

Modular *cis*-regulatory organization of *Endo16*, a gut-specific gene of the sea urchin embryo

Chiou-Hwa Yuh and Eric H. Davidson

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

SUMMARY

The *Endo16* gene of *Strongylocentrotus purpuratus* is expressed at the blastula stage of embryogenesis throughout the vegetal plate, at the gastrula stage in the whole of the archenteron and in postgastrular stages only in the midgut. We showed earlier that a 2300 bp upstream sequence suffices to faithfully recreate this pattern of expression when fused to a CAT reporter gene. Here we define the functional organization of this *cis*-regulatory domain, which includes over thirty high specificity binding sites, serviced by at least thirteen different putative transcription factors, in addition to >20 sites for a factor commonly found in the regulatory sequences of other sea urchin genes as well (SpGCF1). The *Endo16 cis*-regulatory domain consists of several different functional elements, or modules, each containing one or two unique DNA-binding factor target sites, plus sites for factors binding in other modules as well. Modular regulatory function was defined in experiments in which regions of the *cis*-regulatory DNA containing specific clusters of sites were tested in isolation, combined with one another, or by selective deletion, and the effects on expression of the CAT reporter were determined by whole-mount in situ hybridization or CAT enzyme activity measurements. The most proximal module (A) is mainly responsible for early embryonic expression, and module A alone suffices to locate expression in the vegetal plate and archenteron. The adjacent module (B) is responsible for a steep postgastrular rise in expression, when the gene is transcribed only in the midgut and, prior

to this, module B alone also suffices to promote expression in the vegetal plate and archenteron. The most distal module, G, acts as a booster for either A or B modules. However, no combination of A, B and G modules generates vegetal plate or gut expression exclusively. Ectopic expression of A-, B- and G-CAT fusion constructs occurs in the adjacent (*veg1*-derived) ectoderm and in skeletogenic mesenchyme cells. For expression to be confined to endoderm requires negative regulatory functions mediated by modules E, F and DC. Modules E and F each repress ectopic expression specifically in *veg1* ectoderm. Module DC represses ectopic expression specifically in skeletogenic mesenchyme. Expression of some *Endo16* constructs is dramatically increased by treatment with LiCl, which expands the territory in which the endogenous *Endo16* gene is expressed at the expense of *veg1* ectoderm. The same modules that act to repress ectopic expression in untreated embryos are required for enhanced expression of constructs after LiCl treatment. Furthermore, both the negative spatial control functions and response to LiCl require the presence of module A. The total regulatory requirements of the *Endo16* gene during embryogenesis can be expressed in terms of the positive and negative functions of the individual modules and the interactions between modules that are identified in this study.

Key words: vegetal plate specification, *cis*-regulatory module, spatial control, temporal control, gene regulation, *Endo16*

INTRODUCTION

Early in sea urchin development the blastomeres are segregated into polyclonal territories of specific embryonic fate. The cleavage stage specification processes by which the territorial founder cells are defined result in differential spatial presentation of sets of transcription factors. By early blastula stage, territorial transcription of genes encoding terminal differentiation products has been initiated (reviewed by Davidson, 1989, 1990, 1991). Expression of each such gene is controlled by multiple, specific interactions between DNA-binding transcriptional regulators and the target site sequences that they recognize within its *cis*-regulatory domain. *Cis*-regulatory domains can be thought of as 'hardwired' information processing systems. The

functions of these systems in embryogenesis are, initially, to interpret the outcome of the territorial specification processes and, later, to mediate spatial states of gene expression that reflect ongoing morphogenetic interactions. This paper concerns the internal functional organization of the *cis*-regulatory domain that controls the embryonic expression of the *Endo16* gene of *Strongylocentrotus purpuratus*.

Endo16 encodes a polyfunctional cell surface glycoprotein that in late embryos is secreted by the cells of the midgut (Nocente-McGrath et al., 1989, 1991; Soltysik-Espanola et al., 1994). The gene is activated at the early blastula stage, and whole-mount in situ hybridization (WMISH) carried out in blastulae reveals *Endo16* transcripts in all the cells of the eight clones of blastomeres constituting the vegetal plate (Ransick

et al., 1993). These cells are the progeny of the *veg2* ring of Hörstadius (1973; i.e., of the founder cells VOM1 and 2, VAM 1 and 2, and right and left VLM1 and 2 of Cameron et al., 1987). Specification of the vegetal plate founder cells depends at least in part on signals emanating from the large micromeres, which are the sister cells of the grandparents of the vegetal plate founder cells (VOM, VAM, and right and left VLM), and/or the immediate descendants of the micromeres. Thus transplantation of micromeres to the animal pole produces an ectopic second vegetal plate and then a complete second gut, which expresses *Endo16* in the 'normal' pattern (Ransick et al., 1993). Removal of the micromeres at 4th or 5th cleavage severely inhibits vegetal plate specification and abolishes or depresses normal endogenous *Endo16* expression in the vegetal plate (Ransick and Davidson, 1995). After invagination, *Endo16* is expressed throughout the archenteron, but expression is extinguished in the secondary mesenchyme as these cells delaminate from the archenteron tip (30–48 hours), and expression then disappears from the embryonic foregut (60 hours). By the pluteus stage, expression has also disappeared from the hindgut, while becoming stronger in the midgut (Ransick et al., 1993). It appears that cells of the definitive hindgut initially express *Endo16* and then turn the gene off, as do the cells of the foregut (Ransick et al., 1993; A. Ransick, personal communication). A major component of the hindgut consists of derivatives of the *veg1* blastomeres (C. Logan and D. McClay, personal communication; A. Ransick and R. Cameron, personal communication), which roll in over the blastopore late in gastrulation. In normal embryos, *veg1* derivatives, which remain outside the blastopore and do not express *Endo16*, constitute the region of the aboral ectoderm lying immediately adjacent to the blastopore (whether there is a *veg1* contribution to ectoderm on the oral side in this species also remains unclear).

In recent studies, we showed that a fragment of DNA extending 2300 bp upstream of the transcription start site suffices to generate the complete embryonic pattern of *Endo16* expression, when associated with a CAT reporter gene (Yuh et al., 1994). High specificity DNA-protein interactions within this *cis*-regulatory domain were identified and catalogued. By means of fine-scale oligonucleotide competition mapping, combined with the use of affinity chromatography as an analytical tool to distinguish each DNA-binding protein by its size and its target site sequence, we identified 13 different protein factors that interact specifically at target sites within the *Endo16* regulatory DNA sequence. Nine of these factors recognize target sites that occur at unique locations in the *cis*-regulatory domain, while the other four bind at multiple sites located in several different subregions of the domain; the total number of target sites recognized specifically by these 13 factors exceeds 30. In addition, we found >20 sites for the multimerizing, DNA-binding factor SpGCF1 (Zeller et al., 1995a,b; Kirchhamer and Davidson, 1996; Calzone et al., 1988). As shown in Fig. 1A of this paper (reproduced from Yuh et al., 1994), the target binding sites occur in what appears to be a somewhat discontinuous, clustered arrangement within the *Endo16* regulatory domain. Each cluster includes one or two of the unique sites, plus several of the multiply occurring sites. Using this intriguing arrangement as an initial guide, we constructed a large number of *Endo16*•CAT derivatives and determined both their spatial and temporal patterns of expression. We conclude that the *Endo16* *cis*-regulatory system is modular in functional organization, in

that subregions containing sites for specific sets of DNA-protein interaction carry out particular regulatory roles. Some of the modules function positively, mediating expression in vegetal plate and gut, while others function negatively, eliminating expression in the territories that initially lie above and below the boundaries of the vegetal plate, i.e., the *veg1* ectoderm and the skeletogenic mesenchyme, respectively. The overall function of the *Endo16* *cis*-regulatory system is the sum of the functions of the individual modules and of the specific interactions amongst them.

MATERIALS AND METHODS

Animals and embryos

Adult *Strongylocentrotus purpuratus* collected along the Southern California coast were maintained in chilled sea water at Caltech's Kerckhoff Marine Laboratory, so that gametes were available year round (Leahy, 1986). Preparation of gametes, fertilization and embryo culture were carried out according to standard methods. The embryos were microinjected as described (McMahon et al., 1985), using linearized plasmid DNA. Approximately 1500 molecules of the desired plasmid DNA were injected together with a four-fold molar excess of *HindIII*-digested carrier sea urchin DNA per egg in a 2 μ l volume of 30% glycerol and 0.12 M KCl (Franks et al., 1990).

Whole-mount in situ hybridization

The whole-mount in situ hybridization protocol used here is based on the method described by Ransick et al. (1993), modified slightly by the use of Streck Tissue Fixative (Streck Laboratories Inc.). Fixation was carried out overnight at 4°C, followed by a brief wash in sea water. The proteinase K treatment and postfixation procedure of Ransick et al. (1993) were omitted.

Expression constructs

The 38 constructs that are diagrammed in Fig. 1B were assembled in various ways from restriction fragments, bounded either by the restriction sites indicated in Fig. 1A, by polylinker sites present in subclones, or by restriction sites present in an earlier generation of vectors that were derived from the *Endo16*•CAT expression construct described by Yuh et al. (1994). Details are available on request.

CAT enzyme measurements

CAT enzyme activity was determined in lysates of 100 embryos. Samples were collected at various stages in development and spun for 5 minutes in an Eppendorf tube. The sea water was quickly removed from the pellet, to avoid detaching embryos from the bottom. The embryos were resuspended in 100 μ l of 0.25 M Tris (pH 8.0) and freeze-thawed at least three times by cycling from –70°C for 1 minute to 37°C for 2 minutes. The samples were exposed to 65°C for 10 minutes to destroy an activity that inhibits CAT enzyme activity.

CAT enzyme purchased from Pharmacia was used to provide standards of 0.5, 1, 2, 4 and 8×10^{-3} units in total volumes of 100 μ l Tris buffer (0.25 M, pH 8.0). Enzyme reactions were carried out at 37°C for 12 hours, in the presence of 1 μ l of 14 C-chloramphenicol (Amersham, 2.07 GBq/mmol), 20 μ l of 4 mM acetyl Coenzyme A (Sigma) and 29 μ l of 0.25 M Tris (pH 8.0). Following the reaction, 1 ml of ethyl acetate was used to extract the organic phase by vortexing for 30 seconds, followed by centrifugation for 5 minutes. The organic phase was transferred to an Eppendorf tube and dried in a speed vacuum device. The dried samples were resuspended in 20 μ l of ethyl acetate and spotted on TLC plates for chromatography, which was performed in a closed chamber saturated with 95% chloroform and 5% methanol. TLC plates were air dried and autoradiographed.

After developing the film, equal areas of the TLC that contained the acetylated and non-acetylated forms of the substrate were cut out

and counted in a scintillation counter. To convert units of CAT enzyme activity to CAT enzyme molecules per embryo, we utilized as a conversion factor, 2.6×10^7 molecules of CAT enzyme = 10^{-4} units of activity (McMahon et al., 1984).

RESULTS

Spatial expression of a fusion gene controlled by *Endo16* upstream sequence

Our earlier studies (Yuh et al., 1994) were carried out with *Endo16* fusion constructs that included not only the upstream 2300 bp DNA fragment represented in Fig. 1A, but also a 1.4 kb intron that interrupts the N-terminal portion of the protein coding sequence. This construct (*Endo16*•CAT) included the *Endo16* transcription start site, the whole of Exon I, in which *Endo16* translation begins, Intron I and Exon II, down to residue 40. At this point, a fusion with the CAT reporter gene just upstream of its translation initiation site was effected, using a naturally occurring *Sall* site (see Yuh et al., 1994). For the present study, we made use of restriction sites within the 1.4 kb intron of *Endo16*•CAT to remove all but 106 bp from the 5' end and 11 bp from the 3' end of the intron (see legend to Fig. 1A for details). A new fusion gene including the same upstream sequence as *Endo16*•CAT but lacking all but these 117 bp of Intron I was thus constructed (*GFEDCBA*•Bp•CAT, construct 1 of Fig. 1B; in our nomenclature 'Bp' represents the basal promoter fragment indicated in Fig. 1, as discussed in detail below). When construct *GFEDCBA*•Bp•CAT was injected into fertilized eggs and its expression studied by WMISH, it was found to express identically with *Endo16*•CAT. The amount of CAT enzyme produced by these two constructs is also identical (data not shown). Thus the ~1.3 kb of intronic sequence missing from *GFEDCBA*•Bp•CAT contains no regulatory sites the effect of which are noticeable by the test of deletion. All experiments described in the following were carried out with constructs containing only the 117 bp remnant of Intron I.

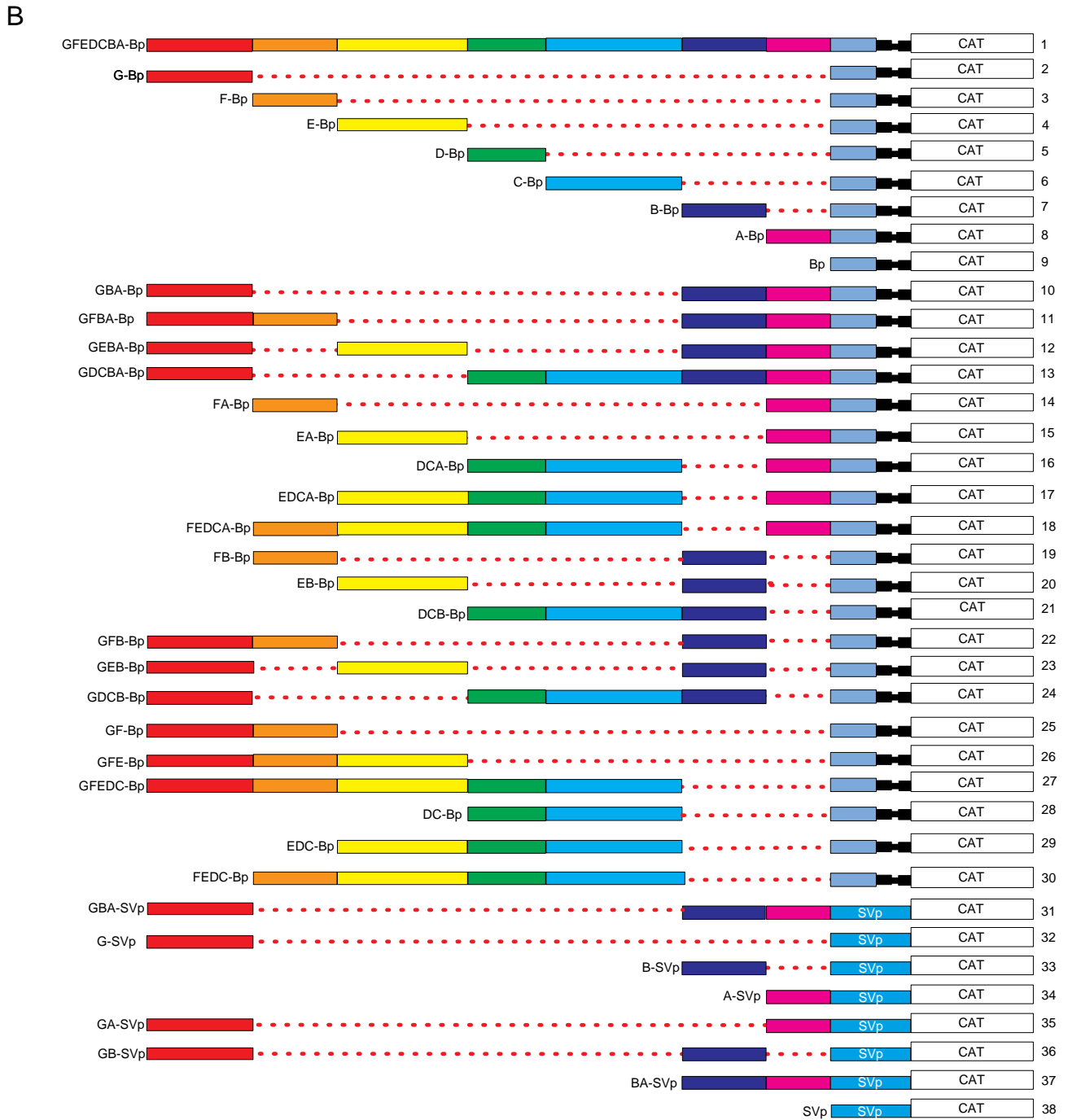
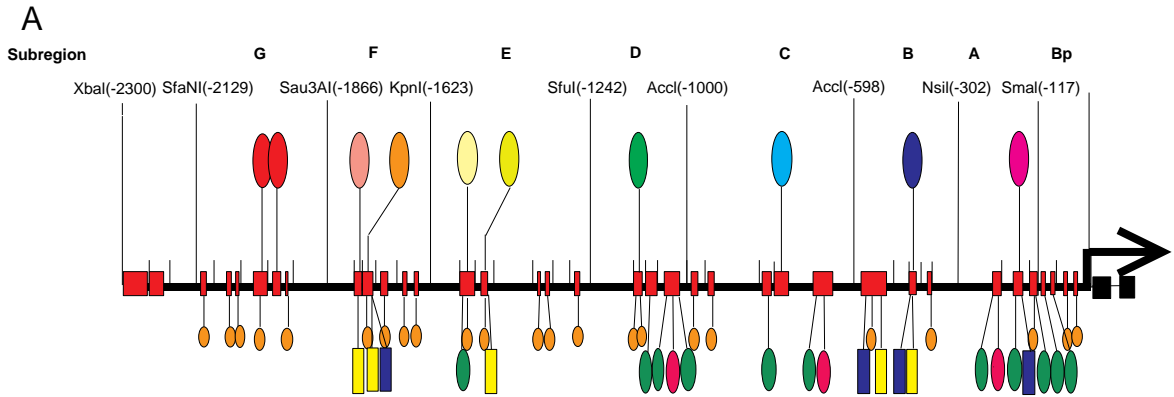
GFEDCBA•Bp•CAT is expressed with good accuracy, producing CAT transcripts in the vegetal plate at blastula stage, in all regions of the archenteron in early gastrulae and in midgut only in pluteus-stage embryos. Since they appear identical to the WMISH hybridizations shown for *Endo16*•CAT in Yuh et al. (1994) we do not illustrate these results here. A quantitative summary of WMISH data for *GFEDCBA*•Bp•CAT in 48 hour late gastrula-stage embryos is shown in the first column of Table 1. About 70% of injected embryos display labeling and, of these, an average of 97% express CAT RNA in the endodermal territories, i.e., the midgut, hindgut and/or blastopore rim. The 70% value for labeling is consistent with expectation, since incorporation of exogenous DNA injected into sea urchin eggs is mosaic (see legend to Table 1). In respect to the territory in which *Endo16* is initially expressed, all 2nd cleavage blastomeres, one-half of 3rd cleavage blastomeres and one-fourth of 4th cleavage blastomeres will contribute a clone of vegetal plate cells (Cameron et al., 1987). Thus 70% labeling indicates an appreciable frequency of early cleavage incorporations, and it follows that the labeled embryos should each display several stained cells, since each vegetal plate founder cell (see Introduction) gives rise to eight vegetal plate cells by the end of cleavage. In the experiments of Table 1A the average number of cells stained per embryo in fact ranged from about five to seven.

Table 1 shows that, although almost every embryo expresses *GFEDCBA*•Bp•CAT in the gut, there is also a low level of ectopic expression. Throughout this work, all staining observed in WMISH experiments was counted only if there were ≥ 2 cells stained per embryo. An average of 5.8% of embryos express CAT RNA in ectoderm and 3.9% in skeletal mesenchyme. Ectopic expression of a few percent of embryos is commonly observed, irrespective of the fusion gene used. This has been observed, for example, for *CyIIIa*•CAT (Kirchhamer and Davidson, 1996), for *Spec2A* fusions with β -Gal reporters (Gan et al., 1990) and for *SM50*•CAT (Makabe et al., 1995). The level of random ectopic expression measured for *GFEDCBA*•Bp•CAT represents the quantitative background against which the results below are to be compared (background ectopic expression consists typically of scattered cells displaying staining outside of the ectodermal domains). Background expression levels characteristically decline with advancing developmental stage and are likely to be correlated with unstable exogenous DNA incorporation.

Modular *Endo16* cis-regulatory elements that individually promote expression in the endoderm

The 2300 bp cis-regulatory domain was divided into seven subregions, named G-A, and a basal promoter (Bp) region, as indicated in Fig. 1A, by means of naturally occurring restriction sites that were predicted by the sequence (Yuh et al., 1994). In all but two cases, these restriction sites lay within gaps in the array of factor binding sites that we had mapped earlier. Except in these two cases, each subregion thus consists of a natural cluster of target sites. Each subregion includes one or two unique sites at which factors bind that interact in no other subregion of the overall domain [these factors are indicated above the line representing the DNA in Fig. 1A, as well as sites for factors that bind as well in other subregions (indicated below the line)]. The two restriction sites that lie within, rather than between, clusters of target sites for DNA-binding factors are that separating subregions D and C, and that separating subregion A from Bp. In fact, we found that subregion C has no function on its own, and is more properly considered part of a module that includes both subregions D and C. Similarly, the upstream boundary of Bp, at position -117, is arbitrary. When truncated to this position (*Bp*•CAT, construct 9 in Fig. 1B), the fusion gene expresses at a low level that is barely detectable by WMISH. That is, only 13% of embryos produced visible staining ≥ 2 cells (Table 1C), although 90-100% of injected embryos contain stably incorporated DNA. CAT enzyme activity measurements presented in Fig. 4 below show that in fact 48 hour embryos bearing *GFEDCBA*•Bp•CAT contain about $7.5 \times$ more of the reporter gene product than do embryos bearing *Bp*•CAT. *Bp*•CAT does not express in gut. The Bp subregion acts as a basal promoter in the sense that it includes the *Endo16* transcription start site, and that it promotes active expression if associated with appropriate upstream regulatory sequences, though it is quite inactive on its own. However, the Bp subregion includes three binding sites for a factor that also interacts in subregions A, DC and E, as well as two SpGCF1 sites (Zeller et al., 1995a,b), and we do not know how many of these sites, if any, are actually essential for its experimentally defined basal promoter function.

Constructs 2-8 of Fig. 1B each consist of a single subregion of the *Endo16* regulatory domain linked to *Bp*•CAT. These were



injected into eggs and assayed by WMISH as above; data are shown in Table 1A,B. There is a dramatic distinction between the single subregion constructs included in Table 1A, all of which promote strong expression of CAT RNA in the gut in 89-95% of labeled embryos, and those of Table 1B, none of which promote any detectable expression in gut. Since subregions A, B and G are all independently capable of generating endoderm (+) expression patterns on linkage to *Bp•CAT*, we henceforth refer to them as *modules A, B and G*. The levels of expression produced by *A-Bp•CAT*, *B-Bp•CAT* and *G-Bp•CAT* are far above the sensitivity limit of the WMISH procedure and thus, as Table 1 shows, all yield ~70% stained embryos, just as does the complete construct *GFEDCBA-Bp•CAT*. However, Table 1 also indicates that *A-Bp•CAT*, *B-Bp•CAT* and *G-Bp•CAT* differ from the complete construct, in that they produce 2-3x the background levels of ectopic expression in both ectoderm and skeletogenic mesenchyme.

Table 1B shows that, when linked to *Bp•CAT*, subregion C is entirely inactive and the activity of subregion D is barely detectable (only 11% of embryos display labeling, the same

fraction as generated by *Bp•CAT* alone). Subregions E or F probably include some very weak, positively acting elements, as they produce detectable staining in about 40-45% of embryos; however, this staining is confined to only a few cells per embryo (av. 2-3). The expression generated by *E-Bp•CAT* and *F-Bp•CAT* is entirely ectopic. Subregion F produces expression almost exclusively in skeletogenic mesenchyme, and subregion E in both ectoderm and skeletogenic mesenchyme.

An important point is that all of the ectopic ectodermal expression beyond background observed by WMISH in the experiments summarized in Table 1 occurs in regions of the embryo immediately adjacent to the vegetal plate, or in later embryos, to the blastopore. Thus in late blastula-early gastrula stage embryos, ectopic ectodermal expression beyond background is always confined to the progeny of the *veg1* cell tier and it never extends above the equator of the embryo. In pluteus-stage embryos, the circumblastoporal ectoderm that is the locus of ectopic expression is also of *veg1* origin. In any case, with respect to the initial processes of specification, when ectopic expression is observed, it occurs exclusively in the progeny of the two early embryonic territories that initially border the founder cells of the vegetal plate territory; i.e., the skeletogenic mesenchyme territory below, and the *veg1* oral and aboral ectoderm territories above.

Distinct quantitative and temporal functions of the endoderm (+) modules A, B and G

When tested in single module constructs the three endoderm (+) modules do not produce identical WMISH patterns, even though they possess in common the ability to promote expression of the CAT reporter in the vegetal plate and archenteron. This became apparent when we began to examine WMISH displays of embryos bearing *A-Bp•CAT*, *B-Bp•CAT* and *G-Bp•CAT* at different stages of development. Fig. 2A-C illustrates a typical experiment with the *B-Bp•CAT* construct. CAT RNA is present in (a) a clone of vegetal plate cells at 30 hours, (b) archenteron cells at 48 hours and (c) a clone of midgut cells at 72 hours. This is essentially the same pattern as displayed by *GFEDCBA-Bp•CAT* (not shown) or by *Endo16•CAT* (Yuh et al., 1994). But *A-Bp•CAT* gives a somewhat different result. This construct is expressed particularly well at 30 hours and 48 hours, as illustrated in Fig. 2D,E; however, very little staining can be discerned at 72 hours (not shown). Expression of *G-Bp•CAT* is significantly weaker than that of either *A-Bp•CAT* or *B-Bp•CAT*; it often occurs in fewer cells, as illustrated in Fig. 2F,G; and it is often difficult to detect at all in later embryos (not shown).

The endodermal patterns of expression produced by *GBA-Bp•CAT* are illustrated in Fig. 2H-J. Again, expression is observed in clones of cells in the vegetal plate, the archenteron and, finally, the midgut at 30 hours, 48 hours and 72 hours, respectively. However, just as do the individual A, B and G constructs, *GBA-Bp•CAT* produces a significant incidence of ectopic expression, over twice background, in both ectoderm and skeletogenic mesenchyme. Ectopic ectodermal expression typically occurs in mesenchyme, probably in patches or clones of cells. WMISH data are summarized in column 1 of Table 2. Examples of embryos expressing *GBA-Bp•CAT* ectopically in skeletogenic mesenchyme, in addition to the appropriate expression in archenteron, are shown in Fig. 2N-P and ectopic expression of *GBA-Bp•CAT* in adjacent patches of ectoderm,

Fig. 1. Cis-regulatory domain of the *Endo16* gene and fusion constructs utilized in this study. (A) Cis-regulatory domain with DNA-binding factors indicated, from Yuh et al. (1994). The nine different factors, indicated by colored ovals, above the heavy solid line representing the DNA, bind at sites which occur uniquely within the indicated subregions; the four different factors indicated below the line by green and pink ovals and by blue and yellow rectangles, each bind at multiple target sites which occur in more than one of the indicated subregions. These subregions are named G-A at the top of the Figure. Bp, basal promoter (see text). The subregions are bounded by the indicated restriction sites, the coordinates of which, with respect to the transcription start site (bent arrow), are shown in parentheses. Orange ovals indicate SpGCF1 target sites (Zeller et al., 1995a,b). Locations of target binding sites and diversity of DNA-binding factors were established by gel shift, fine-scale oligonucleotide competition mapping, affinity chromatography and a variety of other procedures, as described by Yuh et al. (1994). The solid boxes at the right-hand end of the figure represent, respectively, the 26 bp *Endo16* non-translated leader sequence of exon I (exon extends for 61 bp in total) and the 59 bp portion of exon II that precedes the CAT gene in all of the constructs shown in part B of this figure. Exon I and exon II are shown separated by the truncated 117 bp intron that was included in all of these constructs, modified from the natural 1.4 kb intron. The intron interrupts the *Endo16* coding sequence in codon 21. We do not know whether the remaining intron sequence is spliced out efficiently on expression. If so, translation of the fusion product would continue to begin at the *Endo16* start codon in Exon I; if not, the start codon of the CAT gene must be used, since there are three in-frame stop codons within the remaining 117 bp of intron sequence. Thus it is not known if the CAT products of the fusion genes described in this figure include the 40 amino acid protein sequence encoded in exon I and the fragment of exon II that these constructs contain. (B) CAT fusion constructs. Nomenclature, indicating the subregions included in each construct (see A) is given at the left end of each diagram. The subregions are color coded, as indicated in construct 1, *GFEDCBA-Bp•CAT*. Where present, the dotted red lines indicate the positions of the regulatory domain that are missing from each construct. Constructs 1-30 all include the *Endo16* basal promoter (Bp) shown in light blue, arbitrarily shown here as the region from -117 to +1. In constructs 31-38, Bp, exon I, the truncated intron, and the position of exon II preceding the CAT gene sequence are replaced by the SV40 early region promoter (see Materials and Methods), here denoted SVp, shown in darker blue.

Table 1. WMISH data for each module of ENDO16 promoter (48 hour embryos)

CONSTRUCTS†	A. ENDO (+) constructs*				B. ENDO (-) constructs*				C. Bp*
	GFEDCBA-Bp	A-Bp	B-Bp	G-Bp	C-Bp	D-Bp	E-Bp	F-Bp	Bp
TOTAL EMBRYOS‡	147	232	340	113	48	239	167	111	85
EXPTS§	1,2,3	1,2,6	2,5,6,7	2,7,8	8	6,7	5,6	1	1
TOTAL LABELED(%)¶	70	74	72	74	0	(11)‡‡	(41)‡‡	(45)‡‡	(13)
LABELED IN:**									
MG/HG, BP	92	142	163	70	0	0	0	0	0
ECTO	1	6	12	2	0	10	15	0	2
MES	1	3	11	3	0	16	49	48	8
ENDO+ECTO	5	13	25	7	0	0	0	0	0
ENDO+MES	3	5	13	1	0	0	0	0	0
ECTO+MES	0	0	3	0	0	1	4	2	1
ENDO+ECTO+MES	0	2	4	1	0	0	0	0	0
LABELED EMBRYOS (%):‡‡									
ENDO	97	95	89	94	0	0	0	0	0
ECTO	5.8	12.3	19	12	0	41	28	4	27
MES	3.9	5.9	13	6	0	63	78	100	82

*Endo (+) indicates subregions of the regulatory domain that when associated with Bp•CAT produce CAT RNA in vegetal plate and gut; Endo (-) indicates subregions that fail to produce CAT RNA in endodermal cells. Bp, arbitrarily defined "basal promoter," including sequence from -117 to +190 (see text).

†Constructs included in this table are 1-9 of Fig. 1B.

‡Total embryos: All embryos deriving from injected eggs, labeled plus nonlabeled.

§Individual experiments are denoted by numbers, each carried out on a single batch of eggs.

¶Labeled/total embryos × 100. Labeled embryos were counted as such only if two or more cells were stained. As noted in text, the values shown in A are consistent with expectations on the basis that incorporation is mosaic: within minutes the DNA forms one or a few large concatamers and, in 75% of cases, the concatamers are incorporated stably into one nucleus of a 2nd, 3rd or 4th cleavage blastomere, in an entirely random fashion with regard to lineage (McMahon et al., 1985; Flytzanis et al., 1985; Franks et al., 1988; Hough-Evans et al., 1988; Livant et al., 1991). The exogenous DNA then replicates clonally together with that of the host cell lineage.

**MG, midgut; HG, hindgut; BP, blastopore; ECTO, ectoderm (i.e., veg1 derived ectoderm; see text); MES, skeletogenic mesenchyme. MG/HG, BP denotes the sum of embryos labeled in MG and/or HG and/or BP. Labeled cells in any of these regions were counted only if there were two or more cells per region.

‡‡Percentage of total labeled embryos displaying staining in indicated compartment of embryo, irrespective of where else they display staining. Thus, e.g., ENDO indicates the sum of (MG/HG, BP) + (ENDO + ECTO) + (ENDO + MES) + (ENDO + ECTO + MES).

‡‡‡These values are in parentheses because in contrast to the results obtained with the endoderm (+) constructs only two to three cells per embryo displayed labeling. Thus as shown below (Fig. 4), no significant increase in activity is actually produced by subregions D, E and F when linked to Bp.

as well as in archenteron, is shown in Fig. 2Q-T. These and the foregoing experiments suggest that the portions of the *Endo16* regulatory domain missing from all of these constructs, i.e., subregions F, E, D and C, must include targets for negative regulators that in the complete construct act to repress expression outside of the endodermal domains of the embryo.

To obtain quantitative time course data for the expression of these constructs, we turned to measurement of CAT enzyme activity. Fig. 3A illustrates the temporal functions of each of the positive single module constructs, plus that of *GFEDCBA-Bp•CAT* and of a combination of the three endoderm (+) modules with *Bp•CAT*, viz *GBA-Bp•CAT* (construct 10 of Fig. 1B). Two major conclusions devolve from the experiments shown in Fig. 3A: each module indeed displays a unique temporal profile of CAT expression; and when combined in *GBA-Bp•CAT*, the time function of their activity is indistinguishable from that of the complete construct, *GFEDCBA-Bp•CAT*.

The data of Fig. 3A are replotted on a per-cell basis in Fig. 3B. For this calculation, the number of cells expressing *Endo16*, i.e., the number of cells in vegetal plate, archenteron and midgut at early and late blastula, gastrula and postgastrular stages, were compiled from Ransick et al. (1993), Ransick and Davidson (1995), Burke (1980) and unpublished data. These values are shown in the inset at the top of Fig. 3B. It can now be seen clearly that only module B possesses the ability to

reproduce the late embryo surge in per-cell expression, though module B drives some detectable early expression as well. At the 72 hour datapoint, on a per-cell basis, module B is evidently the main contributor. Expression of *Endo16* or *GFEDCBA-Bp•CAT* from this period onward is confined to the midgut, and therefore module B must contain target sites for transcription factors that are present and active in the midgut. In contrast, the contribution of module A is most significant in the earlier periods of embryogenesis. At 20 hours, the 64 cells constituting the vegetal plate have just begun to express *Endo16* and, by 30 hours, the number of expressing cells in the vegetal plate has increased to about 105; during this period the CAT enzyme activity generated by *A-Bp•CAT* is almost the same as that of *GBA-Bp•CAT* or of the complete construct. Thus, during this earliest phase of *Endo16* expression, module A appears to function as the major positive regulator. Module A remains important into the late gastrula stage and, at 48 hours, the level of activity of *A-Bp•CAT* still accounts for about two-thirds of the total activity displayed by *GBA-Bp•CAT* (Fig. 3B). The curves describing the expression of *A-Bp•CAT* and *B-Bp•CAT* cross only at 60 hours. Thus, to summarize, the periods of development when module A functions are during the processes of vegetal plate specification and invagination of the archenteron, while module B becomes dominant as expression is limited to the differentiating midgut and is stepped up to significantly higher per-cell levels. As can be seen in Fig. 3C, when

Table 2. WMISH data for GBA, GFBA, GEBA, GDCBA-BpCAT (48 hour embryos)*

CONSTRUCTS	GBA-Bp	GFBA-Bp	GEBA-Bp	GDCBA-Bp
TOTAL EMBRYOS	210	93	168	200
EXPTS	2,3,4	3,4	9, 10	9, 10
TOTAL LABELED(%)	76	72	41	35
LABELED IN:				
MG/HG, BP	121	58	47	50
ECTO	6	0	0	4
MES	4	1	4	0
ENDO+ECTO	16	3	1	14
ENDO+MES	10	5	15	1
ECTO+MES	1	0	0	0
ENDO+ECTO+MES	1	1	2	0
LABELED EMBRYOS (%):				
ENDO	94	99	94	94
ECTO	15	5.7	4.3	26
MES	9.4	10.4	30.4	1.4

*See legend to Table 1 for abbreviations and presentation. Constructs used in these experiments are 10-13 of Fig. 1B.

tested in isolation module G acts as a relatively weak positive regulator throughout.

The synergistic function of module G

The experiments of Fig. 3A and B do not greatly illuminate the significance of module G, and to explore this issue further we utilized constructs employing a heterologous basal promoter. Fig. 2K-M shows that, when linked to the enhancerless SV40 early region promoter (SVp) (Khoury and Gruss, 1983; Briggs et al., 1986), modules G, B and A together (*GBA-SVp•CAT*; Fig. 1B, construct 31) produce endoderm (+) expression that is identical with that generated by *GBA-Bp•CAT*. Under our assay conditions, by either WMISH or CAT enzyme activity measurements, *SVp•CAT* alone (Fig. 1B, construct 38) is almost inactive. However, both the absolute peak value and the time course of expression of CAT enzyme production by *GBA-SVp•CAT* are quantitatively similar to those of *GBA-Bp•CAT* (compare the expression of *GBA-SVp•CAT* in Fig. 3D to that of *GBA-Bp•CAT* in Fig. 3A). Therefore, SVp serves as an adequate basal promoter for the positively acting elements of the *Endo16* regulatory system.

When linked individually with *SVp•CAT*, however, G, B and A generate levels of activity that are only about 10% of those produced by the same module when tested with *Bp•CAT*. Absolute CAT enzyme values are shown for constructs *A-SVp•CAT* (Fig. 1B, construct 34), *B-SVp•CAT* (construct 33) and *G-SVp•CAT* (construct 32) in Fig. 3C, where they can be compared directly with the corresponding data for *A-Bp•CAT*, *B-Bp•CAT* and *G-Bp•CAT* in Fig. 3A. It is remarkable that the ratios of the activities generated by the SVp constructs to those generated by the corresponding Bp constructs are about the same for all three modules. Furthermore, despite their low levels of activity, the temporal profiles of all the SVp constructs closely resemble those of the equivalent Bp constructs. Once again, module B responds late in development; the activity of module A peaks at 40-50 hours, and then declines; and module G produces only a low and almost constant activity profile. It follows that the temporal profiles of expression of

these three modules are entirely due to the interactions occurring within them and not to interactions occurring within Bp. In contrast, Bp is about 10× more active than is SVp with all three upstream modules. Some or all of the interactions for which Bp includes target sites but SVp does not, must contribute powerfully to the probability of transcriptional initiation. SVp inadvertently includes two consensus sites for the sea urchin factor SpGCF1 (Zeller et al., 1995a). But since Bp also contains two SpGCF1 sites, the enhanced transcriptional activity displayed by the Bp constructs is probably due not to SpGCF1 interactions, but to the factor indicated by the green ovals in Fig. 1A; this factor binds at three specific sites within the Bp sequence (see Yuh et al., 1994 for characterization).

The low background activity of SVp permitted us to observe easily what we believe to be the key property of module G (we have carried out similar experiments with equivalent constructs based on the *Endo16* Bp, which lead to the same conclusions; data not shown). As shown in Fig. 3D, when module G is combined with module A (*GA-SVp•CAT*, construct 35 of Fig. 1B), the peak level of activity is raised about fourfold (compare the peak value for *GA-SVp•CAT* with that of *A-SVp•CAT* in Fig. 3C). This effect is much greater than a simple addition of the levels of activity achieved by *G-SVp•CAT* and *A-SVp•CAT* (Fig. 3C) would predict. Similarly, in absolute terms, *GB-SVp•CAT* is more than twice as active in late embryos than is *B-SVp•CAT*. We conclude that module G is able to synergistically step up the level of expression of both module A and module B. When modules B and A are combined in *BA-SVp•CAT* (construct 37 of Fig. 1B), the main features of the temporal activity profiles of both individual modules remain evident in that both the early expression peak of module A and the late activity peak of module B can still be seen. However, the absolute value of CAT enzyme production by *BA-SVp•CAT* is >10-fold greater than that of either *B-SVp•CAT* or *A-SVp•CAT* at any point in the time course. Therefore module A and module B also function synergistically. Furthermore, since *GBA-SVp•CAT* is two fold more active than is *BA-SVp•CAT*, this effect of module G is evident in this combination as well. These intermodule effects fully compensate, in terms of absolute per-cell activity, for the relative weakness of SVp when assayed with only one of the upstream elements at a time. We examine the quantitative nature of the relationships between modules A, B and G in further detail elsewhere; for our present purposes, the main conclusions are that module G synergistically enhances the activities of both modules A and B without changing their different temporal profiles and that modules A and B similarly enhance each other's activity level, again without changing the temporal character of either activity profile.

Negative functions of modules E, F and DC

To explore directly the implication in the foregoing that subregions F, E and/or DC might contain negative regulators of spatial expression, we carried out WMISH experiments with several additional constructs. These were designed to utilize the combined endoderm (+) construct *GBA-Bp•CAT* as a test system for ectopic expression, by addition of each of the subregion fragments F, E and DC. These elements were inserted in their natural positions, i.e., between modules G and B (see Fig. 1B for diagrams of *GFBA-Bp•CAT*, construct 11; *GEBA-Bp•CAT*, construct 12; *GDCBA-Bp•CAT*, construct 13). Table 2 reveals that the central subregions of the *Endo16 cis-*

regulatory domain indeed repress ectopic expression and that they do this specifically with respect to embryonic lineage.

Ectopic expression in ectoderm occurs in an average of 15% of labeled embryos expressing *GBA-Bp•CAT* and addition of either the F or the E subregion returns the incidence of ectopic ectoderm expression to the background level observed with the control *GFEDCBA-Bp•CAT* construct (Table 2; for background values see Table 1). Since the E and F subregions are able to carry out this function independently, we provisionally regard them as separate regulatory modules. Aside from the ubiquitous SpGCF1 target sites, Fig. 1A shows that the only species of target site included in both E and F modules is that for the factor represented by the yellow boxes in this figure (see Yuh et al., 1994 for characterization). Since the same factor also binds in module B, but construct *B-Bp•CAT* also expresses ectopically in the ectoderm (in 19% of labeled embryos; Table 1), the presence of this factor alone is not sufficient for control of ectopic ectoderm expression.

An important result shown in Table 2 is that neither modules F nor E repress ectopic expression in skeletogenic mesenchyme, and indeed module E appears to increase this ectopic expression, probably due to the weak mesenchyme expression produced by this module, as just discussed. The DC subregion, however, completely eliminates all ectopic skeletogenic mesenchyme expression driven by GBA, but it in turn fails to exercise any control over ectopic ectodermal expression. Thus DC functions as a modular repressor, the activity of which is confined to the skeletogenic mesenchyme lineage of the embryo. We note, in addition, that both modules E and DC appear to decrease the overall levels of expression, so that, as Table 2 shows, only 41% or 35% of embryos, respectively, display labeling, rather than the ~70% normally produced by all endoderm (+) constructs. No ectopic expression was ever observed in cells that by their positions could be unequivocally identified as secondary mesenchyme.

We next sought to determine which of the endoderm (+) elements the repressor modules interact with. We carried out a large number of experiments on constructs of the form *X-Bp•CAT*, *GXBA-Bp•CAT*, *XA-Bp•CAT*, *XB-Bp•CAT*, *GX-Bp•CAT* and *GXB-Bp•CAT* where X is F, E, DC or combinations thereof. The constructs were injected in sets, as indicated, and measurements were made on batches of 100 embryos per construct at 48 hours of development. Results are summarized in Fig. 4 (see legend). For these studies, we

measured CAT enzyme production in order to estimate quantitatively the overall decrease in activity, if any, that would be mediated by the repressor linked into each construct. The *X-Bp•CAT* series of constructs in Fig. 4 shows that, while the individual F, E and DC elements affect the background level of Bp function less than ~2-fold, when they are combined, much stronger negative effects are observed. Thus, for example, *FEDC-Bp•CAT* produces only about a tenth the activity of *Bp•CAT*. It follows that the additive negative effects of the F, E and DC modules are at least capable of being exerted directly on the Bp transcriptional apparatus.

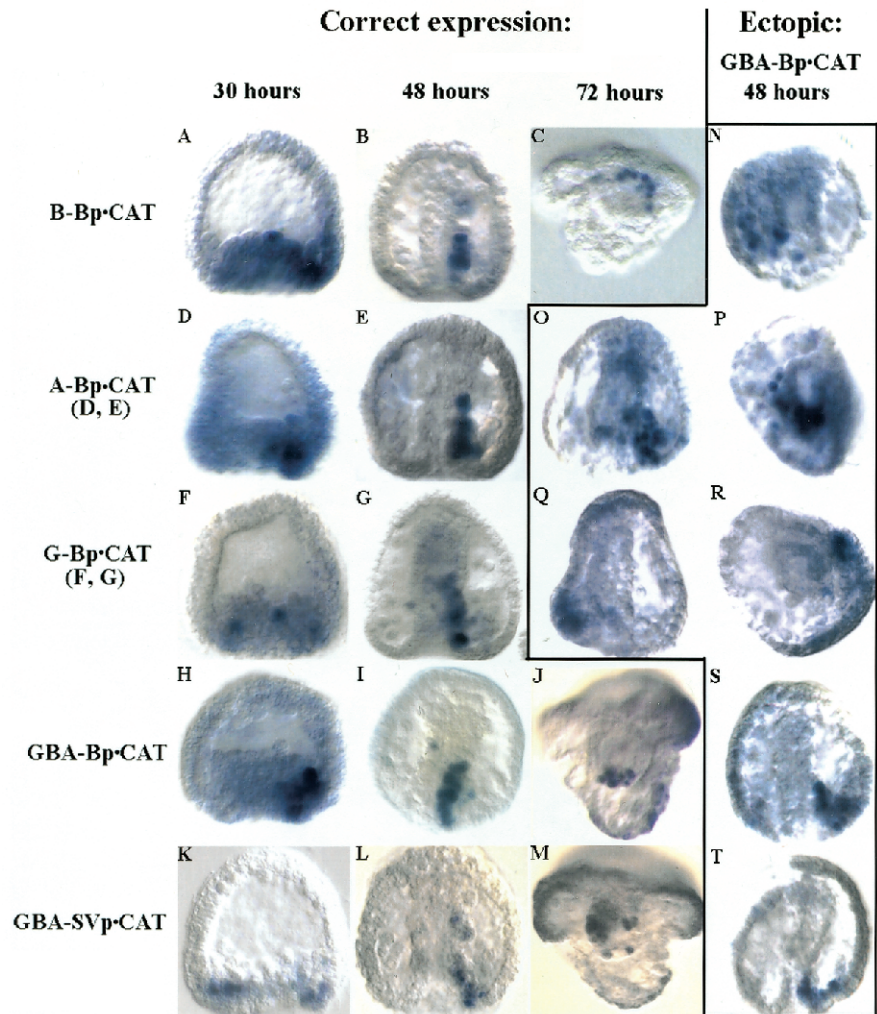


Fig. 2. Representative patterns of expression of *Endo16* constructs visualized by WMISH. Constructs are indicated on left margin (cf. Fig. 1B) and at the top right (referring to examples within heavy solid lines, i.e., N-T). Examples are shown for late blastula/early gastrula (30 hour), late gastrula (48 hour) and pluteus (72 hour) stages. The blue stain indicates the location of CAT RNA generated from these constructs, which were injected into fertilized eggs. Representative illustrations are shown here. Digital images were generated on a Roche Imaging Camera (Prog. Res. 3012) mounted on a Zeiss Axioskop microscope operated in DIC mode and printed on a Nikon CP-3000. A-M all display patterns of expression that are correct for the endogenous *Endo16* gene (Ransick et al., 1993), i.e., expression in the vegetal plate at 30 hours, in the future midgut and hindgut regions of the archenteron at 48 hours, and in the midgut at 72 hours. Ectopic expression is shown within the boxed area. N, O and P display ectopic expression in skeletogenic mesenchyme, as well as expression in the archenteron, and Q, R, S, T display ectopic expression in ectoderm cells close to the blastopore, as well as in the archenteron.

The *GXBA-Bp•CAT* series of Fig. 4 (constructs 11-13 of Fig. 1B) display modest negative effects on expression levels. In each case, the result is to decrease total positive GBA activity about twofold, which is about what would be expected based on the cellular contributions to the respective territories. Slightly less negative effect is obtained with the complete construct, *GFEDCBA-Bp•CAT* (i.e., where X = FEDC). Nonetheless, all experiments of the *GXBA-Bp•CAT* series display a negative effect of the X component that is highly significant, given the small standard errors shown in the Fig. 4. The *XA-Bp•CAT* series (constructs 14-18 of Fig. 1B) demonstrates the same negative effects of the individual modules when the only endoderm (+) element in the construct is A (compare results of *XA-Bp•CAT* constructs to *A-Bp•CAT*). However, much more severe inhibition is observed when the repressor modules are present in combination, as in constructs *EDCA-Bp•CAT* and *FEDCA-Bp•CAT*. We cannot yet rationalize the relative magnitude of the negative effects observed in this series of experiments, although the most clear negative effects are obtained with module A. Thus the *XB-Bp•CAT* constructs and *GX-Bp•CAT* in general express within twofold of the levels of *B-Bp•CAT* and *G-Bp•CAT*, respectively. The small inhibitions obtained with modules B and G could be due to the direct effects of the repressors on Bp function noted above. In summary, major repressive effects are seen only with module A, the activity of which is decreased by factors of 5 (*FEDCA-Bp•CAT*) and 12 (*FDCA-Bp•CAT*) when multiple repressor modules are linked to *A-Bp•CAT*. We conclude that the function of the repressor modules is mediated largely through the endoderm (+) module A. This makes biological sense, since module A is the element mainly responsible for early expression in response to the initial specification events, and it is these events that set the boundaries between the vegetal plate and the contiguous ectodermal and skeletogenic territories. In the following section, we explore the function of the negative regulatory modules when the vegetal plate/ectoderm boundary is specifically perturbed by the teratogenic agent LiCl.

Functional role of modules F, E and DC in LiCl-treated embryos

LiCl treatment of late cleavage-early blastula state sea urchin embryos increases the vegetal plate territory at the expense of surrounding ectoderm (Hörstadius, 1973; Livingston and Wilt, 1993). Ransick et al. (1993) showed that the spatial domain of expression of the *Endo16* gene is correspondingly enlarged in LiCl-treated embryos. Though the exact cell lineage and cell specification changes caused by LiCl treatment remain to be described, all or most of the progeny of the *veg1* cell tier are included in the expanded vegetal plate and archenteron, and also express *Endo16* in the treated embryos. In normal untreated embryos, this whole lineage remains in the ectodermal wall during blastula and early-mid gastrula stages, though toward the end of embryogenesis some of the *veg1* progeny eventually enter the hindgut. Under our conditions, LiCl

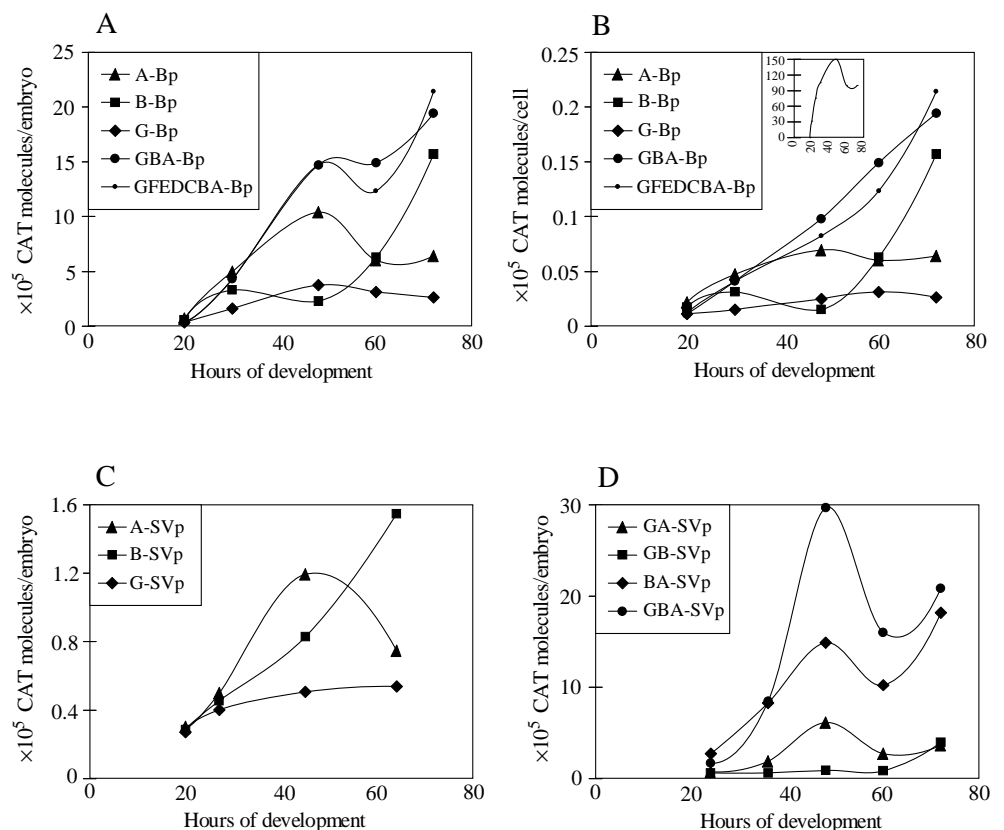


Fig. 3. Quantitative temporal output of positively acting *Endo16* regulatory modules. Indicated constructs (see Fig. 1B) were injected into fertilized eggs and CAT enzyme activity was measured as described in Materials and Methods in samples of 100 embryos at each time period. All the data included in each panel were obtained at once from a single batch of eggs in a single experiment. The experiments shown are representative. (A) Expression profiles of the complete construct (*GFEDCBA-Bp•CAT*), and of the endoderm (+) modules G, B and A tested individually (*A-Bp•CAT*, *B-Bp•CAT* and *G-Bp•CAT*), and in combination (*GBA-Bp•CAT*). Data are shown as average CAT enzyme molecules per embryo. (B) Data from A, but calculated to give average CAT enzyme molecules per cell of the population of cells that express the *Endo16* gene at each time point. These values are shown in the inset. This shows the number of cells expressing *Endo16* gene (ordinate) as a function of time postfertilization (hours, abscissa); see text for sources of data. (C) Average CAT enzyme production per embryo as in A, for modules A, B and G, in association with the SV40 early promoter (i.e., constructs *A-SVp•CAT*, *B-SVp•CAT* and *G-SVp•CAT*). (D) Average CAT enzyme production per embryo for combinations of module G with modules A or B (constructs *GA-SVp•CAT* and *GB-SVp•CAT*); for the combination of modules A and B (construct *BA-SVp•CAT*), and for the combination of all three endoderm (+) modules (*GBA-SVp•CAT*).

treatment does not cause expression of *Endo16* in either skeletogenic or secondary mesenchyme (Ransick et al., 1993).

Fig. 5 shows, remarkably, that the same functional relationships hold within the *Endo16* *cis*-regulatory domain for response to LiCl as for prevention of ectopic ectodermal expression in normal embryos. Thus modules F and E are needed for both functions, and both are mediated *via* module A. The first series of experiments, i.e., those carried out with the *X-Bp•CAT* constructs, demonstrate that Bp alone is insensitive to LiCl, and that neither modules F, E nor DC by themselves confer LiCl sensitivity on Bp. However, when modules F and E are linked to *GBA-Bp•CAT*, as shown in the second series of experiments summarized in Fig. 5, LiCl treatment results in a two- to three-fold increase in expression. In contrast, *GBA-Bp•CAT* itself is entirely insensitive to LiCl. Therefore, the effect of LiCl in expanding the domain of *Endo16* expression is not mediated through an enlargement of the spatial domain in which the positive regulators that service the A, B or G modules are presented. Rather, the expansion of *Endo16* construct depends on interactions mediated by the repressor modules of the regulatory system. Furthermore, as Fig. 5 shows clearly, the *XA-Bp•CAT* constructs all display a strong LiCl enhancement, while *A-Bp•CAT* is again blind to LiCl treatment. However, LiCl has no effect whatsoever on *XB-Bp•CAT*, *XG-Bp•CAT* or *XGB-Bp•CAT* expression.

WMISH observations on LiCl-treated embryos bearing *GFEDCBA-Bp•CAT* (not shown) indicate expression of CAT RNA in all regions of the expanded vegetal plate, as expected from the effects of LiCl on the endogenous *Endo16* gene. We assume that the expansion in number of cells expressing CAT RNA in LiCl-treated embryos directly contributes to the increased amount of CAT enzyme produced by all constructs that include modules F, E and A (Fig. 5). However, the function of module DC is in this regard paradoxical, in that in untreated embryos this module does not in any way affect ectopic expression in ectoderm, but only in skeletogenic mesenchyme. Ectopic mesenchymal expression of *Endo16* constructs including DC was observed by WMISH in LiCl-treated embryos (not shown), but in these morphologically abnormal embryos the identity of the ectopically expressing cells was difficult to ascertain.

DISCUSSION

Modular organization of the *Endo16* *cis*-regulatory domain

At least thirteen different DNA-binding proteins interact specifically in the 2300 bp *Endo16* *cis*-regulatory domain (Yuh et al., 1994). These interactions occur at 33 specific

target binding sites, in addition to the ubiquitous SpGCF1 sites, of which >20 have been identified (Yuh et al., 1994; Zeller et al., 1995a,b). In this work, we have shown that the *cis*-regulatory domain can be subdivided into six clusters of sites. We term these site clusters regulatory modules, because we have identified a specific regulatory role for each. The individual modular functions can be observed separately or in combination with other modules. We do not here address the individual significance of any of the binding factors identified by Yuh

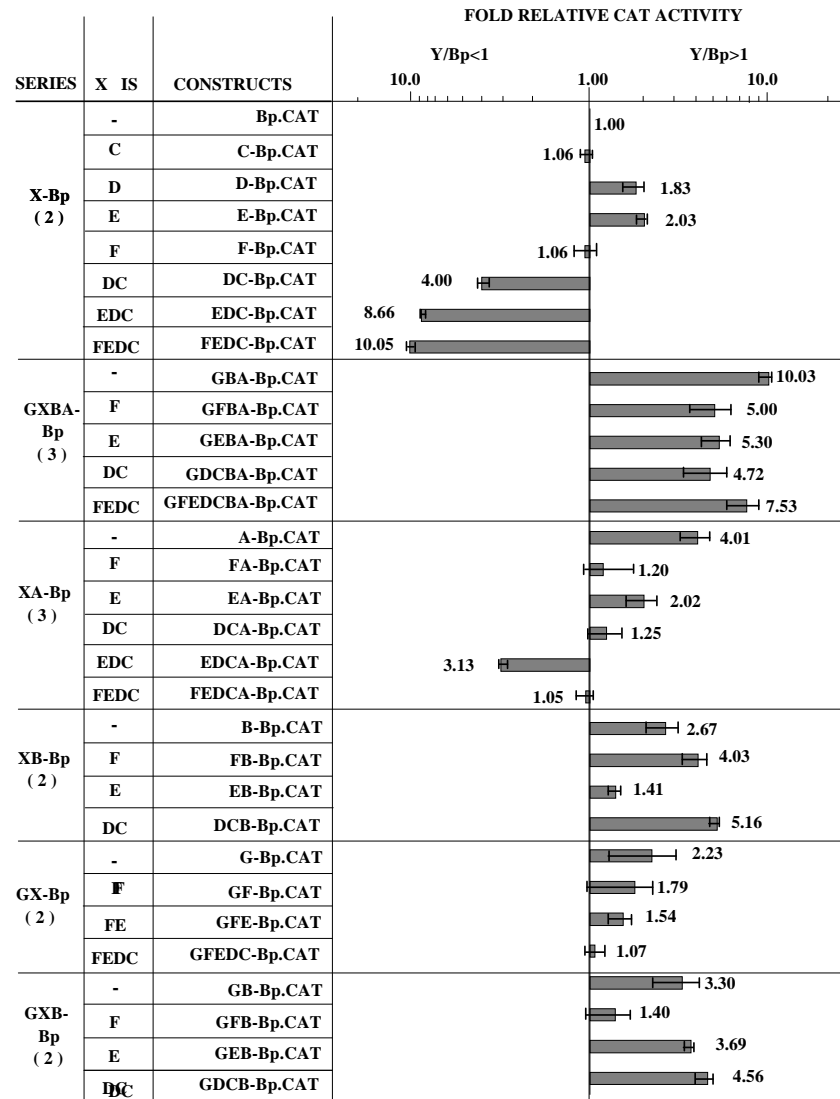


Fig. 4. Negative functions of modules F, E and DC assayed by CAT enzyme production, in context of various endoderm (+) modules. Constructs (see Fig. 1B) are indicated as series of the form given in the first column, where X denotes the modules listed in the second column. Each series of measurements was obtained on a single batch of 48 hour embryos, 100 embryos per measurement sample. Data are shown from two or three experiments on different batches of eggs, as indicated in the first column (parentheses). Results of replicate experiments were closely comparable, as shown by the standard errors indicated. The histogram expresses on a log₁₀ scale the *fold* change in CAT enzyme production relative to the CAT enzyme production of *Bp•CAT*, which was included in each series of measurements. The absolute value of the *Bp•CAT* enzyme level measured in these experiments was 2.5×10^5 molecules of active CAT enzyme per average embryo. Thus where the CAT enzyme production of the construct (Y) was less than that of Bp, the log₁₀ of the ratio Y/Bp appears to the negative side of 0, (i.e., log₁₀ (1)), and where Y/Bp > 1 the value of the log₁₀ of the ratio appears on the positive side.

et al. (1994), and the focus of this study is rather on the functions of the modular site clusters *per se*. It remains very possible that, although the discontinuous distribution of sites served as a guide in physically separating one module from another, some sites could have been included in given modules that functionally belong instead to adjacent modules, and until the individual interactions are each evaluated the boundaries of the modules that we have defined functionally will remain somewhat arbitrary. It might be thought that many of the interactions discovered by Yuh et al. (1994) have no regulatory function, but we regard this as a most unsafe *a priori* assumption. Thus, in a recent study of the *CyIIIa* gene that encodes a cytoskeletal protein of the embryonic aboral ectoderm, we found that a function can indeed be attributed to every specific DNA-protein interaction that was investigated (Kirchhamer and Davidson, 1996). The *CyIIIa* cis-regulatory domain similarly has a modular functional organization, in which each module consists of multiple sites at which occur interactions with diverse transcription factors.

For most of the experiments in this paper, we utilized as a test platform for *in vivo* function expression constructs that included the natural *Endo16* 'basal promoter' (Bp) linked to the CAT reporter gene. The Bp DNA fragment extends out to -117, and probably includes sites for more interactions than are minimally required for function as a promoter element in the strict sense of the term. This follows from the experiments of Fig. 3, in which the activity of Bp is shown to be about 10× greater than that of the SV40 early region promoter (SVp) when combined singly with the A, B or G modules of the *Endo16* regulatory domain. Perhaps the necessary SVp sequences are simply not recognized effectively by the sea urchin transcription apparatus. But this seems unlikely, since the *GBA-SVp*•CAT construct is just as active as is *GBA-Bp*•CAT (Fig. 3). Similarly, Makabe et al. (1995) found SVp to promote transcription as actively as do control constructs when linked to regulatory elements of the *SM50* and *CyIIIa* genes. The greater independence of the *Endo16* Bp element may be due instead to the three sites that it contains for the factor represented by the green ovals in Fig. 1A (Yuh et al., 1994); in any case *Bp*•CAT by itself expresses only at a very low level (Fig. 4), and the expression is not detectable at all by WMISH in any endodermal territory (Table 1C).

The functions that we identified for

modules A-G by means of the *Bp*•CAT constructs shown in Fig. 1B can be summarized as follows. The most proximal module, A, has two roles. This module causes transcription to occur in vegetal plate and archenteron, and its activity accounts quantitatively for most of the expression up to the late gastrula stage. It is also required to mediate the negative spatial control functions of modules E, F and DC. Modules E and F individually repress ectopic expression in the ectoderm surrounding the vegetal plate, and module DC represses ectopic expression in the skeletogenic mesenchyme, but none of these repressor

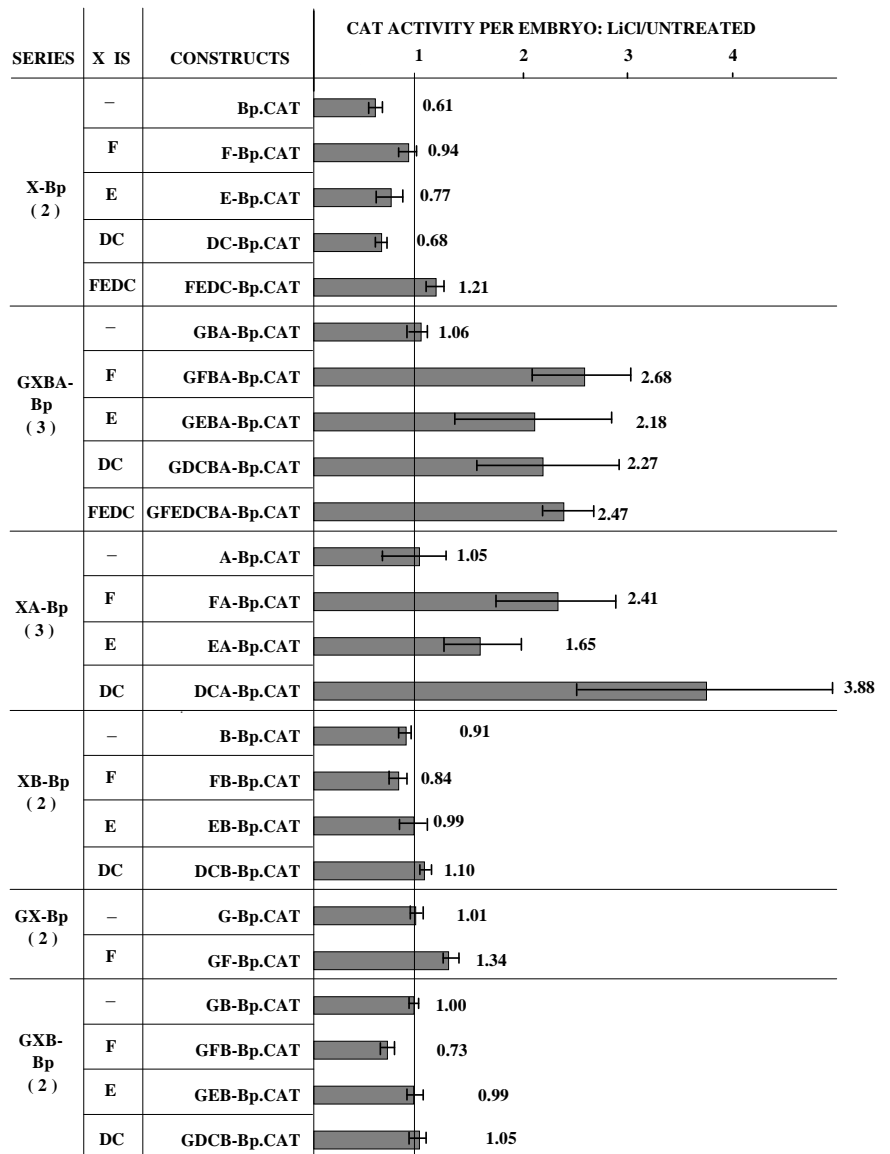
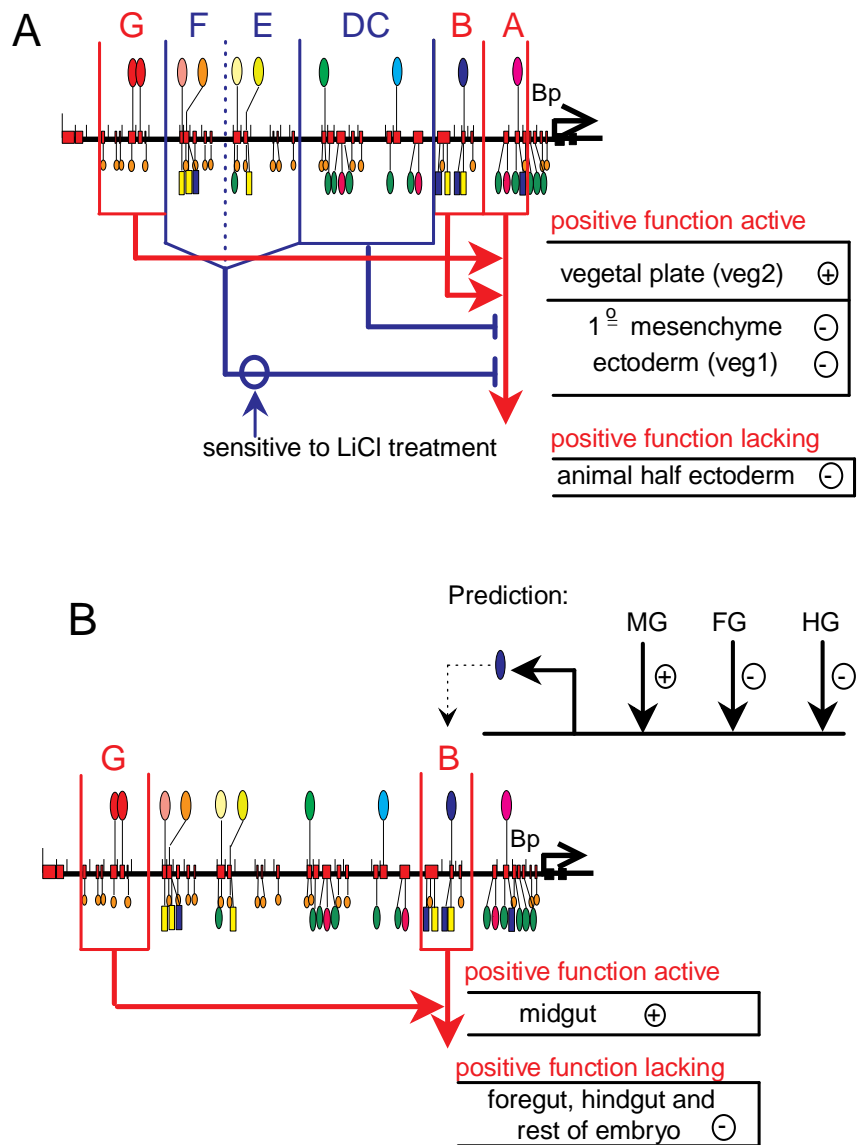


Fig. 5. Effect of LiCl on production of CAT enzyme by various *Endo16* constructs. Six series of experiments are shown, each carried out on a single batch of eggs. Data are reported from experiments for each construct series, as indicated in parentheses in the first column; standard errors are indicated. 100 embryos were utilized for each CAT assay sample (see Materials and Methods). Nomenclature is as in Fig. 4. The histogram shows the *ratio* of CAT enzyme production by normal embryos bearing each construct to CAT enzyme production by embryos treated with LiCl; i.e., where there was no difference between embryos treated with LiCl and control embryos, this ratio is 1.0. LiCl was added at the two-cell stage at a concentration of 15 mM and CAT assays were done at 48 hours of development.

modules operate in the absence of module A. Negative control of ectopic territorial expression has now been observed in several sea urchin genes expressed early in development, e.g.,

the *Cy13a* gene (Kirchhamer and Davidson, 1996, and earlier studies referred to therein) and the *Spec2A* gene (Gan et al., 1990). Later in development, the relative importance of the

Fig. 6. Function and interrelationships of modular elements within the *Endo16* cis-regulatory domain. The cis-regulatory domain and the DNA-binding proteins indicated at the positions of the mapped target sites (Yuh et al., 1994) are reproduced from Fig. 1A. There is no implication intended that all the protein factors shown are bound at once in any particular cell type(s), or at any particular time in development, though they are all present in nuclear extracts of 24 hour embryos. (A) The early specification system. Modular elements (G-A) are as defined in text. Modules the function of which is positive are indicated in red; modules that function as repressors are indicated in blue. The black brackets at right show the spatial territories of the embryo in which the indicated interactions are required, and their result: gene on (+); gene off (-). This diagram refers to the period beginning with vegetal plate specification, which is completed by the end of cleavage and continues through the blastula and gastrula stages up through complete invagination of the archenteron. During this period, the *Endo16* gene is expressed in all the cells of the vegetal plate, then throughout the archenteron, but in no other territories of the embryo. The sources of evidence leading to the individual functions indicated beneath the schematic gene are as follows: The major determinant of early expression (vertical red arrow) is module A, which alone suffices to produce the blastula and gastrula spatial patterns of endoderm expression (Fig. 2), and which accounts quantitatively for the early expression levels (Fig. 3A-C). Module B contributes to overall activity by synergizing with module A (red bent arrow) as shown in Fig. 3D, and in unpublished experiments similar to those of Fig. 3D except that the constructs were based on the Bp rather than the SVp basal promoter elements. Module G also synergizes with module A (red bent arrow); the sources of this evidence are the same as for module B. Without module A neither B nor G nor BG in combination provide normal levels of early expression; but, on a per-cell basis, the expression of the GBA combination is close to that of the complete regulatory system (Fig. 3B), indicating that these are the only modules that possess significant positive function (see also Table 1A,B). Module DC eliminates ectopic expression such as shown in Fig. 2 in the skeletogenic mesenchyme (Table 2), and therefore must repress transcription in cells in this domain of the embryo (blue lines); similarly, Table 2 shows that module E and module F are both able to eliminate ectopic expression in adjacent ectoderm (Fig. 2). That these modules require module A but not modules B or G for their repressive function is shown in Fig. 4. LiCl treatment causes *Endo16* expression to spread to ectoderm (Ransick et al., 1993) (probably the *veg1* ectodermal derivatives, both oral and aboral). Fig. 5 shows that LiCl treatment causes a several-fold enhanced expression of *Endo16* constructs. Module E or module F can each provide enhancement of expression on LiCl treatment (blue circle), but none of the positively acting modules G, B or A themselves respond to LiCl. Fig. 5 also shows that the derepression effect of LiCl requires the presence of module A together with E or F, and this is a unique function of module A, not shared by G or B. This result adds further evidence for the termination of the blue lines emanating from modules F and E at the red arrow representing module A function. Module DC is also stimulated by LiCl treatment (Fig. 5), but this result is not included in the figure since *Endo16* expression is not expanded into the skeletogenic mesenchyme territory by LiCl treatment (Ransick et al., 1993). (B) The postgastrular midgut expression system. Since omission of none of the modules causes ectopic foregut or hindgut expression, no negative spatial regulatory functions required for extinction of foregut or hindgut expression are exercised by *Endo16* cis-regulatory elements. Module B is responsible for the major quantitative expression after 60 hours (Fig. 3A,B) and alone suffices to generate midgut expression in pluteus stage embryos (Fig. 2). Module G continues to act synergistically with module B in late development (Fig. 3D; unpublished evidence). These relations are shown by the red arrows. To explain the extinction of expression in foregut and hindgut, we predict (upper part of diagram) that negative interactions preventing expression in these domains of the archenteron will be found in the cis-regulatory system of the gene encoding the key positive regulator of module B. This is probably Factor I (Yuh et al., 1994), here indicated by a blue oval. FG, foregut and secondary mesenchyme; HG, hindgut; MG, midgut.



positive activity of module A declines. At late gastrula and pluteus stages, when expression of the *Endo16* gene is confined to the midgut, its activity depends mainly on functions of module B, which then drives expression to a new and higher per-cell level (Fig. 3B). The most distal module, G, acts synergistically with both modules B and A when linked to them, strongly increasing their levels of expression, as can be seen even in constructs utilizing SVp as a basal promoter (Fig. 3D). Module B also acts synergistically with module A at the period when the latter is dominant, earlier in development.

The specific roles of the *Endo16* cis-regulatory modules, and the interrelated functions by which they establish and maintain the expression of the *Endo16* gene in the vegetal plate territory and the archenteron are shown in diagrammatic form in Fig. 6A. All six modules are utilized in this developmental regulatory process. Interpretation and processing of the developmental information states which define the cells of the vegetal plate, and the adjacent cells of the ectodermal and skeletogenic regions, evidently requires a surprisingly large, 'hard wired' cis-regulatory array of specific target sites. This work illustrates clearly the modular functional organization of the array, which reflects the biological complexity of the conditional specification processes by which the vegetal plate is initially established.

Conditional specification of the vegetal plate

The micromere transplantation and deletion experiments of Ransick and Davidson (1993, 1995) showed that vegetal plate specification depends at least in part on signals from these four cells (or their immediate progeny), as reviewed above. The early expression of *Endo16* depends mainly on module A, and we think therefore that a terminus of the responsible signal transduction pathway in the vegetal plate cells must lie within module A. There is only one factor that binds *exclusively* within module A (see Fig. 1A), and this may be the most likely candidate.

Table 2 shows that, when combined in the construct *GBA-Bp•CAT*, the positively acting *Endo16* modules cause ectopic expression in both skeletogenic mesenchyme and the surrounding ectoderm derived from the *veg1* blastomeres (see examples in Fig. 2). G, B and A modules all produce ectopic ectoderm expression when tested individually (Table 1), and at least module B by itself produces ectopic expression in skeletogenic cells. Therefore, by the time the *Endo16* gene is expressed, the positive regulators that service these modules are active in the whole lower half of the early embryo, i.e., in the skeletogenic mesenchyme precursors, the vegetal plate and the *veg1* ectoderm territories (Fig. 6; we here ignore the small micromeres). This is why the F, E and DC repressor functions are required. As we have seen, a role of module DC is to repress expression specifically in the skeletogenic territory, and this function may depend on negatively acting factors that are present from the beginning in the skeletogenic precursors, as a consequence of their own autonomous process of specification (Davidson, 1989). However, it is also possible that, once the vegetal plate territory is set up in response to signals from the micromeres, signaling from the vegetal plate cells back across the boundary ensues. The factors servicing the DC module in the skeletogenic mesenchyme could thus also lie at the terminus of a conditional interaction process. Negative signaling in both directions across an interterritorial cellular boundary may be a general feature in development as, e.g., in the particularly well

known case of the parasegmental boundaries of *Drosophila* embryos (Ingham and Martinez-Arias, 1992).

The upper boundary of the vegetal plate, i.e., that separating these cells from the *veg1*-derived ectoderm is likely to be maintained by signaling, since the distinction between these territories is clearly plastic. While no known treatment will convert skeletogenic cells into vegetal plate cells, by early exposure to LiCl, it is easy to convert all *veg1* progeny to a vegetal plate destiny, whereupon they contribute only to the archenteron (normal or exogastrulated). The converted cells also express *Endo16* (Ransick et al., 1993). The effects of LiCl are thought to be due to its interference with intracellular signal transduction systems (Berridge et al., 1989; Livingston and Wilt, 1989). We suppose that, in normal embryos, the progeny of the *veg1* tier require extracellular signals, a consequence of which is to activate the repressors that interact in the E and F modules of the *Endo16* gene, so as to shut down expression of this gene and probably other vegetal-plate-specific genes in these cells (Table 2; Fig. 4). The effect of LiCl would be to interfere with this signal transduction process, releasing the cells of the boundary region to participate in vegetal plate specification. The signals could emanate either from overlying ectoderm cells of the animal pole ectoderm lineage, or from the vegetal plate cells themselves as indicated in Fig. 6. In either case, we would expect that constructs *A-Bp•CAT* and *GBA-Bp•CAT* would be insensitive to LiCl treatment since they lack modules E or F, which are required to prevent ectopic expression in the *veg1* progeny. Fig. 5 shows that the ratio of expression with and without LiCl treatment is indeed close to one for both of these constructs. However, were the *only* effect of LiCl to block the repression of ectopic expression otherwise caused by the activity of the positive regulators in the ectoderm, then the level of expression expected in LiCl treated embryos would be that of the *GBA-Bp•CAT* construct. As expected, Fig. 4 shows that, in untreated embryos, this level is higher than that of the normally repressed constructs bearing E, F or DC modules, or combinations thereof. But paradoxically, LiCl treatment in fact *increases* activity several-fold beyond the level attained by *GBA-Bp•CAT*, and response to LiCl requires the same modules, F, E or DC, plus module A, as are required for responses of (ectopic) expression in untreated embryos. This cannot be a coincidence: both functions must depend on the intracellular signal transduction pathways that LiCl affects. Consistent with this argument is a recent observation that introduction of a gene encoding a dominantly active *ras* protein together with *Endo16* constructs, produces exactly the same effect as does LiCl treatment, and requires exactly the same regulatory modules (unpublished data). We infer that the spatial repressor modules of the *Endo16* gene, viz, E, F and DC, are all likely to lie at the termini of signal transduction pathways. It may be relevant that partial sequence data on one of the factors binding to a unique target site in module F indicates homology to the CREB/ATF family of transcription factors, which are known in other systems to mediate responses to extracellular signals (Yamamoto et al., 1988; de Groot and Sassone-Corsi, 1993; Lalli and Sassone-Corsi, 1994).

Endo16 regulation during regionalization of the archenteron

Remarkably, none of the partial constructs that we tested produced ectopic expression in either the foregut or the hindgut

late in development. If they were expressed at all in the endoderm at this stage, they were expressed only in the midgut. As illustrated in Fig. 2, even a construct as simple as *B-Bp•CAT* is expressed accurately in the midgut late in development, though it lacks any of the repressive modules E, F or DC required to prevent ectopic expression in the adjacent embryonic territories. It follows that the *cis*-regulatory domain of the *Endo16* gene does not include the spatial control elements that are responsible for extinguishing expression in foregut, the secondary mesenchyme derivatives and hindgut. Instead, in the final phase of development, spatial control of *Endo16* expression must depend on the regulatory interactions that control the regions of the archenteron where the positively acting transcription factors that bind in the B module are presented; that is, the responsible interactions would have to be sought in the *cis*-regulatory domains of the genes encoding these transcription factors, as indicated in Fig. 6B.

In this light, we see that the major part of the functional complexity of the *Endo16 cis*-regulatory system is required for interpretation of the spatial information presented in early development. Once the embryo has been initially organized the need for information processing at the gene and for an elaborate regulatory apparatus to carry out this task, gives way to a much simpler functional requirement. During regionalization and differentiation of the gut, control is shifted to a higher level of genetic regulatory hierarchy and, thereafter, the *Endo16* gene responds simply to the availability of one or a few cell-type-specific, positively acting transcription factors.

This work was supported by a grant from the Institutes of Child Health and Human Development of the NIH (HD-05753). We are grateful to Drs Andrew Ransick, Andrew Cameron and Carman Kirchhamer of this laboratory, and to Professor Ellen Rothenberg, for their extremely perspicacious critical reviews of drafts of this manuscript.

REFERENCES

- Berridge, M. J., Downes, C. P. and Hanley, M. R. (1989). Neural and developmental actions of lithium: A unifying hypothesis. *Cell* **59**, 411-419.
- 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.
- Burke, R. D. (1980). Morphogenesis of the digestive tract of the pluteus larva of *Strongylocentrotus purpuratus*: Shaping and bending. *Int. J. Invert. Reprod.* **2**, 13-21.
- 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.
- 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.
- 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.
- Davidson, E. H. (1991). Spatial mechanisms of gene regulation in metazoan embryos. *Development* **113**, 1-26.
- de Groot, R. P. and Sassone-Corsi, P. (1993). Hormonal control of gene expression: Multiplicity and versatility of cAMP-responsive nuclear regulators. *Mol. Endocrinol.* **7**, 145-153.
- 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.
- 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.
- 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.
- Hörstadius, S. (1973). *Experimental Embryology of Echinoderms*. Oxford: Clarendon Press.
- Hough-Evans, B. R., Britten, R. J. and Davidson, E. H. (1988). Mosaic incorporation and regulated expression of an exogenous gene in the sea urchin embryo. *Dev. Biol.* **129**, 198-208.
- Ingham, P. W. and Martínez-Arias, A. (1992). Boundaries and fields in early embryos. *Cell* **68**, 221-235.
- Khoury, G. and Gruss, P. (1983). Enhancer elements. *Cell* **33**, 313-314.
- Kirchhamer, C. V. and Davidson, E. H. (1996). Spatial and temporal information processing in the sea urchin embryo: Modular and intramodular organization of the *Cy13a* gene *cis*-regulatory system. *Development* (in press).
- Lalli, E. and Sassone-Corsi, P. (1994). Signal transduction and gene regulation: The nuclear response to cAMP. *J. Biol. Chem.* **269**, 17359-17362.
- Leahy, P. S. (1986). Laboratory culture of *Strongylocentrotus purpuratus* adults, embryos, and larvae. *Meth. Cell Biol.* **27**, 1-13.
- 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.
- Makabe, K. W., Kirchhamer, C. V., Britten, R. J. and Davidson, E. H. (1995). *Cis*-regulatory control of the *SM50* gene, an early marker of skeletogenic lineage specification in the sea urchin embryo. *Development* **121**, 1957-1970.
- McMahon, A. P., Novak, T. J., Britten, R. J. and Davidson, E. H. (1984). Inducible expression of a cloned heat shock fusion gene in sea urchin embryos. *Proc. Natl. Acad. Sci. USA* **81**, 7490-7494.
- 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.* **108**, 420-430.
- Nocente-McGrath, C., Brenner, C. A. and Ernst, S. G. (1989). *Endo16*, a lineage-specific protein of the sea urchin embryo, is first expressed just prior to gastrulation. *Dev. Biol.* **136**, 264-272.
- Nocente-McGrath, C., McIsaac, R. and Ernst, S. G. (1991). Altered cell fate in LiCl-treated sea urchin embryos. *Dev. Biol.* **147**, 445-450.
- Ransick, A. and Davidson, E. H. (1995). Micromeres are required for normal vegetal plate specification in sea urchin embryos. *Development* **121**, 3215-3222.
- 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.
- Soltysik-Espanola, M., Klinzing, D. C., Pfarr, K., Burke, R. D. and Ernst, S. G. (1994). *Endo16*, a large multidomain protein found on the surface and ECM of endodermal cells during sea urchin gastrulation, binds calcium. *Dev. Biol.* **165**, 73-85.
- Yamamoto, K. K., Gonzalez, G. A., Biggs III, W. H. and Montminy, M. R. (1988). Phosphorylation-induced binding and transcriptional efficacy of nuclear factor CREB. *Nature* **334**, 494-498.
- Yuh, C.-H., Ransick, A., Martinez, P., Britten, R. J. and Davidson, E. H. (1994). Complexity and organization of DNA-protein interactions in the 5' regulatory region of an endoderm-specific marker gene in the sea urchin embryo. *Mech. Dev.* **47**, 165-186.
- Zeller, R. W., Coffman, J. A., Harrington, M. G., Britten, R. J. and Davidson, E. H. (1995a). SpGCF1, a sea urchin embryo transcription factor, exists as five nested variants encoded by a single mRNA. *Dev. Biol.* **169**, 713-727.
- Zeller, R. W., Griffith, J. D., Moore, J. G., Kirchhamer, C. V., Britten, R. J. and Davidson, E. H. (1995b). A multimerizing transcription factor of sea urchin embryos capable of looping DNA. *Proc. Natl. Acad. Sci. USA* **92**, 2989-2993.