Cis-regulatory control of the *SM50* gene, an early marker of skeletogenic lineage specification in the sea urchin embryo

Kazuhiro W. Makabe*, Carmen V. Kirchhamer, Roy J. Britten and Eric H. Davidson

Division of Biology 156-29, California Institute of Technology, Pasadena, CA 91125, USA

*Present address: Department of Zoology, Faculty of Science, Kyoto University, Kyoto 606, Japan

SUMMARY

The *SM50* gene encodes a minor matrix protein of the sea urchin embryo spicule. We carried out a detailed functional analysis of a *cis*-regulatory region of this gene, extending 440 bp upstream and 120 bp downstream of the transcription start site, that had been shown earlier to confer accurate skeletogenic expression of an injected expression vector. The distal portion of this fragment contains elements controlling amplitude of expression, while the region from −200 to +105 contains spatial control elements that position expression accurately in the skeletogenic lineages of the embryo. A systematic mutagenesis analysis of this region revealed four adjacent regulatory elements, *viz* two copies of a positively acting sequence (element D) that are positioned just upstream of the transcription start site; an indispensable spatial control element (element C) that is positioned downstream of the start site; and further downstream, a second positively acting sequence (element A). We then constructed a series of synthetic expression constructs. These contained oligonucleotides representing normal and mutated versions of elements D, C, and A, in various combinations. We also changed the promoter of the *SM50* gene from a TATA-less to a canonical TATA box form, without any effect on function. Perfect spatial regulation was also produced by a final series of constructs that consisted entirely of heterologous enhancers from the *CyIIIa* gene, the SV40 early promoter, and synthetic D, C, and A elements. We demonstrate that element C exercises the primary spatial control function of the region we analyzed. We term this a 'locator' element. This differs from conventional 'tissue-specific enhancers' in that while it is essential for expression, it has no transcriptional activity on its own, and it requires other, separable, positive regulatory elements for activity. In the normal configuration these ancillary positive functions are mediated by elements A and D. Only positively acting control elements were observed in the *SM50* regulatory domain throughout this analysis.

Key words: *Strongylocentrotus purpuratus, SM50* gene, mutagenesis analysis, gene regulation, locator element, skeletogenic lineage

INTRODUCTION

Cis-regulatory target sites encode the recognition elements by which a gene interprets its spatial position in development. Here we demonstrate the individual functions of specific target site sequence elements that are required to generate accurate transcriptional expression of a gene utilized exclusively in the skeletogenic lineages of the sea urchin embryo. From a developmental point of view, the gene we have chosen for this study, *SM50*, provides an excellent molecular marker of the autonomous specification process by which the skeletogenic lineages are specified, as we discuss below. We discovered that a synthetic oligonucleotide representing a single short *cis*-regulatory site is capable of imposing skeletogenic lineage-specific expression on an entirely heterologous expression vector.

In regularly developing sea urchins, the skeletogenic lineages derive from four polar founder cells, the large micromeres. These segregate from their non-skeletogenic sister cells, the small micromeres, at fifth cleavage. Three further divisions of

these lineages ensue during late cleavage, so that the embryo now contains 32 skeletogenic precursor cells. After gastrulation another division of these cells occurs so that the embryo ultimately contains 64 skeletogenic mesenchyme cells (Cameron et al., 1987; Ruffins and Ettensohn, 1993). At the early blastula stage, the skeletogenic precursor cells are located symmetrically in the vegetal wall of the embryo, surrounding the descendants of the small micromeres. They are surrounded by the vegetal plate precursors of the archenteron, with which they are in close contact. Midway into blastulation the skeletogenic precursors undergo a stereotypic differentiation process, as a result of which they acquire mobility, alter their cell surface constituents and properties, and reorganize their cytoskeletal structure. They then ingress singly into the blastocoel and assume a mesenchymal character (for review see Davidson, 1986; Ettensohn and Ingersoll, 1992; McClay et al., 1992). At the early gastrula stage the skeletogenic mesenchyme cells aggregate around the base of the invaginating archenteron. Skeletal elements, or spicules, then begin to be secreted, in two bilateral clusters of skeletogenic cells that form on the oral side

of the embryo. Many genes that are specific to the skeletogenic process have been cloned (cf. reviews cited above; also Drager et al., 1989). At present the best known of these genes are *msp130*, which encodes a cell surface glycoprotein (Leaf et al., 1987; Anstrom et al., 1987; Harkey et al., 1992; Kabakoff et al., 1992), and the *SM30* (George et al., 1991; Akasaka et al., 1994) and *SM50* genes (Benson et al., 1987; Sucov et al., 1987, 1988; Katoh-Fukui et al., 1991). Both *SM30* and *SM50* encode proteins of the matrix within which the mineral elements of the skeletal structures are embedded. *SM50* is of particular interest as a marker of the initial process by which the skeletogenic lineages are specified. Its transcripts begin to accumulate in the skeletogenic precursor cells within only two to three cleavages after the definitive 5th cleavage segregation of these lineages (Benson et al., 1987; Killian and Wilt, 1989). This is almost a whole day before the process of skeletogenesis *per se* begins to take place in the species with which this work was carried out, *Strongylocentrotus purpuratus*, and is also far in advance of any of the blastula-stage differentiation events that lead to ingression and expression of mesenchymal functions. We regard *SM50* gene expression as an early marker of skeletogenic lineage specification.

Specification of the skeletogenic founder cells in regularly developing sea urchins is, by all available tests, an autonomous process (Davidson, 1989, 1990; McClay et al., 1992). No immediate intercellular interactions are required for this specification to occur. Thus, once the skeletogenic micromeres have formed, they can be transplanted to ectopic locations in the embryo, whence their progeny ingress and carry out skeletogenesis, irrespective of their new location (Hörstadius, 1939; reviewed by Davidson, 1989). These cells are even unable to produce progeny that express any other differentiated fates, in contrast to all other blastomeres of the cleavage-stage sea urchin embryo (excepting their small micromere sister cells). Furthermore, as first shown by Okazaki (1975), if placed in culture, isolated 4th cleavage micromeres will carry out skeletogenesis in vitro, after undergoing the stereotypic number of divisions (reviewed by Davidson, 1986, p. 221); the only requirement is the addition of a small amount of horse serum to the sea water. The micromeres that harbor the autonomous potential to give rise to the skeletogenic precursor lineages are formed by the horizontal 4th cleavage plane, which is positioned so that they together include only the polar 7.5% of the cytoplasmic volume of the egg (Ernst et al., 1980). The implication of this autonomous specification process (Davidson, 1989, 1990) is that the activation of the earliest cohort of genes expressed exclusively in the skeletogenic lineage might depend on maternal factors localized at the vegetal pole of the egg.

MATERIALS AND METHODS

Animals and embryos

Adult *Strongylocentrotus purpuratus* collected along the Southern California coast were maintained in chilled sea water tables so that gametes were available year round at the Caltech Kerckhoff Marine Laboratory (Leahy, 1986). Preparations of gametes, fertilization and embryo culture were carried out according to standard methods. The embryos were microinjected as described by McMahon et al. (1985) with gel-purified, linearized plasmid DNA. Approximately 2500 molecules of the desired plasmid DNA were introduced per embryo,

together with a four-fold molar excess of *Pst*I-digested carrier sea urchin DNA (Franks et al., 1990).

All of the expression constructs used in this work are illustrated diagrammatically in Figs 1 and 4, to which the reader is referred for nomenclature used in the following. Constructs Δ(−200), Δ(−155) and Δ(−10) were gel-purified after digestion of Δ(−440) DNA with *Hpa*II, *Ava*II and *Dde*I, respectively. Δ(−440) is the same as *pB2S•CAT* of Sucov et al. (1988), which contains a *Bgl*II-*Sal*I fragment of the original *SM50* gene. Fusions Δ(d), Δ(b), Δ(a) and Δ(ab) were reconstructed from fragments of *pB2S•CAT* double digested with *Ava*II and *Dde*I; *Nsi*I and *Sfu*I; *Sfu*I and *Sal*I; and *Nsi*I and *Sal*I, respectively (see Fig. 3 for sequence; relevant restriction sites in this region of the *SM50* gene are indicated in Fig. 1).

Mutations were introduced in the form of synthetic oligonucleotides. Essentially, the region to be mutated was in each case replaced by synthetic double-stranded oligonucleotides in which the sequence had been transversionally changed (i.e., A↔C; T↔G). Some oligonucleotides contained flanking restriction enzyme target sites also present in the *SM50* gene so that they could be ligated into the expression vector. In other cases five-nucleotide tags were built onto the termini of the oligonucleotides, and complementary tags onto the termini of other oligonucleotides, for use in constructs bearing multiple mutated sites (cf. Fig. 1). The exact sequences of the *SM50* regulatory domain that were transversionally mutated in this fashion are indicated in Fig. 3.

Other fusion genes utilized (Fig. 1D,E; and all constructs shown in Fig. 4) were assembled from vector elements and positive regulatory elements derived from different genes, together with oligonucleotides representing natural elements of the *SM50* sequence. The sequences reproduced in these oligonucleotides are indicated as open boxes in Fig. 3. These oligonucleotides also contained terminal tags to allow their insertion into the vector or ligation to contiguous oligonucleotides. The sequence of the plus-strand oligonucleotide used for region D was cgcgCCAGGGTTACGacaccct; for region C,

Fig. 1. Schematic diagrams of injected constructs and efficiency of expression as monitored by whole-mount in situ hybridization. Regions of the regulatory domain referred to in the text are indicated at the top of the figure. Background levels of ectopic expression were about 2% (0-4% depending on batch of eggs used). Every construct shown that was expressed significantly over background was expressed only in cells of the skeletogenic mesenchyme lineages, i.e., above the average 2% background level. The fraction of embryos developing from eggs injected with each construct that expressed CAT DNA is shown immediately to the right of each colored diagram. Following this is the number of embryos (n) that constituted the total sample of embryos scored for that construct. The number of different experiments constituting that sample is indicated by the batch designations in the right-hand column. Each batch represents eggs from a single female on a given day. Each individual experiment, carried out on a single batch of eggs, is indicated by a capital letter. (A) Δ(−440), a construct containing the entire regulatory region between −440 and +120. This is composed of the 'distal' (yellow), 'proximal' (green), 'initiator' (purple), and 'plus' (blue) regions. (B) Constructs used to analyze the proximal region. Regions that were transversionally mutated are shown in pink. (C) Constructs used to analyze the plus region. (D) *m(Inr)TATA*, a construct in which the original initiator was replaced by a synthetic TATA box (T) and a synthetic initiator (red) (see text). The sequence of the changed region of the gene is shown in Fig. 3(B). (E) *SV40(DCA)*, a construct containing the SV40 early region promoter and synthetic DNA sequences representing the D, C, and A sequence elements of the original *SM50* gene (see text). '% expression' is the percentage of embryos expressing the CAT reporter, as observed by whole-mount in situ hybridization in ≥two skeletogenic mesenchyme cells (see text).

cgaggTGGTAGTCGTGAATGCATCGATCTCcgggtctag; and for region A, atcccAAAGTCTAGTGAGATCGCAACACATTTGA-GAAGCAg. The minus strand of the region A oligonucleotide, which is not shown here, contained a *Sal*I site at its 5′ end. The elements to

which these synthetic *SM50* sequences were attached were obtained as follows. The SV40 early promoter was obtained from Promega, and was used as a *Bgl*II-*Hin*dIII fragment which has no enhancer (Gorman et al., 1982). Construct RTB-1 (Fig. 4) consisted of the CAT
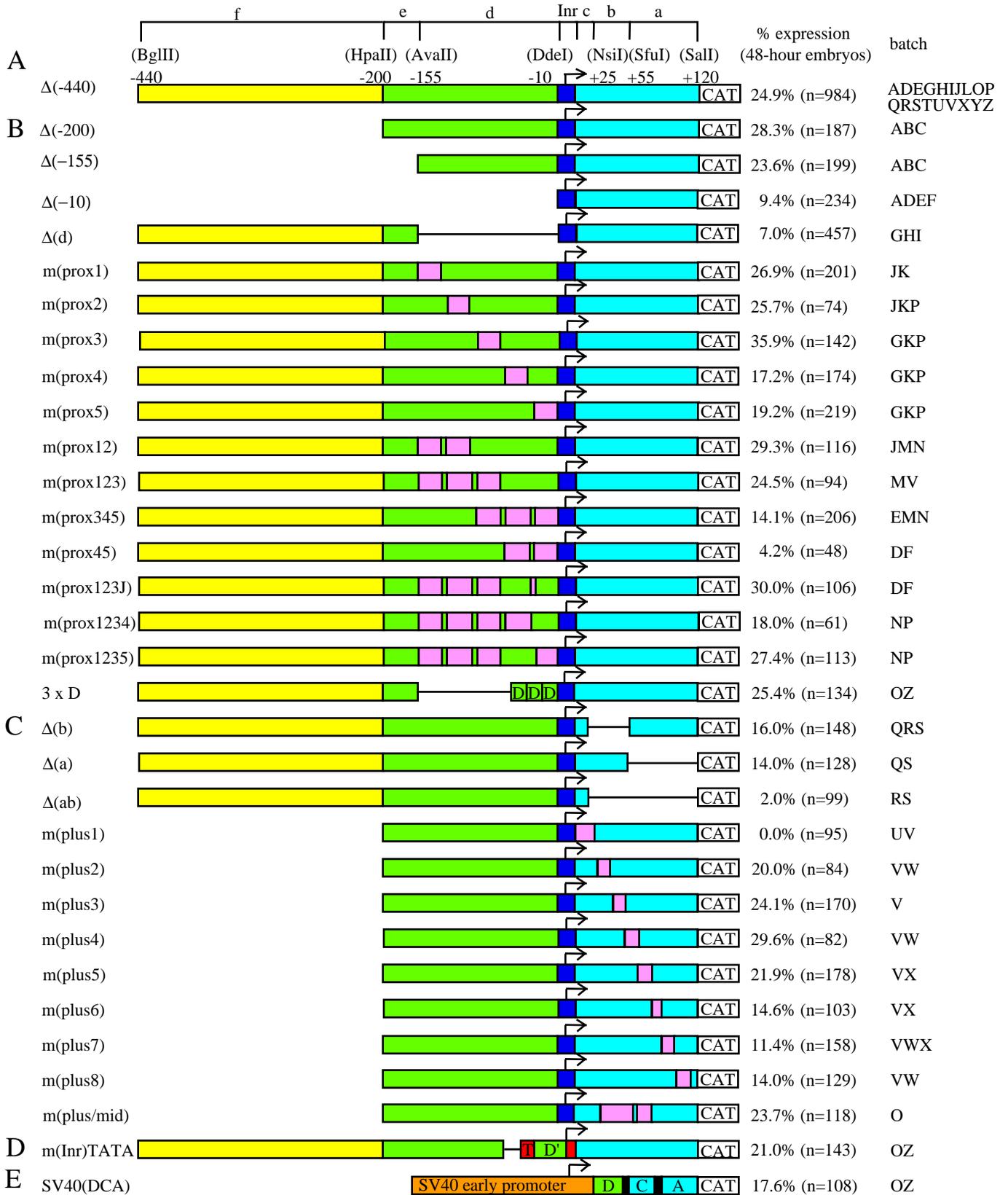
expression vector as above, the SV40 early promoter, and two fragments from the *CyIIIa* gene containing the distal SpGCF1 (P8) cluster of enhancer sites (Thézé et al., 1990; Zeller et al., 1995), and the sites at which the positively acting factors SpCTF1 (P4), SpTCF1 (P5), and SpOct1 (P3B) bind (reviewed by Coffman and Davidson, 1992). These three sites were obtained on a single restriction fragment from a deletion mutant of *CyIIIa•CAT* which lacks the negatively acting P3A2 target site (unpublished experiments). *RTB-1* also includes a synthetic polylinker, PL1/2, containing *Apa*I, *Nhe*I, *Asc*I, *Acc*65I, and *Bam*HI sites, positioned at the 5′ side of the CAT coding sequence. A second synthetic polylinker, PL3/4, that includes *Hin*dIII, *Xho*I, *Apa*I, *Pst*I, *Sph*I, *Bgl*II, *Bsi*WI, *Mlu*I, *Xba*I, *Eag*I, *Eco*RV, and *Sal*I sites, was positioned at the 5′ end of the SpGCF1 site cluster.

Synthetic TATA and initiator box elements were also used in some constructs. Based on consensus compilations (Kaufmann and Smale, 1994) the oligonucleotides bearing these included the sequences CATATTA, and ACACC, respectively. These sequences were built into a D′ or D oligonucleotide (cf. Fig. 3) and ligated together with other oligonucleotides and/or the vector, as required, by the same approach as above. In the synthetic TATA box constructs the junction point between the natural *SM50* sequence and the TATA-D oligonucleotide was after the T at −55 of the natural sequence (Fig. 3).

Whole-mount in situ hybridization and CAT assay

The whole-mount in situ hybridization protocol used here is based on the method described by Ransick et al. (1993). However, the procedure was slightly modified by the use of Streck Fixative (Streck Laboratories Inc., NE). Fixation was carried out overnight at 4°C, followed by a brief wash in 50 mM Tris buffered-saline, containing 0.1% Tween 20. The proteinase K treatment and post-fixation procedure of Ransick et al. (1993) were omitted. CAT assays were carried out according to Seed and Sheen (1988).

RESULTS

Temporal and spatial expression of Δ*(−440)*, an *SM50•CAT* fusion gene

Spatially accurate embryonic expression of exogenous *SM50•CAT* fusion constructs was demonstrated earlier by Sucov et al. (1988). Their starting construct, which was expressed exclusively in the skeletogenic mesenchyme cells of mesenchyme blastula-stage embryos, included *SM50* sequences extending from −2200 downstream to +120, with respect to the transcription start site. Expression was assayed by radioactive in situ hybridizations carried out on serially sectioned samples. No change in level or locus of expression was observed when the upstream sequence was deleted down to −440, but further deletions sharply decreased expression (Sucov et al., 1988). This evidence provided the starting point for our present experiments. As we describe below, the minimal *SM50* regulatory domain defined by Sucov et al. (1988) contains a number of functionally significant sequence elements, and their further analysis required the use of fine-scale deletions, mutations, and a variety of synthetic constructs. Fig. 1 displays a series of mutationally altered *SM50•CAT* constructs that we injected into zygotes and analyzed quantitatively for expression during embryogenesis. We had the advantage in this work of more sensitive methods of analysis than were available to Sucov et al. (1988), *viz* the whole-mount in situ hybridization procedure of Ransick et al. (1993; slightly modified, as described in Materials and Methods, by the use of a different fixative).

The construct shown at the top of Fig. 1, Δ*(−440)*, serves as the control for the remaining constructs. In Fig. 1 the regulatory sequence included in Δ*(−440)* is divided into four color-coded subregions, to which we refer in the following. These are the 'distal' region (−440 to −200), shown in yellow; the 'proximal' region (−200 to −10), shown in green; the 'initiator' region (−10 to +10) shown in purple; and the 'plus' region (+10 to +120) shown in blue. There may well exist additional *cis*-regulatory sequence upstream of −440 in the native *SM50* gene, particularly since we have not investigated requirements for larval or post-metamorphosis expression of this gene, which is utilized in adult test and spine-producing cells as well as in embryonic skeletogenic cells (Richardson et al., 1989). With respect to embryonic expression, however, we have confirmed and extended the observations of Sucov et al. (1988) that the *SM50* sequence between −440 and +120 suffices to generate a normal pattern of expression. Thus, injected Δ*(−440)* is expressed almost exclusively in skeletogenic mesenchyme cells. Illustrative embryos in which skeletogenic mesenchyme expression of Δ*(−440)* is visualized by whole-mount in situ hybridization are shown in Fig. 2A1-A3.

When DNA is injected into sea urchin zygotes it rapidly concatenates in the cytoplasm, and the concatenate becomes stably incorporated in a blastomere nucleus, in most cases at 2nd, 3rd, or 4th cleavage, though later incorporations also occur (McMahon et al., 1985; Flytzanis et al., 1985; Hough-Evans et al., 1988; Livant et al., 1991). Thereafter the exogenous DNA replicates together with the host cell DNA (Franks et al., 1988; Livant et al., 1991), and is thus present in all of the cells of the clone to which the blastomere initially incorporating it gives rise. The transcriptional capacity of injected constructs is amplified by this process, because the exogenous DNA in each embryonic cell consists of a replica of the original concatenate. Thus as we found earlier (McMahon et al., 1985; Flytzanis et al., 1985; Livant et al., 1988), each cell bearing the exogenous DNA typically contains a sufficient number of copies of the introduced regulatory sequences to sequester a significant fraction of the transcription factors that service it. This means that the observed expression is contributed by many copies of the injected gene.

The mosaic incorporation pattern that results can be observed in Fig. 2A, for example. Eight of the 32 skeletogenic mesenchyme cells in the embryo at this stage are stained (these cannot all be distinguished clearly in the plane of focus shown), indicating that the incorporation occurred either in one of the first four blastomeres, in a vegetal 3rd-cleavage blastomere, or in a 4th-cleavage micromere, but not later. Embryos were considered positively stained only if two or more cells displayed the blue reaction product, and if the embryos were morphologically normal. In our experience the single stained cells that are occasionally observed usually indicate unstable or transient exogenous DNA that will disappear later in development, and that is often ectopically expressed irrespective of the construct injected. Single stained cells can also be produced as a result of experimental damage or abnormal developmental process that causes cells to delaminate from the ectoderm and fall into the blastocoel. The top line of data in Fig. 1 shows that by these criteria Δ*(−440)* was expressed in an average of close to 25% of embryos, based on examination of almost a thousand embryos belonging to 19 different batches of eggs (each batch is derived from a single female). Some batches were more

active than others, and in these as much as 30-35% of embryos displayed labeled mesenchyme cells. The majority of batches expressed in about 24-28% of embryos. However, these values are all about two-fold lower than what would be expected on a purely random basis, judging from results obtained with constructs that are expressed in ectodermal or vegetal plate territories, in which 60-80% of embryos commonly express the injected constructs (C.-H. Yuh, C. Kirchhamer, E. Davidson, unpublished data). It follows that there is probably a bias against incorporation into skeletogenic precursors, i.e., the 4th or 5th cleavage micromeres. Such a bias might arise simply from the reduced chance these cells have of inheriting the exogenous DNA concatenate because of the disproportionately small fraction of zygote cytoplasm they inherit, or perhaps because of their delayed cleavage. An indication of a bias toward later incorporation than seen with other territories is that the *average* number of stained cells per positive embryo was, for Δ*(−440)*, only three to four (again there was batch variation in this statistic). We note that Sucov et al. (1988) reported a similar result, about five cells labeled per positive embryo, based on a very much smaller sample.

Spatial expression of Δ*(−440)* is quite accurate, in that <2% of injected embryos ever displayed any ectopic expression, i.e., expression other than in skeletogenic mesenchyme. Δ*(−440)* also contains *cis*-regulatory sites that mediate the temporal activation of the reporter as development proceeds. Transcripts of the endogenous *SM50* gene accumulate sharply during development, from a few per skeletogenic precursor cell in late cleavage to 100-200 per cell from the mesenchyme blastula stage onward (Killian and Wilt, 1989). Measurements of the relative transcription rate by the run-off method also show that the *SM50* gene is transcriptionally activated by mesenchyme blastula stage (Killian and Wilt, 1989), but no comparisons between mesenchyme blastula and later stages are available. Table 1 shows a several-fold increase in CAT enzyme content in embryos expressing Δ*(−440)* between mesenchyme blastula and 30 hour (early gastrula) or 48 hour (late gastrula) stages. The same fraction of embryos was observed to be expressing Δ*(−440)* at 21 hours, as at the 48 hour late gastrula stage for which data are recorded in Fig. 1. Therefore the number of expressing embryos at 48 hours is not limited by the sensitivity of the in situ hybridization (see also below). The developmental increase in CAT levels between mesenchyme blastula and gastrula stages requires a proportionately increased rate of transcription from Δ*(−440)* over this period, since both CAT mRNA and CAT enzyme protein are unstable in sea urchin embryos. Their half-lives were estimated earlier at ≤40 minutes (Flytzanis et al., 1987).

The distal region of Δ*(−440)* affects the level of expression

Preliminary experiments of Sucov (1989) showed that deletions extending proximally of −440 depress the level of expression of *SM50•CAT* constructs. Table 1 demonstrates that deletion of the distal region (construct Δ*(−200)* in Fig. 1) in fact results in a two- to three-fold decrease in CAT expression, as monitored at early and late gastrula stages. This, however, does not bring the level of CAT expression down near the limit of detection of whole-mount in situ hybridization, since, as shown at the right side of Fig. 1, about the same fraction of embryos bearing the Δ*(−200)* construct score as positive as of

Table 1. CAT enzyme activity in transgenic embryos bearing *SM50•CAT* mutants

| | CAT activity normalized to value obtained for Δ*(−440)* at 48 hr | | |
Constructs	21 hours pf	30 hours pf	48 hours pf
Δ*(−440)*	18.2	86.4	100.0
Δ*(−200)*	—	25.9	45.8
SV40(DCA)	2.7	17.3	15.8

Each value represents the average of two or more samples of 100 injected eggs, except for the 21 hour *SV40(DCA)* sample, which consisted of 700 injected eggs. pf, post-fertilization.

embryos bearing Δ*(−440)*. Expression of Δ*(−200)* remains entirely confined to skeletogenic mesenchyme cells, as illustrated in the embryo shown in Fig. 2B1,2. Thus the only function that we observed for the distal region is quantitative. We did not explore further the distal region in this work, and thus it remains undetermined whether its positive function is specific to skeletogenic mesenchyme lineages. However, a clear conclusion from the Δ*(−200)* experiments can be drawn, *viz* that skeletogenic mesenchyme-specific regulatory elements are present in the *SM50* sequence that lies between −200 and +120.

Two copies of an important regulatory sequence, the 'D-repeat', are present in the proximal region

A further deletion to −155 gave essentially the same result as did Δ*(−200)*, but when the sequences inward to −10 were also deleted, expression dropped precipitously (Fig. 1; Δ*(−10)*). Less than 10% of embryos now expressed CAT DNA at a level sufficient for detection by whole-mount in situ hybridization. Nonetheless the remaining expression was still confined to the skeletogenic mesenchyme. Therefore *cis*-regulatory elements specifying the territorial spatial function of the *SM50* gene must exist still further downstream, between −10 and +120.

The positive regulatory function indicated by this experiment for the sequence between −155 and −10 (for convenience this region is labeled 'd' at the top of Fig. 1) cannot be substituted by the positive function resident in the distal region. Thus expression of construct Δ*(d)*, which consists of Δ*(−440)* minus region d, is depressed as much as expression of Δ*(−10)* (see Fig. 1). To identify the responsible sequences we generated the twelve mutations labeled *m(prox1)* to *m(prox1235)* in Fig. 1. These were assembled from a series of overlapping synthetic oligonucleotides, as described in Materials and Methods. The exact sequences that were mutated by this means to produce the constructs shown in Fig. 1 are indicated in Fig. 3. As summarized in Fig. 1B, none of the single block mutations (i.e., *m(prox1)* to *m(prox5)*) affected expression as severely as observed with Δ*(−10)* or Δ*(d)*, and mutants *m(prox1)* to *m(prox3)* were expressed as well as the control Δ*(−440)*. Again in these experiments expression was always confined to cells of the skeletogenic mesenchyme lineages. Double, triple and quadruple block mutants were then tested. We found (Fig. 1B) that only in mutants in which the sequence of regions 4 and/or 5 were altered was the percentage of positive embryos significantly depressed (i.e., *m(prox4)*, *m(prox345)*, *m(prox45)*, *m(prox1234)*, and *m(prox5)*, *m(prox345)*, though not *m(prox1235)*). This result indicated that a strong positively acting element is present in both subregions 4 and 5. When both of these subregions are mutated (e.g. in *m(prox4,5)*,
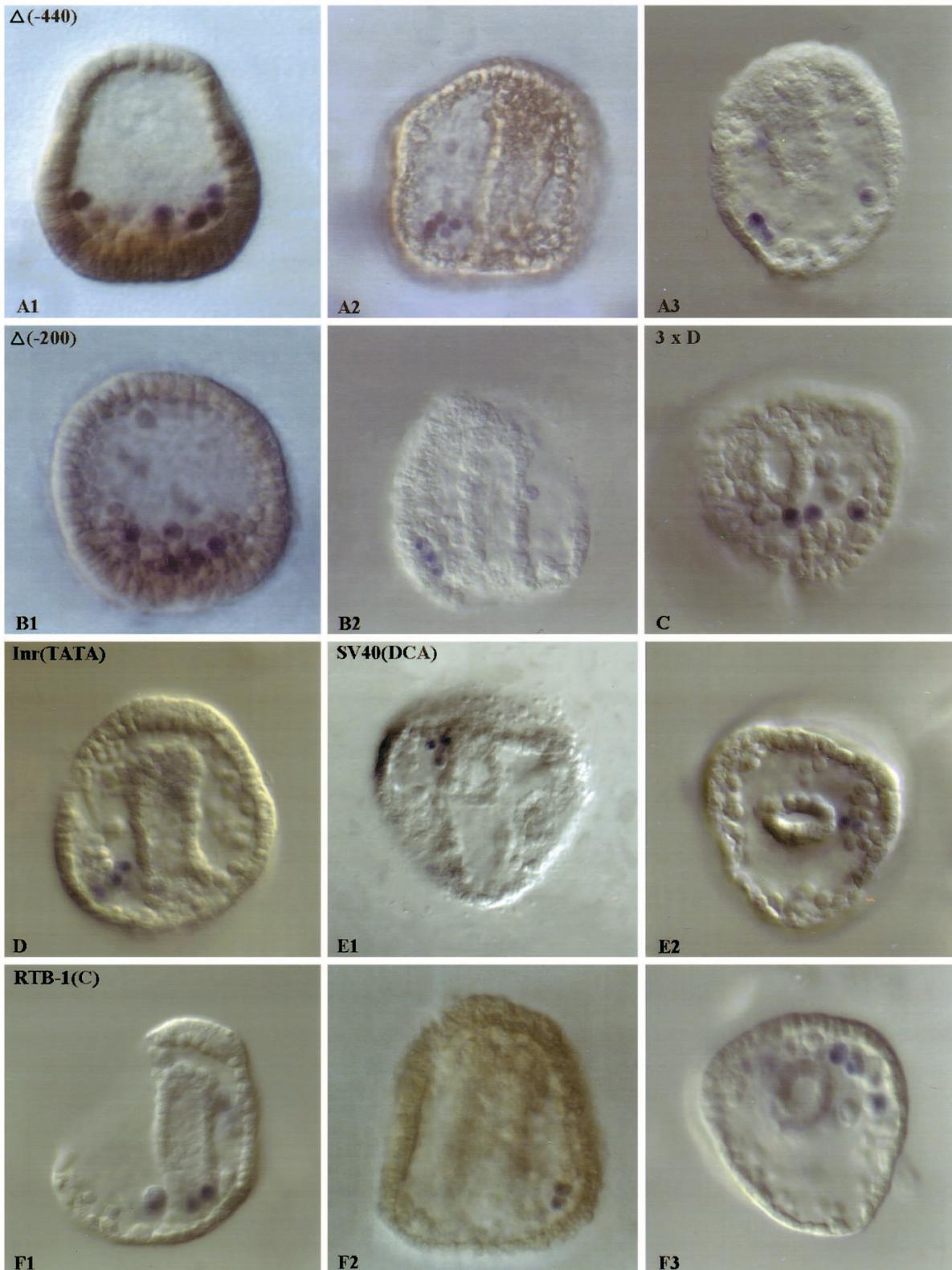
Fig. 2. Illustrative whole-mount in situ hybridizations of embryos bearing various expression constructs, indicated in each panel. Injected zygotes were cultured for 24 hours (mesenchyme blastula stage embryos) or 48 hours (late gastrula stage embryos). The blue staining is seen only in skeletogenic mesenchyme cells (the dark peripheral region observed in Fig. 2(E1) is an optical shadow rather than bona fide staining). Fig. A1, A2, B1, and F2 are photographs that were scanned to convert them to digital images; the remainder of the panels display digital images that were generated on a Roche Imaging Camera (ProgRes 3012) mounted on a Zeiss Axioskop. The figure was printed on a Nikon CP-3000 printer.

expression is most severely affected, approaching the low level observed when the whole of region d is deleted. Another mutation, in which the sequence lying between subregions 4 and 5 was changed (*m(prox123J)*; J, junction sequence) was expressed at the control level. It follows that subregions 4 and 5 may each include a target site for a positively acting factor, separated by the non-essential intervening J sequence (see Fig. 3). Subregions 4 and 5 apparently both contribute to the overall positive function of the proximal region. In context, subregion 4 might be the more essential, since *m(prox1234)* displays a depressed level of expression, while *m(prox1235)* expresses normally.

Fig. 3 shows that subregions 4 and 5 in fact both contain, on either side of the J sequence subelement, one occurrence of the short, directly repeated sequence AGGGTT. These sequences are labeled 'D$_{(4)}$ repeat' and 'D$_{(5)}$ repeat' in Fig. 3A. To test the proposition that the D repeat is responsible for the regulatory function of the proximal region revealed in these experiments, we constructed a synthetic oligonucleotide consisting of three copies of the sequence labeled D$_{(4)}$ in Fig. 3A (this includes two flanking nucleotides on the 5′ end of the D$_{(4)}$ repeat and three additional nucleotides on the 3′ end; these differ between D$_{(4)}$ and D$_{(5)}$). The synthetic sequence was then inserted into the Δ(d) construct (see Fig. 1, Construct *3xD*), and injected into eggs. This construct behaved exactly as did the

control Δ(−440), producing skeletogenic mesenchyme-specific expression in 27.4% of experimental embryos. An example is shown in Fig. 2C. The synthetic D$_{(4)}$ sequences thus account completely for the function of the proximal region of Δ(−440).

Regulatory elements in the plus region of the *SM50* regulatory domain

A similar analysis of the 'plus' region was carried out next. Results from the initial three constructs, shown diagrammatically in Fig. 1C as Δ(b), Δ(c) and Δ(ab), suggested that there are at least two subregions downstream of the transcription start site that are individually required for full expression, and that in their combined absence the construct is dead. We then built and tested the series of block mutations shown in Fig. 1C (see Fig. 3A for exact sequences), *viz* constructs *m(plus1)* to *m(plus8)*. Mutations of a sequence extending from +10 to +30 entirely obliterated expression (construct *m(plus1)*, while mutation of the adjacent seven nucleotides (construct *m(plus2)* depressed expression. In the following we designate the combined sequence that was mutated in these two constructs (i.e., +13 to +37) as region C. At least a major fraction of this sequence is absolutely required for *SM50* fusion gene expression.

Proceeding downstream, the regions mutated in constructs *m(plus3)* and *m(plus4)*, i.e., from +40 to +60 (Fig. 3), are apparently not essential. Thus as summarized in Fig. 1C, these constructs were expressed at normal levels, i.e., in about 24-30% of injected embryos. However, mutations in the region +60 to +103 all displayed reduced expression. This is shown in Fig. 1C for constructs *m(plus5)* to *m(plus8)*. The sequence from +60 to +103 is designated region A in the following; like
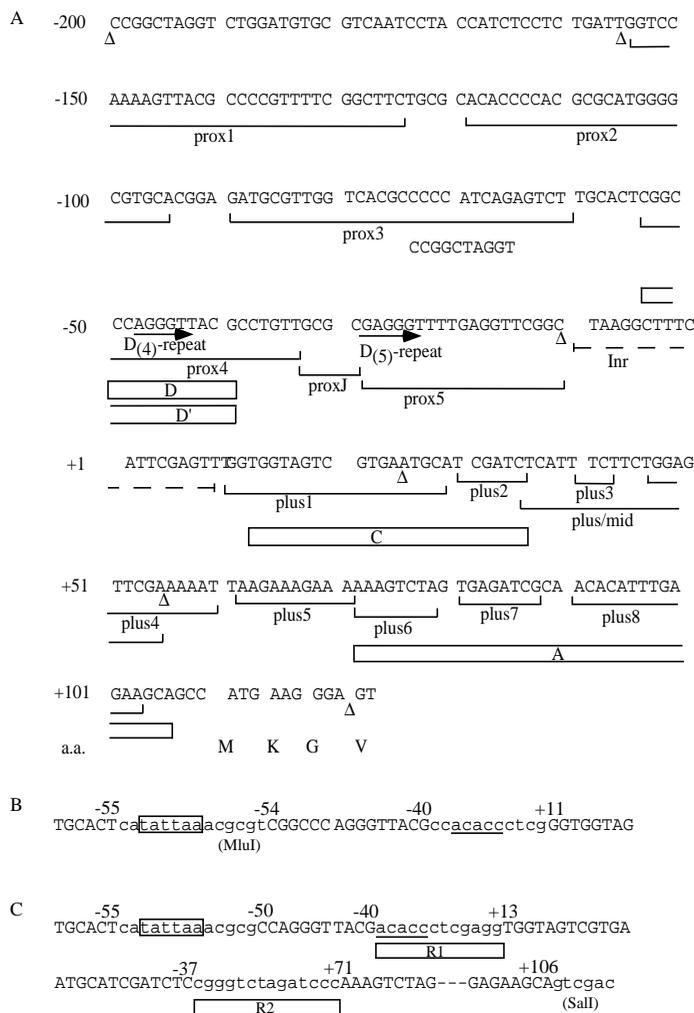
Fig. 3. Sequences of regions of the *SM50* regulatory domains that were deleted or mutated, or that were represented by synthetic double-stranded oligonucleotides in expression vectors described in Figs 1 and 4. (A) Sequence of *SM50* regulatory domain from −200 to +120. The final twelve nucleotides of this sequence encode the first four amino acids of the *SM50* protein, including the initial methionine. Exact endpoints of the deletion mutants shown in Fig. 1 are shown by triangles. Transversionally mutated regions are bracketed, with the names of the mutated sequences below each bracket. The D repeats in the proximal region (prox) are shown by a pair of horizontal arrows. The initiator region (Inr) that was replaced in the *m(Inr)TATA* construct is indicated by a dashed underline. An oligonucleotide, D′, that was used in the *m(Inr)TATA* construct, and oligonucleotides that constituted the synthetic D, C, and A elements used in the constructs of Fig. 4 are shown by open horizontal rectangles beneath the respective sequences. Note that the specific D, C, and A sequence elements shown here are not identical to any of the colored regions denoted by a-d at the top of Fig. 1. (B) Sequence of a changed portion of the *SM50* gene in the *m(Inr)TATA* construct. Capital letters represent the original sequence from the *SM50* gene, and lower case letters the introduced sequence where it differs from the native sequence (see A). Numbers represent positions on the original sequence as in A. The TATA sequence is boxed, and the new initiator sequence is underlined. (C) Sequence of the initiator and plus regions of the *SM50-TATA(DCA)* construct. Capital letters represent the original sequences from the *SM50* gene, that were transversionally changed in the mutants shown in Fig. 4A. The introduced TATA box and initiator sequences are indicated as above. The *Mlu*I and *Sal*I sites that were used for the construction of *SM50-TATA(DCA)* from *m(Inr)TATA* are indicated. Spacer regions containing restriction sites R1 and R2 are shown by rectangles.

region D this sequence acts positively, but neither deletion nor mutation of region A affects the nearly perfect localization of the residual expression in skeletogenic mesenchyme cells. Normal expression of the construct shown at the bottom of Fig. 1C, *m(plus/mid)*, confirms that regions C and A alone can substitute for the whole plus region, and also that the sequence between the C and A elements is dispensable.

By process of elimination it would appear from these experiments that the sequence element conferring skeletal mesenchyme-specific spatial expression must be either region C or the initiator region. All other deletions and mutations so far reviewed, i.e., the distal region deletions, the mutations that identified the D region of the proximal domain, and the A region mutations of the plus domain, preserve skeletogenic mesenchyme specificity. This test could of course not be applied to the mutation defining the C region, because without the C sequence no expression that can be observed by whole-mount in situ hybridization is obtained at all.

The initiator sequence is not required for skeletogenic mesenchyme specificity

The *SM50* promoter lacks a TATA box (Sucov et al., 1988), and instead belongs to the class of genes that utilize only an initiator sequence. This is the element that in a TATA-less promoter determines the transcription start site by serving as the assembly point for the TFIID complex (Kaufmann and Smale, 1994). To determine if the region in *SM50* that extends from -10 to $+10$ with respect to the transcription start site might also be required for the spatial regulation of *SM50•CAT* constructs, we rebuilt this region of the gene, replacing it with a synthetic oligonucleotide encoding a consensus initiator site (Kaufmann and Smale, 1994). As shown in Fig. 3B, the natural *SM50* sequence and the sequence including the initiator element (ACACC) differ completely for the 10 bp following the *SM50* transcriptional initiation site ($+1$). A consensus TATA box sequence was also inserted immediately upstream of a synthetic D repeat element at position -25 (Fig. 3B). The *SM50* sequence was thus changed into that of a typical TATA box-containing gene. The expression vector carrying this alteration, *m(inr)TATA* (Fig. 1D), was injected into eggs and its expression assessed by whole-mount in situ hybridization. Expression was again observed only in skeletogenic mesenchyme, as illustrated in Fig. 2D. Furthermore, the efficiency with which *m(Inr)TATA* was expressed (21.0%) is only a few percent below that of the control $\Delta(-440)$. The original initiator sequence is therefore not the locus of spatial specificity in constructs such as $\Delta(-10)$ or *m(plus/mid)* (cf. Fig. 1).

To confirm that the D, C, and A elements function as well with the TATA-containing construct as with the natural initiation sequence, we carried out the experiments summarized in Fig. 4A. Here the entire sequence of $\Delta(-440)$ from -55 to $+120$, was replaced with synthetic DNA. As Fig. 3 indicates, -55 is the position at which the D repeats of the proximal region begin in the natural *SM50* sequence. A 7 bp CATATTA sequence was inserted at this point, followed by the synthetic D sequence, and 25 bp from the TATA sequence, the initiator sequence (ACACC) was inserted, together with a restriction site linker to join it to the following C sequence, as above. The sequence of the junctional region of this vector, *SM50•TATA*, is shown in Fig. 3C. In it were mounted either normal, or transversionally mutated D, C, and A sequence

elements in various combinations, as shown in Fig. 4A. When normal D, C, and A sequences were present, the construct functioned exactly as does $\Delta(-440)$, producing skeletogenic mesenchyme-specific expression in about 29% of experimental embryos. This result has two consequences. First, it proves that there is no detectable difference in transcriptional regulatory function whether the gene does or does not contain a TATA box sequence; second, it eliminates any possibility that sequence other than that included in the synthetic D, C, and A oligonucleotides is required for the regulatory function we observe. The further experiments, detailed in Fig. 4A, reproduce, with these entirely synthetic sequence elements, the same relationships seen in Fig. 1C. Thus, without a normal C sequence, expression is reduced to the level of the ectopic background expression, about 2% (this is shown by construct *SM50•TATA(DA)*). Addition of either the A or D oligonucleotide (*SM50•TATA(DC)* and *SM50•TATA(AC)*) to C partially restores expression, though D is more effective than A. Without either D or A, the C sequence element alone is not functional (*SM50•TATA(C)*). Thus it appears that for the C sequence element to promote expression, an ancillary positive function mediated by the D and/or A sequences is required. This is so even though the *SM50•TATA* constructs all include the distal region enhancer.

Regulatory function of synthetic D, C, and A elements in a vector composed entirely of sequences from other genes

We next examined the function of synthetic D, C, and A elements in vectors that contained no other *SM50* sequence. The construct *SV40(DCA)* (Fig. 1E) consists solely of the 200 bp enhancer-less SV40 early region promoter, which includes a TATA box and six Sp1 sites (Khoury and Gruss, 1983; Briggs et al., 1986), plus the synthetic D, C, and A oligonucleotides. While several transcription factors have been isolated from sea urchin embryo nuclear extracts that bind GC-rich target sites (Zeller et al., 1995; Hapgood and Patterson, 1994; Xiang et al., 1991) a sea urchin Sp1 factor has so far not been identified. Three of the Sp1 sites in the SV40 element, however, are also target sites for the sea urchin SpGCF1 factor (Zeller et al., 1995), which in context of various other factors functions as a positively acting regulator. Nonetheless, a vector consisting only of SV40 and the CAT reporter (*SV40•CAT*) has very little activity. The *SV40(DCA)* construct does express, however, and its expression is confined to skeletogenic mesenchyme. A quantitative comparison of CAT enzyme activity in embryos bearing *SV40(DCA)* and $\Delta(-440)$ is shown in Table 2. While the level of *SV40(DCA)* expression is only 16% of that produced by $\Delta(-440)$, it is still sufficient to produce significant whole-mount in situ hybridization signals in about 18% of the embryos. An embryo bearing *SV40(DCA)* is shown in Fig. 2E1,2.

In the course of the foregoing experiments we never observed significant ectopic expression; that is, none of the deletions or mutations we studied provided even a hint of the existence of negative spatial regulatory functions in the *SM50* cis-control system. The fact that *SV40(DCA)* is also expressed specifically in skeletogenic mesenchyme confirms that the spatial regulatory function carried out by the D, C, and A sequence elements is entirely positive. There are no other *SM50* sequences present in this construct. Therefore there is no
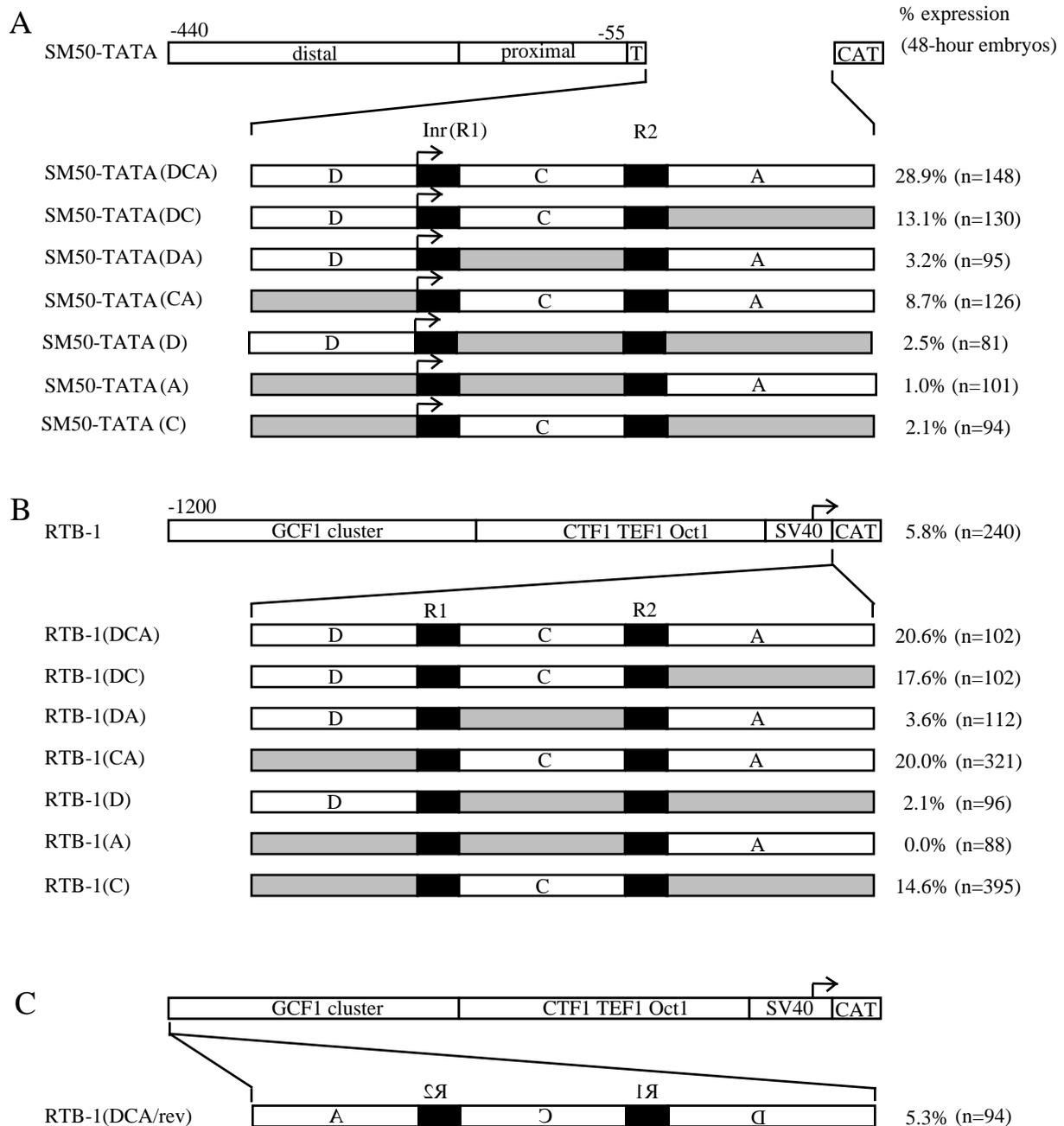
Fig. 4. Schematic diagram of injected constructs and the efficiency of their expression, as monitored by whole-mount in situ hybridization. Symbolism is as in Figs 1 and 3. The stippled regions in each construct indicate oligonucleotides representing the transversionally mutated natural sequence. (A) Constructs including one to three of the D, C, and A elements in context of the *SM50-TATA* vector. The synthetic TATA box is located at −55 (T), or −22 with respect to the beginning of the introduced initiator sequence. The exact sequence of this region is shown in Fig. 3C. (B,C) Constructs including combinations of normal and mutated D, C, and A elements in the context provided by the *RTB-1* vector. The synthetic A, C, D or mutated A, C, D elements were inserted between the SV40 promoter and the CAT gene (B), or into the most upstream position in a reversed orientation (C). Values in B represent summed data from three independent experiments which produced almost identical results. In the RTB-1 control group the 5.8% of embryos staining were in approximately equal parts stained in mesenchyme, in ectoderm, and in both. Thus the same small amount of ectopic (i.e., ectoderm) staining was also observed in the RTB-1(C) samples; i.e., it was not enhanced as was the mesenchyme staining by the addition of the C element. However about 10% of embryos bearing RTB-1(C) displayed some ectoderm staining, suggesting that the combination of 'A' factor with the RTB-1 enhancers produces a weak ectoderm localizing activity.

possibility that the accurate expression that it produces could be due to sequences outside of D, C and A that mediate negative control functions which we somehow missed in the foregoing experiments.

The control system isolated in *SV40(DCA)* also mediates the temporal activation of transcription seen earlier with Δ*(−440)*. Thus as Table 1 shows, the CAT enzyme content of embryos expressing this construct is 5- to 7-fold higher at gastrula stage

than at 21 hour mesenchyme blastula stage. Note that at all three stages there is about a six-fold ratio of expression of Δ(−440) to SV40(DCA). This suggests that the change in expression that we observe over this developmental period is entirely due to the DCA element, and not to other regions of Δ(−440), while the absolute level of expression at each stage does depend on the positive elements located elsewhere in the regulatory domain. From the experiments reviewed above, these positive elements are likely to be the second D element that is present in Δ(−440), and the distal region element.

To compensate for the relatively weak expression of SV40(DCA) we added to it two DNA fragments bearing positively acting enhancers from the regulatory domain of the CyIIIa gene, which is normally expressed only in the aboral ectoderm of the embryo. These are, a distal region of CyIIIa bearing a cluster of the SpGCF1 sites, and a second fragment bearing sites for three different positively acting factors, SpOct1, SpCTF1, and SpTEF1 (Thézé et al., 1990; Calzone et al., 1988; Coffman and Davidson, 1992; C. Kirchhamer, J. Xian and E. Davidson, unpublished data). None of these regulatory factors are believed to convey any territorial spatial regulatory information or to be spatially confined in their activity. These positive regulatory elements were associated with SV40(DCA) with the use of an inserted polylinker (see Materials and Methods). The resulting construct, RTB-1(DCA), expressed almost as well as Δ(−440), as shown in Table 2. Expression was observed in 21% of experimental embryos (Fig. 4B), and was again specific to the skeletogenic mesenchyme. In the absence of the DCA insert, RTB-1 produces only a low level of (mainly ectodermal) expression in a few percent of embryos developing from injected eggs. When the synthetic DCA insert was reversed in orientation, and positioned at the distal end of the RTB-1 construct (see Fig. 4C), no activity above that of RTB-1 alone was observed.

Finally we sought to determine whether the D, C, or A element alone could confer spatial specificity, based on the theory that RTB-1 might include sufficient positive enhancer functions to display the spatial specificity of any added element if it had one. The experiments summarized in Fig. 4B provide the answer. RTB-1(C), in which the only non-mutated SM50 sequence is a single C element, is expressed specifically in skeletogenic mesenchyme, in 14.6% of embryos. Examples of such embryos are shown in Fig. 2F1-3. In the absence of element C no significant expression at all is observed (RTB-1(DA), RTB-1(D), RTB-1(A); see Table 2, and Fig. 4B). Thus, the role of elements A and D in RTB-1(DCA) is merely to increase the activity of element C, and together with element C each of these is individually capable of increasing the level of expression (Table 2; constructs RTB-1(DC) and RTB-1(CA)). These experiments demonstrate that element C alone is able to convey spatial specificity for transcriptional expression in the skeletal mesenchyme lineages.

DISCUSSION

Functions of the D, C, and A cis-regulatory elements

The region of the SM50 regulatory domain that most of this study has focused on extends from −55 to about +105 with respect to the transcriptional start site. This sequence is capable

Table 2. CAT enzyme activity in embryos bearing various heterologous SM50 expression constructs

Construct	Enzyme units/embryo×10⁻⁵*
SM50Δ(−440)	1.8
SV40(DCA)	0.3
RTB-1(DCA)	2.4
RTB-1(D)	not detected
RTB-(C)	0.3
RTB-1(A)	not detected
RTB-1(DC)	1.5
RTB-1(DA)	not detected
RTB-1(CA)	0.7

*Each value represents the average of two independent samples, of 100 injected eggs per sample.

of directing skeletogenic mesenchyme-specific expression of a reporter gene, at a level of activity that is only a few-fold below that of any of the longer constructs studied by Sucov et al. (1988) or ourselves. We show that the essential elements within this region are the directly repeated D sequences just upstream of the initiator element; the C element, which lies within the sequence from +12 to +37, and the A element, which lies within the region from +71 to +105. Thus, the C and A sequence elements are included in the transcribed leader sequence of the gene, and the translational initiation site occurs only a few bp beyond the end of the A regulatory element. We have established the functions of the D, C, and A elements by the use of synthetic oligonucleotides that represent either normal or mutated variants of the sequences. These were used to rebuild the −55 to +105 region of the expression constructs in a variety of combinations, in the context of both the normal upstream SM50 sequence and of an entirely heterologous expression vector that contained no SM50 sequences other than the synthetic D, C, and A elements. In addition, we altered the promoter of the SM50 gene from a TATA-less to a TATA-box conformation by insertion of canonical TATA and initiator sequences, and we showed that this transformation has no effect whatsoever on the functions mediated by the D, C, and A sequence elements.

The C element is the single irreplaceable regulatory sequence of the minimal, skeletogenic lineage-specific cis-control system extending from −155 to +105. It is only one of the three SM50 elements which is alone capable of directing expression to skeletogenic mesenchyme cells, provided that the vector contains other elements that serve as enhancers. Entirely heterologous enhancers can perform this function, as shown by the experiments with the heterologous RTB-1 vector (Fig. 4B and Table 2). The RTB-1 enhancers cannot be skeletogenic mesenchyme-specific, since they derive from the CyIIIa gene, in which they promote expression in the aboral ectoderm. In the natural SM50 gene the A and D sequence elements provide the ancillary enhancer functions that the C element requires for activity; the A and D elements are dispensable only if other positively acting elements are provided (compare in Fig. 4, SM50-TATA(C) with RTB-1(C)). The measurements in Table 2, combined with the prior evidence obtained with specific mutations of D and/or A, show that both of these elements can work with C, and that each can function independently of the other. C is also the only element mutation which completely kills the ability of SM50 constructs to express. Although in the natural SM50 sequence the C element lies immediately adjacent

to the initiator sequence this is not essential for its function. Thus, in the *RTB-1* series of constructs the C element is separated from the initiator sequence by almost 100 bp of SV40 sequence, plus the length of the synthetic normal or mutated D sequence element, plus a linker added to provide convenient restriction sites, i.e., by about 130 bp. Furthermore, in the *SM50-TATA* and *RTB-1* constructs the D element is downstream of the initiator site, and is present only singly, while in the normal gene there are two copies of D upstream of the site of transcriptional initiation. Similarly, the A element in these constructs is adjacent to C, except for a linker sequence, while in the normal alignment it is separated from C by about 30 nonessential base pairs. Thus, there is clearly some flexibility in the range of spatial relations among these elements that remain permissive for function, as well as in their positions with respect to the assembly point for the basal transcription apparatus.

The C element must bind a transcription factor that is present or is functionally active only in skeletogenic mesenchyme lineages, or it could bind a generally present factor that in turn binds a lineage-specific co-factor. The C element in any case serves as the central regulatory information processor, by means of which the *SM50* gene is instructed that it is in a skeletogenic precursor cell. It is not by itself a 'tissue-specific enhancer' since it is completely inactive in the absence of other enhancer elements. We cannot exclude the possibility that the proteins interacting with the D and A elements are also present or functionally active only in skeletogenic mesenchyme cells, but our experiments clearly show that they need not be. It is unlikely that the C element is a binding site for a repressor that in skeletogenic mesenchyme cells is replaced by an activator, while elsewhere in the embryo the repressor serves to keep the *SM50* gene off. Were this the case, constructs such as *RTB-1(DA)*, in which the C element is mutated, would be expected to display enhanced ectopic expression; instead all constructs lacking a C element express very poorly, if at all.

The C sequence as a 'locator' element

It is curious that the C element can function with entirely heterologous enhancers as well as with its own natural retainers, A and D. This makes the idea that the binding of the C proteins serves as a focus for cooperative binding of the ancillary enhancer proteins less likely. An interesting speculation that satisfies our observations is that the lineage-specific protein that binds the C element (or its enhancer-specific co-factor) interacts directly with the basal transcription apparatus. Its function would be to enable or cause the basal transcription apparatus to respond to a range of enhancers, which may be present in several spatial domains of the embryo. That is, a corollary of this idea is that the basal transcription apparatus is not active in mediating upstream enhancer functions in the absence of a signal from the factor binding to the C sequence element (or from a cofactor bound in turn to the C factor). Thus the powerful *CyIIIa* elements that are included in the *RTB-1* construct are almost inactive in the absence of the C element, since by itself this vector produces only very low levels of expression, none of which is in the skeletogenic mesenchyme. In short, we propose that the primary spatial specificity of the *SM50* gene may be determined by a 'locator' element, which acts as an obligatory facilitator of the transcriptional enhancement functions of other DNA binding factors (see Fig. 5). This may not be an uncommon mechanism in regulatory systems that control spatial expression during development. For example, expression of the Drosophila *Adh* gene in the fat body depends on a *cis*-element located near the transcription start site, that like the *SM50* C-element confers responsiveness to more distant enhancers, though in their absence it fails to promote transcription (Fischer and Maniatis, 1988).

Overall organization of the *SM50* regulatory domain

Our knowledge of the functional significance of the sequence upstream of -55 is limited. There do not appear to exist any additional transcriptional control elements in the region beyond the two D sequence elements (i.e., from -55 to -200; this is shown by the results obtained with the *prox* series of mutations summarized in Fig. 1B). Of the distal region we know only that it includes positively acting elements which if present mediate several-fold greater activity. However, our preliminary data indicate that this region contains many closely packed target sites for DNA binding proteins, and the function of these interactions remains entirely unexplored. It is important to note that we have in no way excluded the possibility that the distal region could contain a second skeletogenic mesenchyme-specific positive control system.

To our knowledge *SM50* differs from every other gene so far studied that is expressed in a territory-specific manner in early sea urchin embryos, in that its spatial control system appears to lack any negative regulatory elements that prevent expression in other territories. Such elements are easily detected in gene transfer experiments, since deletions, mutations, or in vivo competitions affecting interactions at these elements result in the often spectacular ectopic spatial expression of test constructs. For example, the *CyIIIa* gene (Hough-Evans et al., 1990; Wang et al., 1995; C. Kirchhamer and E. Davidson, unpublished data), the *Spec1* gene (Gan et al., 1990), the *Endo16* gene (C.-H. Yuh and E. Davidson, unpublished data), and the *SM30* gene (T. Frudakis and F. Wilt, personal communication) are all controlled in part by negative spatial interactions. We saw no case, in this study, of a mutation or deletion that produces ectopic expression. Furthermore, we can state unequivocally that the spatial DCA control system functions only in a positive way. This is a particularly remarkable feature of the *SM50* gene, in that its regulatory domain contains no less than five sites at which bind a regulatory factor, SpP3A2, that is almost certainly a negative regulator of the *CyIIIa* gene (Zeller et al., 1995; Hough-Evans et al., 1990). Interactions at the P3A target sites of the *CyIIIa* gene repress oral ectoderm expression of this gene (C. Kirchhamer and E. Davidson, unpublished data). In the *SM50* gene these sites are at -354, -248, -128, -114, and -34. The double site at -114 and -128 has a particularly high affinity for this factor (Calzone et al., 1991). Yet, as Fig. 1 shows, every one of these sites was removed by one of the deletions or mutations tested, singly or in combination, and no ectopic expression was observed. Note, for example, that the *m(prox123J)* construct lacks both the strong double site and the site at -34, and yet its expression is quantitatively and spatially normal manner. The function of the P3A sites in the *SM50* gene is another mystery that remains to be explored. It is still possible that the *SM50* gene has a negative control system in the distal region, if both a negative functional element, and a positive functional element, the activity of which it modulates, are located together within this region. However, we can conclude that no negative

spatial control system exists between −200 and +120, and yet this *cis*-regulatory sequence suffices to produce very accurate skeletogenic mesenchyme-specific expression.

The *SM50* regulatory system and the embryology of skeletogenic lineage specification

The precursors of the skeletogenic mesenchyme are autonomously specified, as reviewed in the Introduction. This is significant in the present context, because an implication of an autonomous specification process is that the specific gene expressions that result could be mediated by localized maternal regulatory factors. The factor binding to element C has now been isolated by affinity chromatography (which affords the possibility of securing any co-factors as well), and it will soon be possible to test this proposition directly. If the factor binding to element C were indeed localized, or were present in active form only in the polar region of the egg inherited by the skeletogenic precursor cells, this would explain the absence of a negative control system in the *SM50* gene. That is, the initial territorial specification would be spatially confined by cytoplasmic localization process(es), rather than by intercellular signaling across territorial boundaries in the cleavage-stage embryo. The latter argument would hold whether the localized factor is a C-element binding transcription factor, a co-factor, a system that covalently activates these factors, or a factor that promotes the transcription of a C-element binding transcription factor or co-factor.

Under various experimental situations other blastomeres than the precursors of the skeletogenic mesenchyme can be induced to give rise to mesenchyme cells that secrete skeleton and express skeletogenic genes such as *SM50*. Thus isolated macromeres, which normally give rise only to vegetal plate and some ectoderm, will differentiate as embryoids that contain some skeleton (Hörstadius, 1939; Khaner and Wilt, 1991); animal pole mesomeres, if cultured for several days do the

same (Khaner and Wilt, 1990; Henry et al., 1989); and so do various combinations of isolated mesomeres and other blastomeres (Khaner and Wilt, 1990). A well known phenomenon that is particularly germane is the effect of LiCl on isolated animal pole blastomeres (Livingston and Wilt, 1989; Von Ubisch, 1929). This agent induces differentiation in prospective ectoderm cells from the animal half of the embryo of both skeletogenic mesenchyme and gut, and for the effect to occur, exposure to LiCl must take place during cleavage, i.e., the period when blastomere specification is initially occurring. One interpretation (Davidson, 1989) is that transcription factors required for skeletogenic gene expression are present globally but must be activated in order to become functional. Normally these factors would be presented in active form only in the skeletogenic mesomeres and their descendants. All the experimental operations that induce ectopic skeletogenesis may affect intercell signaling in one way or another, and thus might result in ectopic activation of a factor such as that binding to the C element of *SM50*. However, it is possible that in none of these experimental situations is there a *direct* conversion to skeletogenic fate. Instead the *only* relevant conversion may be to vegetal plate fate, which in each of these
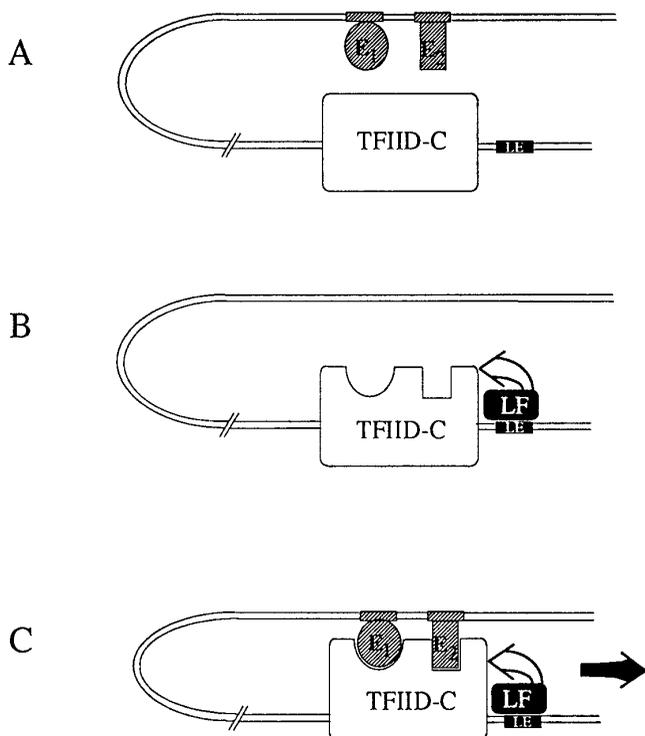
Fig. 5. *Locator function.* A speculative role for the factor binding the C sequence element in the *SM50* gene. E_1 and E_2 are enhancer proteins binding upstream regions of the *cis*-regulatory DNA (curved line); the basal transcription apparatus (TFIID-C), symbolized simply as a rectangle, indicates the holo-TFIID complex, including TBP and the TAF co-factors. A locator target sequence element (LE) is shown downstream of the TFIID complex, because that is the position of the C target site in the *SM50* gene. However, we show (see text) that this site may be either almost adjacent to the transcription initiation site or >100 bp downstream and still function, and thus it might also operate if placed immediately upstream of the start site. In (A), the regulatory system resides in a cell lacking the locator factor and the locator sequence is unoccupied. Enhancers E_1 and E_2 are present in this cell, but they have no effect on the basal transcription apparatus, and detectable transcription does not occur. This would be the case, e.g., in gut or ectoderm cells. Thus the TFIID-C requires an external additional signal in order to mediate upstream enhancer functions. In B, the regulatory system is shown in a cell type of the lineage in which the gene is specified to function, but at a stage in which the enhancer proteins E_1 and E_2 are lacking or inactive. This might be the case, e.g., in skeletogenic precursor cells prior to the stage in late cleavage when the gene is activated (see text). The locator sequence is occupied by the locator factor (LF) which recognizes and binds to its sequence specifically. Either directly or by means of a co-factor (curved double open arrow) the locator factor interacts with TAF co-factors of the TFIID complex, activating them, as symbolized by the specific depressions shown in their contours. However, in the absence of E_1 and E_2 there is still no detectable transcriptional activity. In C, E_1 and E_2 are present in this cell lineage, and they interact with the activated TAF co-factors that are able to service them (specific enhancer factors require different TAF co-factors, each of which may be able to interact with multiple enhancer proteins of given structural classes; see e.g., Chen et al., 1994). This would account for the ability of the activated regulatory system to respond to a wide range of different enhancer proteins. The consequence is detectable transcription (solid rightward arrow). Alternative versions of this model include the possibility that the LF actually recruits the appropriate TAFs to the TFIID complex; and/or that the LF interacts with the TFIID complex by means of an adaptor protein which is the factor that is lineage specific, rather than the DNA binding LF, as shown here.

examples is indeed also known to be induced by the experimental circumstances imposed. The vegetal plate gives rise to the archenteron and to secondary mesenchyme. When there is a deficiency of the normal primary skeletogenic mesenchyme, cells of the secondary mesenchyme are able to transdifferentiate to skeletogenic mesenchyme. This could account for the appearance of skeletal elements and skeletogenic gene expression, including *SM50* expression, in all of the experimental circumstances that we cite.

Conversion of secondary to skeletogenic mesenchyme fate at the gastrula stage is a remarkable phenomenon. It can be induced directly by experimental removal of skeletogenic mesenchyme cells (reviewed by Ettensohn and Ingersoll, 1992). A possible implication is that the *SM50* gene must be under some sort of negative control in the secondary mesenchyme since it is derepressed in these cells when conversion takes place. On the surface this is a proposition that runs counter to our suspicion that the *SM50* spatial control system lacks negative control elements. However, at this late period of embryogenesis the presentation of skeletogenic transcription factors almost certainly depends on their transcription *de novo*. Therefore it would be this process that is under negative control in the secondary mesenchyme.

In summary, we wish to stress the view that detailed functional knowledge of the *cis*-regulatory systems controlling spatial gene expression in the embryo will lead immediately to the generation of an invaluable set of new experimental probes. These can now be used to illuminate the mechanisms of classically defined embryological phenomena, in our case the initial autonomous specification, and the experimentally induced conditional activation, of skeletogenic cell lineages.

REFERENCES

Akasaka, K., Frudakis, T. N., Killian, C. E., George, N. C., Yamasu, K., Khaner, O. and Wilt, F. H. (1994). Genomic organization of a gene encoding the spicule matrix protein SM30 in the sea urchin *Strongylocentrotus purpuratus*. *J. Biol. Chem*. **269**, 20592-20598.

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

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.

Briggs, M. R., Kadonaga, J. T., Bell, S. P. and Tjian, R. (1986). Purification and biochemical characterization of the promoter-specific transcription factor Sp1. *Science* **234**, 47-52.

Calzone, F. J., Thézé, N., Thiebaud, P., Hill, R. L., Britten, R. J. and Davidson, E. H. (1988). Developmental appearance of factors that bind specifically to cis-regulatory sequences of a gene expressed in the sea urchin embryo. *Genes Dev*. **2**, 1074-1088.

Calzone, F. J., Höög, C., Teplow, D. B., Cutting, A. E., Zeller, R. W., Britten, R. J. and Davidson, E. H. (1991). Gene regulatory factors of the sea urchin embryo. I. Purification by affinity chromatography and cloning of P3A2, a novel DNA-binding protein. *Development* **112**, 335-350.

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

Chen, J.-L., Attardi, L. D., Verrijzer, C. P., Yokomori, K. and Tjian, R. (1994). Assembly of recombinant TFIID reveals differential coactivator requirements for distinct transcriptional activators. *Cell* **79**, 93-105.

Coffman, J. A. and Davidson, E. H. (1992). Expression of spatially regulated genes in the sea urchin embryo. *Curr. Opinion Genet. Dev*. **2**, 260-268.

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

Davidson, E. H. (1989). Lineage-specific gene expression and the regulative capacities of the sea urchin embryo: a proposed mechanism. *Development* **105**, 421-445.

Davidson, E. H. (1990). How embryos work: A comparative view of diverse modes of cell fate specification. *Development* **108**, 365-389.

Drager, B. J., Harkey, M. A., Iwata, M. and Whiteley, A. H. (1989). The expression of embryonic primary mesenchyme genes of the sea urchin, *Strongylocentrotus purpuratus*, in the adult skeletogenic tissues of this and other species of echinoderms. *Dev. Biol*. **133**, 14-23

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

Ettensohn, C. A. and Ingersoll, E. P. (1992). Morphogenesis of the sea urchin embryo. In *Morphogenesis* (eds. E. F. Rossomando and S. Alexander), pp. 189-262. New York: Marcel Dekker, Inc.

Fischer, J. A. and Maniatis, T. (1988). Drosophila *Adh*: a promoter element expands the tissue specificity of an enhancer. *Cell* **53**, 451-461.

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

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

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

Franks, R. R., Anderson, R., Moore, J. G., Hough-Evans, B. R., Britten, R. J. and Davidson, E. H. (1990). Competitive titration in living sea urchin embryos of regulatory factors required for expression of the *CyIIIa* actin gene. *Development* **110**, 31-40.

Gan, L., Wessel, G. M. and Klein, W. H. (1990). Regulatory elements from the related Spec genes of *Strongylocentrotus purpuratus* yield different spatial patterns with a *lacZ* reporter gene. *Dev. Biol*. **142**, 346-359.

George, N. C., Killian, C. E. and Wilt, F. H. (1991). Characterization and expression of a gene encoding a 30.6 kDa *Strongylocentrotus purpuratus* spicule matrix protein. *Dev. Biol*. **147**, 334-342.

Gorman, C., Moffat, L. F. and Howard, B. H. (1982). Recombinant genomes which express chloramphenicol acetyltransferase in mammalian cells. *Mol. Cell. Biol*. **2**, 1044-1051.

Hapgood, H. and Patterton, D. (1994). Purification of an oligo(dG)•oligo(dC)-binding sea urchin nuclear protein, suGF1: a family of G-string factors involved in gene regulation during development. *Mol. Cell Biol*. **14**, 1402-1409.

Harkey, M. A., Whiteley, H. R. and Whiteley, A. H. (1992). Differential expression of the msp130 gene among skeletal lineage cells in the sea urchin embryo: A three dimensional in situ hybridization analysis. *Mech. Dev*. **37**, 173-184.

Henry, J. J., Amemiya, S., Wray, G. A. and Raff, R. A. (1989). Early inductive interactions are involved in restricting cell fates of mesomeres in sea urchin embryos. *Dev. Biol*. **136**, 140-153.

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

Hough-Evans, B. R., Britten, R. J. and Davidson, E. H. (1988). Mosaic incorporation of an exogenous fusion gene expressed exclusively in aboral ectoderm cells of the sea urchin embryo. *Dev. Biol*. **129**, 198-208.

Hough-Evans, B. R., Franks, R. R., Zeller, R. W., Britten, R. J. and Davidson, E. H. (1990). Negative spatial regulation of the lineage specific *CyIIIa* actin gene in the sea urchin embryo. *Development* **108**, 41-50.

Kabakoff, B., Hwang, S.-P. L. and Lennarz, W. J. (1992). Characterization of post-translational modifications common to three primary mesenchyme cell-specific glycoproteins involved in sea urchin embryonic skeleton formation. *Dev. Biol*. **150**, 294-305.

Katoh-Fukui, Y., Noce, T., Ueda, T., Fujiwara, Y., Hashimoto, N., Higashinakagawa, T., Killian, C. E., Livingston, B. T., Wilt, F. H., Benson, S. C., Sucov, H. M. and Davidson, E. H. (1991). The corrected structure of the *SM50* spicule matrix protein of *Strongylocentrotus purpuratus*. *Dev. Biol*. **145**, 201-202.

Kaufmann, J. and Smale, S. T. (1994). Direct recognition of initiator elements

by a component of the transcription factor IID complex. *Genes Dev.* **8**, 821-829.

Khaner, O. and Wilt, F. (1990). The influence of cell interactions and tissue mass on differentiation of sea urchin mesomeres. *Development* **109**, 625-634.

Khaner, O. and Wilt, F. (1991). Interactions of different vegetal cells with mesomeres during early stages of sea urchin development. *Development* **112**, 881-890.

Khoury, G. and Gruss, P. (1983). Enhancer elements. *Cell* **33**, 313-314.

Killian, C. E. and Wilt, F. H. (1989). The accumulation and translation of a spicule matrix protein mRNA during sea urchin embryo development. *Dev. Biol.* **133**, 148-156.

Leaf, D. S., Anstrom, J. A., Chin, J. E., Harkey, M. A., Showman, R. M. and Raff, R. A. (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.

Leahy, P. S. (1986). Laboratory culture of *Strongylocentrotus purpuratus* adults, embryos, and larvae. *Meth. Cell Biol.* **27**, 1-13.

Livant, D., Cutting, A., Britten, R. J. and Davidson, E. H. (1988). An in vivo titration of regulatory factors required for expression of a fusion gene in transgenic sea urchin embryos. *Proc. Natl. Acad. Sci. USA* **85**, 7607-7611.

Livant, D. L., Hough-Evans, B. R., Moore, J. G., Britten, R. J. and Davidson, E. H. (1991). Differential stability of expression of similarly specified endogenous and exogenous genes in the sea urchin embryo. *Development* **113**, 385-398.

Livingston, B. T. and Wilt, F. H. (1989). Lithium evokes expression of vegetal-specific molecules in the animal blastomeres of sea urchin embryos. *Proc. Natl. Acad. Sci. USA* **86**, 3669-3673.

McClay, D. R., Armstrong, N. A. and Hardin, J. (1992). Pattern formation during gastrulation in the sea urchin embryo. *Development 1992 Supplement* 33-41.

McMahon, A. P., Flytzanis, C. N., Hough-Evans, B. R., Katula, K. S., Britten, R. J. and Davidson, E. H. (1985). Introduction of cloned DNA into sea urchin egg cytoplasm: Replication and persistence during embryogenesis. *Dev. Biol.* **198**, 420-430.

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

Ransick, A., Ernst, S., Britten, R. J. and Davidson, E. H. (1993). Whole mount in situ hybridization shows *Endo16* to be a marker for the vegetal plate territory in sea urchin embryos. *Mech. Dev.* **42**, 117-124.

Richardson, W., Kitajima, T., Wilt, F. and Benson, S. (1989). Expression of an embryonic spicule matrix gene in calcified tissues of adult sea urchins. *Dev. Biol.* **132**, 266-269.

Ruffins, S. W. and Ettensohn, C. A. (1993). A clonal analysis of secondary mesenchyme cell fates in the sea urchin embryo. *Dev. Biol.* **160**, 285-288.

Seed, B. and Sheen, J.-Y. (1988). A simple phase-extraction assay for chloramphenicol acetyltransferase activity. *Gene* **67**, 271-277.

Sucov, H. M. (1989). Characterization and developmental regulation of a gene expressed specifically in the skeletogenic lineage of the sea urchin embryo. Ph.D. Thesis, California Institute of Technology, Pasadena, CA.

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

Sucov, H. M., Hough-Evans, B. R., Franks, R. R., Britten, R. J. and Davidson, E. H. (1988). A regulatory domain that directs lineage-specific expression of a skeletal matrix protein gene in the sea urchin embryo. *Genes Dev.* **2**, 1238-1250.

Thézé, N., Calzone, F. J., Thiebaud, P., Hill, R. L., Britten, R. J. and Davidson, E. H. (1990). Sequences of the *CyIIIa* actin gene regulatory domain bound specifically by sea urchin embryo nuclear proteins. *Mol. Reprod. Dev.* **25**, 110-122.

Von Ubisch, L. (1929). Über die Determination der larvalen Organe und der Imaginalanlage bei Seeigeln. *Wilhelm Roux's Arch. Entwicklungsmech. Org.* **117**, 81-122.

Wang, D. G.-W., Kirchhamer, C. V., Britten, R. J. and Davidson, E. H. (1995). SpZ12-1, a negative regulator required for spatial control of the territory-specific CyIIIa gene in the sea urchin embryo. *Development* **121**, 1111-1122.

Xiang, M., Lu, S.-Y., Musso, M., Karsenty, G. and Klein, W. H. (1991). A G-string positive *cis*-regulatory element in the LpS1 promoter binds two distinct nuclear factors distributed non-uniformly in *Lytechinus pictus* embryos. *Development* **113**, 1345-1355.

Zeller, R. W., Coffman, J. A., Harrington, M. G., Britten, R. J. and Davidson, E. H. (1995). SpGCF1, a sea urchin embryo transcription factor, exists as five nested variants encoded by a single mRNA. *Dev. Biol*, **169**, in press.