of the plasmid downstream from the insert by digestion with *Hin*dIII.

In situ hybridization. *In situ* hybridization of sections of fixed embryos was carried out exactly as described by Cox *et al.* (1984). ^{3}H-Labeled single-stranded probes were synthesized from the PS65 transcription vector containing the cDNA insert (in both orientations) using [^{3}H]UTP and had a specific activity of 10 Ci/mmole of probe nucleotide. Emulsion-coated slides were exposed for 10 days before development.

RNA blots. RNA blotting was carried out by electrophoresis of RNA in 1.2% agarose gels containing formaldehyde followed by transfer to nitrocellulose (Thomas, 1980). RNA blots were hybridized for 16 hr to single-stranded RNA probes (with a specific activity of about 2×10^8 dpm/μg of probe) in 50% formamide, 5× SSC (SSC = 0.15 *M* NaCl, 0.015 *M* Na-citrate), 1 m*M* EDTA, 50 m*M* phosphate buffer, pH 6.8, 2× Denhardt's solution (Denhardt, 1966) 0.1% SDS, and 200 μg/ml of single-stranded salmon sperm DNA at 55°C. Blots were washed in this same buffer without formamide or salmon sperm DNA, but with the addition of 0.1% sodium pyrophosphate, at 68°C for 1 hr. Subsequently the blots were washed at 68°C for 30 min in 50 m*M* NaCl, 5 m*M* Tris, pH 8, 0.5 m*M* EDTA, 0.1% SDS, and 0.1% pyrophosphate, followed by three 20-min washes in 0.1× SSC at 65°C. Sometimes blots were also treated with 0.05 μg/ml of RNase A in 2× SSC at room temperature for 10 min followed by washing in 0.1× SSC, 0.1% SDS, at 50°C. Blots were autoradiographed either using preflashed Kodak X-O-Mat X-ray film and a Dupont Cronex intensifying screen or without a screen using unexposed film (Laskey and Mills, 1975).

Hybrid-selected translation. Experiments designed to select the mRNA corresponding to pHS72 cDNA from poly(A)$^+$ RNA of different stages of development were based on the methods of Riccardi *et al.* (1979; cf. Maniatis *et al.*, 1982). DNA was applied to nitrocellulose in 10× SSC and then baked in a vacuum oven at 80°C for 2 hr. Either electrophoretically purified 1.3-kb cDNA excised from PS65 by *Eco*RI or the entire plasmid bearing the insert was used, as well as linear vector DNA as a control. Hybridization of poly(A)$^+$ RNA was carried out at a concentration of 20 μg/100 μl of 65% formamide, 10 m*M* Pipes, pH 6.4, 0.2% SDS, 0.4 *M* NaCl, and 100 μg/ml of *E. coli* tRNA. Hybridization was for 3–16 hr at 50°C on filters containing either 2 μg of plasmid DNA or 0.5 μg of purified insert. Filters were washed as described (Maniatis *et al.*, 1982), and the RNA was eluted in water at 98°C for 60 sec followed by fast cooling and ethanol precipitation. In some cases the RNA selected by hybridization was digested with RNase free DNase (Q-1, Promega) at 15 μg/ml for 15 min at room temperature followed by phenol extraction and ethanol precipitation. This improves the signal/noise ratio of the result considerably. The RNA selected by hybridization was translated in a reticulocyte lysate (Pelham and Jackson, 1976) containing [^{35}S]methionine.

Hybridization arrest. Hybrid arrest of translation was carried out following the original method of Patterson *et al.* (1977). The transcription vector, PS65, either with or without the eukaryotic insert, was linearized by digestion with *Eco*RI, which completely excises the 1.3-kb insert from the 3-kb plasmid. Six micrograms of total gastrula RNA was incubated with 0.06 or 0.6 μg of plasmid and the mixture was heated to 100°C for 30 sec, then quick chilled in dry ice/methanol. The samples were adjusted to a final volume of 35 μl containing 80% deionized formamide, 10 m*M* Pipes buffer, pH 6.4, 0.4 *M* NaCl. The mixture was incubated at 48°C for two hr, 5 μg of tRNA was added, and the sample was divided into two equal portions, one of which was heated to 100°C for 60

sec and quick chilled. Both samples were then adjusted to contain 0.2 *M* sodium acetate, pH 5.5, and were precipitated with 2.5 vol of ethanol overnight at −20°C. RNA was collected by centrifugation, washed twice with 70% ethanol, and dried in a vacuum. The RNA was dissolved in 7 μl of sterile distilled water and translated in a nuclease-treated reticulocyte lysate for 90 min.

Immunoprecipitation and gel electrophoresis. Immunoprecipitations were carried out on translation products of the reticulocyte lysate using a procedure originally described by Rohrschneider *et al.* (1979) and modified by Adelson (1986). The 50-μl translation mixture was diluted to 0.5 ml with RIPA (radio immunoprecipitation assay) buffer: 10 m*M* Tris-HCl, pH 7.2, 0.15 *M* NaCl, 1% Na-deoxycholate (DOC), 1% Triton X-100, 0.1% Pepstatin A, 10 m*M* iodoacetamide, 1 m*M* phenylmethylsulfonyl fluoride (PMSF). PMSF was added to the buffer immediately before use. Twenty microliters of the preimmune IgG was added and the mixture was incubated at room temperature for 10 min and on ice for 30 min. Two hundred microliters of a 10% (w/v) suspension of fixed washed *Staphylococcus aureus* (Zysorbin, Zymed Labs, Burlingame, CA) was added to the mixture and incubated on ice for 30 min with frequent agitation. The suspension was centrifuged at 12,000*g* for 5 min and the supernatant was removed. Five microliters of rabbit antispicule matrix IgG (Benson *et al.*, 1986) was added, incubated, and precipitated with *Staphylococcus* protein A as described above for the preimmune IgG. The pellet was washed once with RIPA and once each with buffers R-1, R-2, and R-3. (R-1: 2 *M* NaCl, 0.5% NaDOC, 1% NP-40, 10 m*M* Tris-HCl, pH 7.2. R-2: 1% NP-40 in 0.02 *M* Na-Phosphate, 0.15 *M* NaCl. R-3: 10 m*M* Tris-HCl, pH 7.2, 50 m*M* NaCl, 0.5% NP-40.) After a final wash with sterile distilled water the pellets were resuspended in 35 μl of 1.25× SDS sample buffer and heated to 85°C for 5 min with frequent agitation. The mixture was centrifuged at 16,000*g* for 10 min to remove the bacterial debris and the supernatants were subjected to electrophoresis on 12.5% polyacrylamide gels (Laemmli, 1970), as modified by Dreyfus *et al.* (1984). Visualization of radioactive protein was achieved by fluorography as described by Laskey and Mills (1975).

RESULTS

Hybrid-Selected Translation

The cDNA clone, pHS72, was isolated from the expression vector library using antibodies raised against spicule matrix proteins. This clone included an open reading frame which predicted an amino acid composition consistent with the properties expected of the spicule matrix (Benson *et al.*, 1986; Sucov *et al.*, 1987). The cDNA was hybridized to RNA from several stages of development. Hybridizing RNA was translated, *in vitro*, and the products were immunoprecipitated. Figure 1 shows the results of this kind of experiment. In this instance, poly(A)$^+$ RNA from a stage that does not make spicules (the 8-cell stage) was hybridized to vector alone (lane C) or to vector containing the cDNA insert (lane D). In a similar fashion RNA from a late-stage embryo (midgastrula) containing spicules was hybridized to vector alone (lane A) or to vector containing the eukaryotic insert (lane B). The selected RNA was translated in a reticulocyte lysate. Though the lysate had been nuclease-treated and the background was low, long exposures were required to see an mRNA-encoded product against the background bands of the lysate. The products of the cell-free translation were subjected to immunoprecipitation, thereby purifying the putative spicule matrix protein. Figure 1 shows that only gastrula RNA hybrid selected with vector containing insert yields immunoprecipitated translation product (lane B). This protein has an apparent molecular weight of about 50 kDa, which corresponds well with the molecular weight expected of deglycosylated and unprocessed protein product of the sequence described in the accompanying paper (Sucov *et al.*, 1987). This band is found in experiments using poly(A)$^+$ RNA from early-gastrula, late-gastrula, and prism stages. It is not found in RNA prepared from unfertilized eggs, the 2-cell stage, the 8-cell

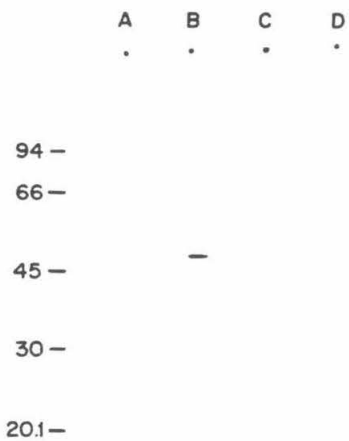


FIG. 1. Translation of RNA hybridized to pHS72 RNA from either the midgastrula (A, B) or 8-cell (C, D) stage was hybridized to nitrocellulose containing either PS65 (vector) DNA or vector containing the eukaryotic insert, pHS72 (B, D). The hybridized RNA was eluted and translated in a reticulocyte lysate containing [^{35}S]methionine, and the products were immunoprecipitated with antibody against spicule matrix protein. The immunoprecipitates were subjected to electrophoresis and fluorography.

stage, or very early blastulae. It is found only in RNA samples hybridized to plasmid containing the insert or to purified insert, but not to vector alone. Hence, the 50-kDa protein band is specific for the cDNA insert and is only detected from mesenchyme blastula onward.

The translation product may be compared to spicule matrix proteins synthesized by the embryo (Benson *et al.*, 1986). Prominent glycoproteins of the spicule matrix migrate at apparent molecular weights of 47, 50, and 57 kDa and after deglycosylation have molecular weights of about 44, 47, and 54 kDa, respectively. The leader sequence of these deglycosylated matrix proteins, if any, would probably be missing. Thus, there is good correspondence between the 50-kDa protein produced *in vitro* that would contain the N-terminal leader and the 47-kDa deglycosylated protein derived from the embryo.

Hybrid Arrest of Translation

An alternative approach to establishing that pHS72 encodes a spicule matrix protein is to arrest translation of the particular poly(A)$^+$ RNA sequence to which the clone is complementary. The techniques of this method are more convenient to use with multiple RNA samples and are less likely to lead to degradation of mRNA. The 1.3-kb cDNA, pHS72, described by Sucov *et al.* (1987) was hybridized to poly(A)$^+$ RNA of the gastrulae after which the mixture was divided in half. One portion was heated to denature any hybrids that might have formed, whereas the other portion of hybrids was left untreated. RNA was then translated *in vitro*. The pattern of proteins synthesized *in vitro* under the direction of poly(A)$^+$ RNA was too complex to reveal the absence of a single polypeptide by a one dimensional gel; therefore, products of translation were immunoprecipitated prior to electrophoresis. Figure 2 shows the results of such an experiment. There is a particularly prominent doublet at approximately 50 kDa. Hybridization to 0.6 μg of pHS72 specifically inhibits translation of the lower band of this doublet which has a molecular weight of 50 kDa. The upper 51-kDa band is unaffected. Hybridization with 10 times less plasmid produced only marginal arrest of translation. Hybridization to vector DNA alone had no effect. It is clear there is a very selective inhibition by pHS72 of only one immunoprecipitable band from the array of proteins encoded by poly(A)$^+$ RNA. This band corresponds with an expected unglycosylated and unprocessed version of the 50-kDa spicule matrix glycoprotein. The identity of the upper band is uncertain. It may represent the translation product of another spicule matrix mRNA or a protein which nonspecifically binds to certain preparations of bacteria bearing protein A.

We also carried out similar experiments using a 24-mer synthetic polynucleotide with the nucleotide sequence 5'-CCA-TGC-CTG-GTT-GTC-GCC-CAC-CCA-3', which is complementary to the sequence that encodes Trp-Val-Gly-Asp-Asn-Gln-Ala-Trp, a consensus sequence that appears 13 times in the C-terminal half of the protein (Sucov *et al.*, 1987). Kawasaki (1985) has shown that hybridization of synthetic oligonucleotides corresponding to the amino, middle, or carboxyl portions of a protein are all effective in arresting translation. We found hybrid arrest experiments with this 24-mer reduced translation of the 50-kDa band (data not shown), but it was not as effective as the 1.3-kb cDNA. Another 24-mer polynucleotide which is upstream of the AUG initiation codon (nucleotides 110–112), the complement of the sequence from nucleotides 4–28 of the cognate gene, was ineffective in producing hybrid arrest of translation. Hence, hybrid arrest experiments with this

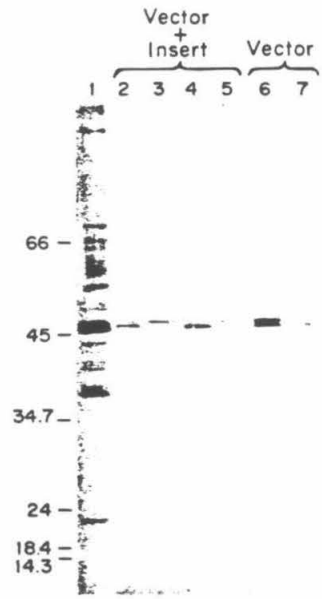


FIG. 2. The specific arrest of translation by hybridization to pHS72. Poly(A)$^+$ RNA from the midgastrula stage was hybridized and subsequently translated in a reticulocyte lysate containing [^{35}S]methionine followed by immunoprecipitation, electrophoresis, and fluorography to visualize the products. Lane 1, 6 μg RNA with no further treatment. Lane 2, 6 μg of RNA was hybridized to 0.6 μg of pHS72 and then boiled and fast cooled to denature any hybrids that may have formed. Lane 3, same conditions as lane 2 but without hybrid denaturation. Thus, hybridization of 6 μg of RNA with 0.6 μg of pHS72 specifically inhibits translation of the lower band of the doublet. Lane 4, 6 μg of RNA was hybridized to 0.06 μg of pHS72 and then boiled and fast cooled. Lane 5, same as lane 4 but without hybrid denaturation. Lane 6, 6 μg of RNA was hybridized to vector (PS 65) DNA and then boiled and fast cooled. Lane 7, same as lane 6 but without hybrid denaturation.

mRNA seem to require hybridization to the coding sequence.

RNA Gel Blot Hybridization

The timing of the expression of the gene encoding the pHS72 transcript was examined by RNA blotting analysis. Total RNA was extracted from the embryos of various stages, subjected to electrophoresis on agarose gels containing formaldehyde, blotted to nitrocellulose, and probed with radioactive cDNA. As revealed by hybridization with a single-stranded RNA probe, the hybrid was first apparent near the time of primary mesenchyme formation, thereafter reaching a plateau in amount by late gastrula. Figure 3A shows that a clear hybridization signal occurs in RNA from hatching blastulae and becomes very prominent by mesenchyme blastula and later stages, including the 72-hr pluteus.

The sensitivity of detection was increased by loading more RNA on the gel, preparing more highly radioactive single-stranded probes, and flashing the film (which also produces a linear response). Figure 3B shows that a very faint signal is detectable in morula-stage embryo RNA (containing between 100 and 150 cells) and in blastulae before hatching (approximately 300 cells). There are no primary mesenchyme cells at this stage; there is some gene expression prior to primary mesenchyme ingression, albeit at a low level. We have been unable to detect any hybridization signal at earlier stages with these techniques. The hatching blastula sample showing some transcript (Figs. 3A and 3B) was taken from a culture which contained embryos with a few primary mesenchyme cells within the blastocoel; hatching and mesenchyme ingression are not always perfectly sequential in this species. RNA from later stages that showed the presence of the transcript were fractionated into poly(A)$^+$ and poly(A)$^-$ RNA and only the poly(A)$^+$ RNA fraction contained the transcript (data not shown).

The data of Fig. 3B were examined by densitometry of images produced by various exposure times, and the results are shown in Fig. 4. The relative amount of accumulation of the transcript of the spicule matrix gene shows a dramatic increase as the primary mesenchyme forms, and this accumulation continues through later development. Development up to the gastrula stage results in at least a 100-fold increase in transcript concentration over that present at the morula stage.

In Situ Hybridization

The cellular distribution of the transcripts was examined by the use of *in situ* hybridization using ^{3}H probes produced by SP6 polymerase transcription of the cloned cDNA. Figure 5 shows representative examples of the results. Embryos that have not hatched and possess no ingressed mesenchyme yield no detectable signal above background (A, B). The very faint signal seen on RNA blots from embryos of this stage is very reproducible but is produced by such an extremely low concentration of transcript that it cannot be localized by this technique, indicating there are probably less than 100 transcripts/embryo at this stage (Cox *et al.*, 1984). The first primary mesenchyme cells that can be identified within the blastocoel display a prominent localization of grains (C, D). As the mesenchyme continues its development to form spicules during gastrulation (E, F)

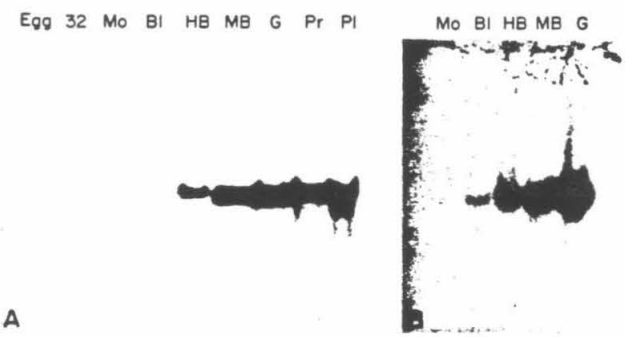


FIG. 3. RNA gel-blotting hybridization to detect transcripts of the SM50 spicule matrix gene. (A) 5 μg of total RNA from each stage was subjected to electrophoresis, blotted, and hybridized with 5×10^6 dpm/ml of ^{32}P-labeled single-stranded antisense RNA probe as described under Methods. The blot was treated with RNase and exposed to unflashed film and intensifying screen for a 1.5-day exposure. (B) 10 μg of the same RNAs was similarly analyzed with 10^7 dpm/ml of probe, washed with RNase, and exposed to flashed film and an intensifying screen for 1 week. The various RNAs are egg, 32 (32 cells), Mo (morula of about 120 cells), Bl (blastulae of about 300 cells), HB (hatching blastula, some of which contained a few primary mesenchyme cells), MB, (mesenchyme blastula), G (gastrula), Pr (prism stage), and Pl (3-day pluteus larva).

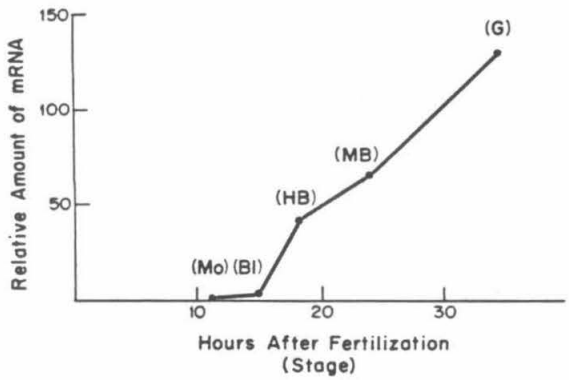


FIG. 4. The accumulation of SM50 transcripts. The signal produced by hybridization in Fig. 3B was scanned in a densitometer and the relative strength of the signal is shown. The value 1.0 was assigned to the signal of the morula RNA.

and in later stages (G, H), the localization of grains is very striking and localized exclusively over primary mesenchyme cells engaged in spicule formation. As a control for nonspecific binding, the sections were also probed with SP6 transcripts from the other strand of the cDNA, producing a "sense" RNA, and this probe gave no reaction whatsoever above background at any stage (I). We conclude that the gene encoding pHS72 transcripts is expressed at very low levels during late cleavage, that this level of transcript increases abruptly in mesenchyme cells as they ingress into the blastocoel, and that the mRNA continues to accumulate during spicule formation.

CONCLUSIONS AND DISCUSSION

The aims of the present set of experiments are to confirm that the cDNA isolated by Sucov *et al.* (1987) encodes a matrix protein (SM50) of the endoskeletal spicule of sea urchin embryos and to describe the time and place of its expression in the developing embryo. The results of several different approaches are consistent with the conclusion that pHS72 does encode such a spicule matrix protein. The antibodies used in the present study were shown to have a high degree of specificity, reacting with a number of matrix spicule proteins (Benson *et al.*, 1986) and staining only demineralized spicule matrix of the pluteus (Wilt *et al.*, 1987). This antibody was used to select phage from the expression vector library and the DNA sequence of the cDNA encodes an organic matrix-type protein; the acidic amino acid composition (p*I* 5.2), single glycosylation site, and putative signal peptide are consistent with the expectations based on prior work (Benson *et al.*, 1986; Benson *et al.*, 1983). Furthermore, the molecular weight of the unprocessed and unglycosylated protein, deduced from the DNA sequence, is close to that expected for the SM50 spicule matrix protein. Removal of a signal peptide and glycosylation at a single site would result in a glycoprotein with a molecular weight corresponding approximately to that of SM50 protein.

Use of cDNA in hybridization selection experiments yielded an RNA that when translated produced a peptide with a migration of about 50 kDa. This peptide is precipitated by antibody against spicule matrix protein. Only poly(A)$^+$ RNA extracted from stages undergoing spiculogenesis results in translation of such an immunoprecipitable peptide. Conversely, hybridization of the total poly(A)$^+$ RNA with the cDNA specifically suppressed the synthesis of one antibody-precipitable peptide, whose migration is identical to the peptide revealed by hybridization selection. Hence, the molecular weight, the stages from which the mRNA can be extracted, and the precipitation by antibody to matrix proteins are all consistent with the conclusion that pHS72 encodes SM50. Furthermore, the use of the cDNA as a cytological probe shows that detectable levels of the transcript are exclusively localized to primary mesenchyme. The number of transcripts per cell increases dramatically during gastrulation as frank spiculogenesis proceeds. This constitutes strong support that the cDNA and its cognate gene do encode spicule matrix protein SM50.

The RNA blotting experiments indicate that the transcript is accumulated at a low level in early development, perhaps only 1–10 copies per cell having accumulated by the 200–300 cell blastula stage; but as mesenchyme forms and spiculogenesis proceeds, a marked increase in accumulation occurs. Further experiments will be required to quantitate the exact time course of accumulation and the actual concentration of transcripts in mesenchyme cells. It is interesting, however, that the known sensitivity of hybridization, *in situ*, and the images we obtained here (Fig. 5) are completely consistent with the estimate in the accompanying paper (Sucov *et al.*, 1987) that primary mesenchyme cells elaborating spicule may contain 10^3–10^4 transcripts per cell.

While we know that the transcript is localized to mesenchyme cells, we do not yet know the cellular localization of the very low levels of transcript found in the prehatching blastula. This low level accumulation precedes the first appearance of the spicule. It is possible that the amount of transcript accumulated at these very early stages is too low to encode much protein, or even that translational control is being exerted at blastula stages. We consider it more likely that a major factor explaining the time between when transcript rapidly accumulates and when the spicule first appears is simply the requirement for sufficient time for accumulation of the polypeptide, its processing, glycosylation, and transport into the vacuole of the spicule cable, and the

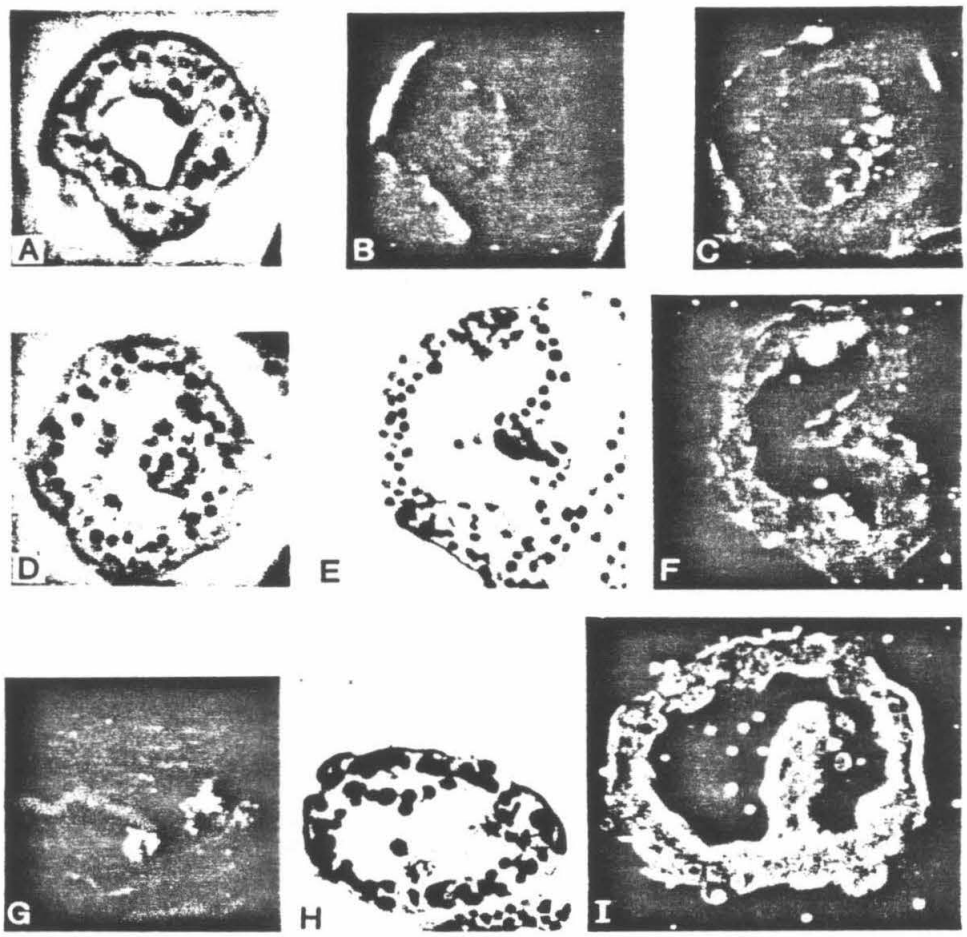


FIG. 5. The cellular localization of SM50 transcript by *in situ* hybridization. Embryos were fixed, processed, and hybridized overnight with ^{3}H-labeled single-stranded RNA probe, as outlined under Methods. Panel I was hybridized with a "sense" RNA probe, while Panels A–G were hybridized with an "antisense" RNA probe. A (bright field) and B (dark field) are the same section of a morula. C (dark field) and D (bright field) are different sections of a mesenchyme blastula. The accumulation of bright grains in C lies over ingressed primary mesenchyme cells, as do the dark grains in D. E (bright field) and F (dark field) are sections of a gastrula. G (dark field) and H (bright field) are transverse sections through a 2-day prism. I (dark field) is a gastrula.

beginning of calcification. Indeed, another lineage-specific protein, Meso 1, a mesenchyme-specific membrane protein, is synthesized hours in advance of its final deposition in the plasma membrane of primary mesenchyme cells (Wessel and McClay, 1985). Electron microscopy using antispicule antibody and colloidal gold indicates some matrix protein is present in the endoplasmic reticulum and Golgi of primary mesenchyme soon after ingression (unpublished observations). Lineage-specific gene expression may begin early in the micromere lineage, and perhaps there is even a major regulatory transition that occurs early in this lineage. Harkey and Whitely (1983) studied qualitative and quantitative changes in protein synthesis by micromere descendents using two-dimensional gel electrophoresis. They observed that the majority of changes occurred during the interval between the blastula and gastrula stages, and a large number of changes in protein synthesis preceded mesenchyme ingression.

The spicule matrix protein gene provides another lin-

eage-specific marker for the study of sea urchin development. Actin Cy II A also accumulates in mesenchyme cells, but in later development is found in the gut, whereas actins Cy III A and B are restricted to the aboral ectoderm. All these gene products show an accumulation phase beginning at the 20-hr mesenchyme blastula stage or later (Cox *et al.*, 1986). The Spec 1 genes encoding calcium-binding proteins are also restricted to aboral ectoderm and accumulate in postblastula stages (Bruskin *et al.*, 1982). The availability of a primary mesenchyme-specific gene is a useful addition to the existent collection of isolated genes, and it may be particularly useful for studies of determination and differentiation during sea urchin development. This gene is also potentially useful for studies on the mechanisms of biomineralization during development.

We appreciate the suggestions and materials generously given to us by Robert and Lynn Angerer, Doug Melton, and Ed Triplett. The technical assistance of Jody Tomic and Joanne Leung is gratefully acknowledged.

REFERENCES

ADELSON, D. (1986). Dissertation. University of Hawaii.

AVIV, H., and LEDER, P. (1972). Purification of biologically active globin messenger RNA by chromatography on oligothymidilic acid-cellulose. *Proc. Natl. Acad. Sci. USA* **69**, 1408–1413.

BENSON, S., JONES, E. M. E., CRISE-BENSON, N., and WILT, F. (1983). Morphology of the organic matrix of the spicule of sea urchin larva. *Exp. Cell. Res.* **148**, 249–253.

BENSON, S. C., BENSON, N., and WILT, F. (1986). The organic matrix of the skeletal spicule of sea urchin embryos. *J. Cell Biol.* **102**, 1878–1886.

BIRNBOIM, H. C., and DOLY, J. (1979). A rapid alkaline extraction procedure for screening recombinant plasmid DNA. *Nucleic Acids Res.* **7**, 1513–1523.

BRUSKIN, A. M., BEDARD, P. A., TYNER, A. L., SHOWMAN, R. M., and KLEIN, W. H. (1982). A family of proteins accumulating in ectoderm of sea urchin embryos specified by two related cDNA clones. *Dev. Biol.* **91**, 317–324.

CHILDS, G., MAXSON, R., and KEDES, L. H. (1979). Histone gene expression during sea urchin embryogenesis: Isolation and characterization of early and late mRNA of *S. purpuratus* by gene specific hybridization and template activity. *Dev. Biol.* **73**, 153–173.

COX, K. H., ANGERER, L. M., LEE, J. J., DAVIDSON, E. H., and ANGERER, R. C. (1986). Cell lineage-specific programs of expression of multiple actin genes during sea urchin embryogenesis. *J. Mol. Biol.* **188**, 159–172.

COX, K. H., DELEON, D. V., ANGERER, L. M., and ANGERER, R. C. (1984). Detection of mRNAs in sea urchin embryos by *in situ* hybridization using asymmetric RNA probes. *Dev. Biol.* **101**, 485–502.

DENHARDT, D. (1966). A membrane filter technique for the detection of complementary DNA. *Biochem. Biophys. Res. Commun.* **23**, 641–646.

DREYFUS, G., ADAM, S. A., and CHOI, Y. D. (1984). Physical change in cytoplasmic messenger ribonucleoprotein in cells treated with inhibitors of mRNA transcription. *Mol. Cell Biol.* **4**, 415–423.

GIBBINS, J. R., TILNEY, L. G., and PORTER, K. R. (1969). Microtubules in the formation and development of the primary mesenchyme in *Arbacia punctulata*. *J. Cell Biol.* **41**, 201–226.

HARKEY, M. A., and WHITELY, A. H. (1983). The program of protein synthesis during the development of the micromere-primary mesenchyme cell line in the sea urchin embryo. *Dev. Biol.* **100**, 12–28.

HINEGARDNER, R. (1967). Echinoderms. In "Methods in Developmental Biology" (F. Wilt and N. K. Wessels, Eds.), pp. 139–155. T. Y. Crowell, New York.

HORSTADIUS, S. (1973). "Experimental Embryology of Echinoderms." Clarendon, Oxford.

KAWASAKI, E. S. (1985). Quantitative hybridization arrest of mRNA in *Xenopus* oocytes using single stranded complementary DNA or oligonucleotide probes. *Nucleic Acids Res.* **13**, 4991–5004.

LAEMMLI, U. K. (1970). Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature London* **227**, 680–685.

LASKEY, R. A., and MILLS, A. D. (1975). Quantitative film detection of ^{3}H and ^{14}C in polyacrylamide gels by fluorography. *Eur. J. Biochem.* **56**, 341–355.

LYNN, D. A., ANGERER, L. M., BRUSKIN, A. M., KLEIN, W. H., and ANGERER, R. C. (1983). Localization of a family of mRNAs in a single cell type and its precursors in sea urchin embryos. *Proc. Natl. Acad. Sci. USA* **80**, 2656–2660.

MELTON, D., KRIEG, P., REBAGLIATI, M., MANIATIS, T., ZINN, K., and GREEN, M. (1984). Efficient *in vitro* synthesis of biologically active RNA and RNA hybridization probes from plasmids containing a bacteriophage SP6 promoter. *Nucleic Acids Res.* **12**, 7035–7056.

MANIATIS, T., FRITSCH, E. F., and SAMBROOK, J. (1982). "Molecular Cloning: A Laboratory Manual." Cold Spring Harbor Laboratories, Cold Spring Harbor, NY.

OKAZAKI, K. (1975a). Normal development to metamorphosis. In "The Sea Urchin Embryo." (G. Czihak, Ed.), pp. 177–233. Springer-Verlag, New York.

OKAZAKI, K. (1975b). Spicule formation by isolated micromeres of the sea urchin embryo. *Amer. Zool.* **15**, 567–581.

PATTERSON, B., ROBERTS, B. E., and KUFF, E. (1977). Structural gene identification and mapping by DNA-RNA hybrid-arrested cell-free translation. *Proc. Natl. Acad. Sci. USA* **74**, 4370–4374.

PELHAM, H. R. B., and JACKSON, R. J. (1976). An efficient mRNA dependent translation system from reticulocyte lysates. *Eur. J. Biochem.* **67**, 247–256.

RICCIARDI, R. P., MILLER, J. S., and ROBERTS, B. E. (1979). Purification and mapping of specific mRNAs by hybridization selection and cell free translation. *Proc. Natl. Acad. Sci. USA* **76**, 4927–4931.

ROHRSCHNEIDER, L., EISENMAN, R., and LEITCH, C. (1979). Identification of a Rous sarcoma virus transformation related protein in normal avian and mammalian cells. *Proc. Natl. Acad. Sci. USA* **76**, 4479–4483.

SUCOV, H., BENSON, S. C., ROBINSON, J. J., BRITTEN, R. J., WILT, F. H., and DAVIDSON, E. H. (1987). A lineage-specific gene encoding a major matrix protein of the sea urchin embryo spicule. II. A Structure of the gene and derived sequence of protein. *Dev. Biol.* **120**, 507–519.

THOMAS, P. S. Hybridization of denatured RNA and small DNA fragments transferred to nitrocellulose. (1980). *Proc. Natl. Acad. Sci. USA* **77**, 5201–5205.

WESSEL, G. M., and MCCLAY, D. R. (1985). Sequential expression of germ-layer specific molecules in the sea urchin embryo. *Dev. Biol.* **111**, 451–463.

WILT, F., BENSON, S. C., and BENSON, N. R. (1987). Spicule formation in sea urchin embryos. In "Molecular Approaches to Developmental Biology," UCLA Symposia on Molecular and Cellular Biology (R. A. Firtel and E. H. Davidson, Eds.), Vol. 51 Alan R. Liss, New York, in press.