**Cis-regulation downstream of cell type specification: a single compact element controls the complex expression of the CyIIa gene in sea urchin embryos**

Maria I. Arnone¹, Ellen L. Martin² and Eric H. Davidson²,*

¹Stowers Institute for Medical Research, 4949 Rockhill Road, Kansas City, Missouri 64110, USA
²Division of Biology, California Institute of Technology, Pasadena, CA 91125, USA

*Author for correspondence (e-mail: davidson@mirsky.caltech.edu)

Accepted 30 January; published on WWW 18 March 1998

**SUMMARY**

CyIIa, a cytoskeletal actin gene of *Strongylocentrotus purpuratus*, is expressed specifically though transiently in the embryonic skeletogenic and secondary mesenchyme and, later in development, is permanently activated in the hindgut and midgut. CyIIa transcription follows, and is therefore downstream of, the initial specification of these embryonic domains. A detailed functional analysis of the cis-regulatory system governing the rate and the location of CyIIa expression during development was carried out using GFP expression constructs. About 4.4 kb of CyIIa sequence including a leader intron were examined for cis-regulatory function. Distal elements scattered over several kb account for 60% of the quantitative output of the expression construct and a strong amplifier of expression is located within the leader intron. However, the complex spatial pattern of CyIIa expression is completely reproduced by a compact upstream regulatory element <450 bp in length. We found no evidence anywhere in the 4.4 kb sequence examined for negative regulators required to repress ectopic expression. The specific site that mediates CyIIa expression in the midgut in late embryos and larvae was identified. This site is the same as that necessary and sufficient for midgut expression of the *Endo16* gene late in development, and was shown to bind the same transcription factor. Except for some temporal and quantitative features, the *S. purpuratus* expression construct is expressed accurately and specifically in the same diverse cell types when introduced into embryos of *Lytechinus pictus*, which belongs to a different echinoid order. No ectopic expression was observed, in contrast to the result of a similar interspecific gene transfer experiment carried out earlier on a different cytoskeletal actin gene that is expressed much earlier in development. Presentation of the set of transcription factors that activate CyIIa in the differentiated cells in which it is expressed is apparently a conserved feature of these cell types.

Key words: Sea urchin, *Strongylocentrotus purpuratus*, CyIIa, Green Fluorescent Protein (GFP), Actin, Gene regulation, Regulation

**INTRODUCTION**

This work was undertaken to explore the relation between cis-regulatory organization and mode of developmental gene expression. The CyIIa cytoskeletal actin gene of *S. purpuratus* is transcribed during development in a sequence of diverse differentiating cell types, belonging to several different spatial territories. CyIIa is expressed only following embryonic specification of these territories (Lee et al., 1986). Our objective was to determine the characteristics of the cis-regulatory system that generates the unusually complex and dynamic mode of expression of this gene.

A contrast in developmental mode of gene expression is afforded by the CyIIIa cytoskeletal actin gene, which except for a region of nine amino acids encodes a protein almost identical to that encoded by the CyIIa gene (Shott et al., 1984). Unlike CyIIa, CyIIIa is activated during cleavage in the course of the initial specification events that distinguish oral from aboral ectoderm. Thereafter it is expressed exclusively in all aboral ectoderm lineages. The 2300 bp cis-regulatory system governing CyIIIa expression has been extensively characterized and provides an illuminating standard of comparison (Calzone et al., 1988; Thézé et al., 1990; Hough-Evans et al., 1990; Franks et al., 1990; Kirchhamer and Davidson 1996; Coffman et al., 1997). The early expression pattern of the CyIIIa gene is mediated by a proximal cis-regulatory module, and its expression during later development by a more distal cis-regulatory module, each of which contains essential target sites for both positively and negatively acting regulators, as well as for other ancillary factors. The CyIIIa activators are functional over embryonic spatial domains that extend beyond the aboral ectoderm, and correct spatial expression depends crucially on repressive functions that confine transcription to the aboral ectoderm territory.
As shown in earlier studies (Cox et al., 1986; Miller et al., 1996), and described in further detail below, CyIIa transcripts appear in succession in skeletogenic mesenchyme, in some delaminating secondary mesenchyme cells and their precursors at the tip of the invaginating archenteron, and finally in mid gut and hind gut. We demonstrate that all of these aspects of CyIIa expression are mediated by a single compact cis-regulatory module a few hundred bp in length. Unlike CyIIb or most other well-analyzed embryonic cis-regulatory systems that operate during spatial specification of embryonic territories (see Arnone and Davidson, 1997 for review), the CyIIa gene appears to depend for spatial control only on positively acting inputs; we have as yet found no evidence for negative spatial control functions. We conclude that the structural organization of the CyIIa cis-regulatory system reflects its position in the developmental gene regulatory network.

MATERIALS AND METHODS

Embryo culture, microinjection, whole-mount in situ hybridization and microscopy

S. purpuratus embryos were cultured and DNA injected into fertilized eggs as described (McMahon et al., 1985). Prior to injection, plasmids were linearized at unique restriction sites located either downstream of reporter gene and SV40 poly(A) sequences or upstream of cis-regulatory elements. After injection, embryos were collected at the appropriate stage and processed for chloramphenicol acetyltransferase (CAT) measurements as described (Yuh and Davidson, 1996) or for whole-mount in situ hybridization (WMISH). For epifluorescence observations, embryos were imaged as described by Arnone et al. (1997). WMISH was performed following the procedures described by Ransick et al. (1993) and modified by Kirchhammer and Davidson (1996), using 0.02 μg/ml CyIIa (Lee et al., 1986) and 0.005 μg/ml Spec1 (Lynn et al., 1983) probe concentrations, respectively.

Reporter gene constructs

The 8.8 kb CyIIa-CAT.1 plasmid described by Zeller et al. (1992) (here designated EDCBA-CAT) was the source of CyIIa regulatory sequence for all the expression constructs used in this study. The EDCBA-GFP fusion construct was created by inserting the 2.7 kb Apol fragment of EDCBA-CAT into the CyIIa-GFP plasmid used in our earlier study (Arnone et al., 1997). A partial restriction map of this plasmid is shown in Fig. 3. Both EDCBA-CAT and -GFP plasmids contain 4.4 kb of upstream CyIIa sequence. The transcription initiation site is followed by a leader sequence which, in the primary transcript, includes a 560 bp intron. After excision of the intron, the leader is 84 bp in length. The CAT- or GFP-coding sequences were fused in-frame with the CyIIa coding sequence 39 bp downstream of the translation initiation site. Genomic CyIIa sequence upstream of the translation initiation site and including regions A and B (Fig. 3) was obtained by K. Katula (GenBank Accession no. AF032880), to whom we are grateful for making it available to us in advance of publication. The remainder of the upstream sequence (regions E–C, Fig. 3) was obtained by us (GenBank Accession no. AF034254). For the sequence of the coding region of the CyIIa gene and the 3′ trailer, see Durica et al. (1988).

The CAT constructs numbered 1 to 8 and 10 to 13 in Fig. 3 were generated from the plasmid EDCBA-CAT either by deletion or fusion of restriction fragments. GFP constructs 7, 8 and 10 were similarly generated from the plasmid EDCBA-GFP. To create plasmid B-GFP (construct 9 of Fig. 3) the PstI fragment of EDCBA-CAT, which contains the whole B element, was isolated and cloned into the PstI site of pGL-4, a derivative of CyIIa-GL (Arnone et al., 1997), which bears only 6 bp upstream of the CyIIa ATG followed by 39 bp of coding sequence. Plasmid mEnBA-GFP and mEnB2B1A-GFP (construct numbers 14 and 15 of Fig. 3) were obtained by point mutation of the BA-GFP and B2B1A-GFP constructs, respectively. Double-strand plasmids were synthesized from each template in a thermal cycler by Pfu DNA polymerase using the following oligonucleotide primers:

MUIU: 5′-AACAGGTCTAGTCGgcaCATATAACTAAAC-TCTGTAATGCC-3′

MULU: 5′-CGAGTTTGAATAGAAAATCTTATGTCGgcaCATGAC-CTGTTTTATCC-3′

in which 6 bp were substituted in the form of the NruI restriction site (sequence in lower case letters). Digestion of parental DNA with DpnI and subsequent plasmid manipulations were performed using the Quick Change Site-Directed Mutagenesis Kit (Stratagene) following manufacturers’ instructions. Both plasmids were checked by restriction at the unique NruI site generated by mutagenesis.

The construct termed B3B2bEp-GFP (Fig. 3, no. 16) was cloned by digesting B-GFP with BglII and HindIII, which removes the 181 bp B element containing the CyIIa transcription start site. The Endo16 basal promoter element was prepared by PCR of the Endo16 cis-regulatory system (Yuh et al., 1994) using primers CYEPU 5′-AAAAGTCTATACCCGGAATAGACgccgTataAACTTTT-GAGTTTCGCTC-3′ and ENBPL 5′-CCGGAAGCTTAAAATATAGCAATATG-3′. These primers amplify only the Endo16 basal promoter region (as defined by Yuh and Davidson, 1996), i.e., from −117 to +20. The CYEPU primer includes the first 20 bp of B1 (here designated as b), so that the amplified fragment also contains this element 5′ of the Endo16 promoter. The PCR fragment was flanked with BglII and HindIII sites and thus cloned into the construct in place of the original B1 fragment.

RESULTS

Expression of the endogenous CyIIa gene during development

The CyIIa cytoskeletal actin gene displays the most complex pattern of expression of any of the five S. purpuratus cytoskeletal actin genes. The principal domains of CyIIa expression were established initially by Cox et al. (1986), on the basis of radioactive in situ hybridization on sectioned material and additional observations obtained by whole-mount in situ hybridizations (WMISH) were published recently by Miller et al. (1996). These studies showed that the spatial expression pattern changes dramatically during development. It was clear that in order to be able to interpret the output of CyIIa reporter genes accurately, detailed information on expression at each stage would be required. Fig. 1 displays the location of CyIIa transcripts in S. purpuratus embryos at closely spaced time intervals, from late blastula to pluteus stages, as revealed by WMISH.

CyIIa transcription is activated toward the end of the blastula stage. As determined by probe excess mRNA titration, transcripts of this gene can first be observed at about 20 hours postfertilization (pf), when about 10³ molecules per embryo were recorded (Lee et al., 1986). At 24 hours, the transcription rate is still very low, not exceeding five molecules per embryo-minute (Lee et al., 1992). In line with these observations, only very faint WMISH staining can be observed in the 24 hour mesenchyme blastula shown in Fig. 1A. The initial locus of CyIIa expression is apparently the ingressing skeletogenic mesenchyme cells, as also noted by Cox et al. (1986). Shortly thereafter (Fig. 1B, C, 28 and 34
Expression of the *CyIIa* gene in sea urchin embryos

Fig. 1. Expression of *CyIIa* in embryogenesis of *S. purpuratus*. (B-H) Typical whole-mount in situ hybridization results are shown for embryos viewed frontally: (A) the aspect of the embryo is indeterminate and (I) is a lateral view. (A) Mesenchyme blastula (24 hour). Faint though specific staining is observed in most though not all ingressing skeletogenic mesenchyme cells. (B) Very early gastrula (28 hour). *CyIIa* is expressed in some cells of the leading edge of the invaginating archenteron. (C) Early gastrula (34 hour) showing expression at the tip of the archenteron. Note the absence of staining in skeletogenic (arrowhead) and some secondary (arrow) mesenchyme cells visible in this image. (D) Later gastrula (40 hour), displaying renewed *CyIIa* expression in skeletogenic mesenchyme cells, which, by this stage, have formed the subequatorial ring (arrowheads). *CyIIa* transcripts are also accumulating in some secondary mesenchyme cells, which are migrating away from the archenteron tip (arrow). (E) Late gastrula (46 hour). *CyIIa* is now strongly expressed in skeletogenic mesenchyme cells, of which the two clusters where skeletogenesis has begun are visible on either side of the archenteron. Labeled transcripts have entirely disappeared from the tip of the archenteron, which has now attained maximal elongation. (F) Early prism (52 hour), showing *CyIIa* expression in the nascent coelomic pouches (arrows) and the blastopore (arrowhead). Staining of skeletogenic cells is weaker but can still be observed. (G) Prism-stage embryo (60 hour). *CyIIa* expression is restricted primarily to midgut and hindgut regions of the archenteron. There is residual staining in one of the two coelomic pouches. (H,I) Pluteus-stage larvae (72 hour). *CyIIa* expression is restricted to midgut and hindgut; no significant staining is observed in foregut. Two different views of the same embryo are displayed, oral (H) and lateral (I), to display absence of expression either in coelomic pouches or in any region of the ectoderm.

hours pf), staining disappears from the skeletogenic mesenchyme, as also observed by Miller et al. (1996), and a more intense domain of expression appears in the invaginating vegetal plate. Expression is maintained at the tip of the archenteron as it extends across the blastocoel throughout early to mid gastrulation. Miller et al. (1996) noticed that this staining is confined to one side of the archenteron tip and we confirm this unusual pattern of expression (Fig. 2A1, A2). To demonstrate the orientation of this asymmetry, we displayed *Spec1* as well as *CyIIa* transcripts in the same early gastrula (32 hour pf) embryos. *Spec1* expression is confined to the aboral ectoderm (Lynn et al., 1983; Tomlinson et al., 1990) and, in Fig. 2B1 and B2, the domain of *CyIIa* expression can be seen to lie directly opposite to that of *Spec1* expression. Thus it is cells on the oral side of the invaginating archenteron tip that express *CyIIa*, just as surmised by Miller et al. (1996). Based on their earlier fate map of secondary mesenchyme cell (SMC) types (Ruffins and Ettensohn, 1996), Miller et al. (1996) suggested that the cells expressing *CyIIa* are the precursors of the blastocoelar type of SMCs. It remains doubtful, but cannot be excluded from any of the three studies, whether endodermal cells ever express *CyIIa* in the blastula-early gastrula period of development.

Towards the end of gastrulation, the pattern of *CyIIa* expression again changes. By 40 hours (Fig. 1D), transcripts begin to accumulate strongly in skeletogenic mesenchyme cells. At this stage, these are arranged in a ring at the vegetal end of the embryo connecting the two clusters on the oral side where spicule formation is initiated. *CyIIa* expression at the tip of the archenteron is almost extinguished, while some migratory SMCs are now also labeled. A few hours later (Fig. 1E, 46 hours pf), *CyIIa* transcripts are seen only in skeletogenic cells. A new pattern then appears at the prism stage (52 hours). Fig. 1F shows that *CyIIa* expression now extends to the coelomic pouches (arrows) and to the circumblastopore region (arrowhead). In pluteus-stage embryos from 60 hours onward, expression is extinguished in skeletogenic mesenchymal cells.
and coelomic pouches, and the final definitive endodermal pattern of expression emerges. Both midgut and hindgut now strongly express the gene (Fig. 1G-I), and continue to do so far into larval development (Arnone et al., 1997). Cox et al. (1986) and Miller et al. (1996) both report in addition that scattered blastocoelear SMCs continue to express the CyIIa gene in late pluteus-stage embryos. However, neither Cox et al. (1986) nor we in this work could confirm Miller et al. (1996) that pigment cells express CyIIa at the pluteus stage.

An overview of the complex pattern of CyIIa expression is shown in Table 1. Three conclusions can be drawn. (1) The CyIIa gene is utilized in cells of entirely separate embryonic lineage and different type, viz in the skeletogenic or primary mesenchyme, in the blastocoelear SMCs which delaminate from the archenteron, in the coelomic pouches and finally in the midgut and hindgut. The gene is not, however, expressed in any ectodermal territory at any time, nor in the foregut. (2) CyIIa transcripts do not persist in any mesenchymal cells (or their precursors) for more than 12 hours, as indicated by the shaded blocks in Table 1. (3) CyIIa activation follows cell initial specification. The earliest expression in ingressing skeletogenic mesenchyme long follows the autonomous specification of these cells (Davidson, 1989), and the early expression in SMC lineages follows their segregation from endoderm lineages, which occurs before invagination (Ruffins and Ettensohn, 1996). This point is particularly clear with reference to the two major later domains of CyIIa expression: in the skeletogenic mesenchyme, CyIIa transcripts accumulate only after 40 hours, a period when these cells are already executing their differentiated function of laying down the spicule rods and, in the midgut and hindgut, where CyIIa transcripts accumulate only during the terminal phases of differentiation, after 60 hours pf. Since CyIIa encodes a cytoskeletal actin, its activity in skeletogenic and secondary mesenchyme may reflect the migratory activity of these cells (Cox et al., 1986) or their projection of filapodial extensions (Malinda et al., 1995; Miller et al., 1995). No doubt some of the cis-regulatory elements that we describe in the following respond to physiological cues affecting the intensity of these activities. The developmental components of the regulatory system may be considered those that cause the CyIIa gene to be expressed in the specific cell types indicated in Table 1, as embryogenesis proceeds.

**Table 1. CyIIa mRNA distribution in different regions of the* S. purpuratus *embryo during development**

<table>
<thead>
<tr>
<th>Stage</th>
<th>hpf</th>
<th>PMC</th>
<th>Archenteron tip</th>
<th>Blastocoelear SMC or precursors</th>
<th>Coelomic pouches</th>
<th>Mid + hindgut</th>
</tr>
</thead>
<tbody>
<tr>
<td>Blastula</td>
<td>24</td>
<td>+</td>
<td>++</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Gastrula</td>
<td>28</td>
<td>-</td>
<td>++</td>
<td>+</td>
<td>++</td>
<td>++</td>
</tr>
<tr>
<td></td>
<td>34</td>
<td>++</td>
<td>++</td>
<td>+</td>
<td>++</td>
<td>++</td>
</tr>
<tr>
<td>Prism</td>
<td>46</td>
<td>++</td>
<td>-</td>
<td>-</td>
<td>++</td>
<td>-</td>
</tr>
<tr>
<td>Pluteus</td>
<td>52</td>
<td>++</td>
<td>-</td>
<td>-</td>
<td>++</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>60</td>
<td>++</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>++</td>
</tr>
<tr>
<td></td>
<td>72</td>
<td>++</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>++</td>
</tr>
</tbody>
</table>

*The pattern of CyIIa RNA localization is reported at different stages, from Fig. 1 and earlier studies (Cox et al., 1984; Miller et al., 1986). Presence of expression in a given region is reported with a single or double cross, according to the relative intensity. The relative distance between entries is proportional to developmental time. Shaded regions represent continuity of expression in a given cell type or region of the embryo.

Abbreviations: archent., archenteron; hpf, hours post fertilization; PMC, primary mesenchyme cells; SMC, secondary mesenchyme cells.

**Spatial and temporal expression mediated by a 4.4 kb CyIIa fusion construct**

The cis-regulatory analysis described in this paper was carried out using two different reporter systems. GFP reporters were employed for most of the spatial analyses, as they are specially advantageous for this gene due to the transience of its expression phases (Arnone et al., 1997). CAT reporters were utilized mainly for quantitative and temporal analyses, though our initial WMISH experiments were carried out with a CAT construct containing 4.4 kb of genomic CyIIa sequence. The spatial expression of this fusion gene had been preliminarily characterized by Zeller et al. (1992). This construct, here referred to as EDCBA-CAT, and the GFP construct derived from it, EDCBA-GFP, are considered to include the ‘complete’ CyIIa cis-regulatory system. As we now demonstrate, the EDCBA constructs indeed reproduce essentially all the main spatial and temporal features of endogenous CyIIa gene expression.

A map of the EDCBA-CAT and EDCBA-GFP constructs is shown diagrammatically at the top of Fig. 3. These constructs begin about 3.7 kb upstream of the CyIIa transcription initiation site, which is followed by a 54 or 59 bp leader exon (there are two closely spaced initiation sites), a 560 bp intron, 25 bp of remaining leader sequence and the beginning of the CyIIa coding sequence. The CAT and GFP reporters were fused in-frame at a site 37 bp downstream of the CyIIa translation initiation codon. The transcription initiation site was determined in a conventional primer extension experiment (not shown), which confirmed an assignment derived from a comparison of cDNA and genomic sequences (K. Katula, personal communication).

The quantitative temporal output of EDCBA-CAT was measured by the levels of CAT enzyme present in embryos grown from eggs injected with this plasmid and then collected at different developmental times. Because of the relatively short half-lives of CAT mRNA and protein relative to the developmental time scale, CAT enzyme content is approximately proportional to the transcription rate of the expression construct (Yuh et al., 1996). As shown in Fig. 4A (solid symbols), the fusion gene is barely active at mesenchyme blastula stage but is expressed at an increasing rate from this stage onward, until the levels of CAT enzyme activity attain a plateau at pluteus stage. This temporal profile reconstructs that described by the steady-state levels of the endogenous CyIIa.
transcripts during development, as measured by Lee et al. (1986) and summarized by the bars in Fig. 4A. The two data sets are not directly comparable, in that late in development the half-life of endogenous CyIIa mRNA is only about 30 minutes (Lee et al., 1992) when the gene is expressed mainly in gut cells (Fig. 1), while that of CAT protein in these same cells is several hours (Armone et al., 1997). Nonetheless, it can be seen that expression of the construct is initiated at the appropriate time and that its rate of expression then increases, just as does that of the endogenous gene. Based on the accumulation of CAT mRNA in late development shown in Fig. 4A and the rate of translation (Davidson, 1986), the embryo contains \( \geq 1.3 \times 10^4 \) molecules of CAT mRNA; from the CAT mRNA \( t_1 \) measured by Armone et al. (1997) in endoderm and the average clone size of 6.5 expressing cells, the rate of transcription would be 4.2 molecules CAT mRNA•cell•min\(^{-1}\). This may be compared to 0.21 molecules•cell•min\(^{-1}\) for the endogenous CyIIa gene of this developmental period (Lee et al., 1992). The 20-fold difference directly reflects the multiple copies of incorporated exogenous DNA (McMahon et al., 1985) and indicates that the constructs are expressing at reasonable rates, perhaps on a per active gene basis not too different from the endogenous CyIIa genes.

Spatial expression mediated by the EDCBA expression constructs is illustrated in Fig. 5. At mesenchyme blastula stage (Fig. 5, A1, B1) expression is observed in skeletogenic mesenchyme cells and in vegetal regions where some skeletogenic mesenchyme cells are still in process of ingression. This pattern is observed in almost all injected embryos that display staining (Table 2, line 1; note that injected embryos are delayed about 2 hours in development, so the ‘mesenchyme blastula’ category consists of embryos equivalent to normal embryos at 24-25 hours pf). Precursors of SMCs may also contribute to this early vegetal labeling, since SMCs may also contribute to this early vegetal labeling, since the total fraction of injected embryos displaying transgene expression is somewhat higher than observed with strictly skeletogenic mesenchyme-specific transgenes (i.e., \( \leq 50\% \); Makabe et al., 1995; Armone et al., 1997). Except for a small, though reproducibly observed, component of ectopic ectodermal expression at this stage (13.2%; Table 2), EDCBA-CAT and EDCBA-GFP behave just as does the endogenous CyIIa gene in its earliest phases of activity, as summarized in Table 1. The low ectopic ectodermal expression is transient and Table 2 shows that it soon disappears. Previous studies indicate this phenomenon is likely to be an artifact which is due to exogenous DNA that has not been stably incorporated (Yuh et

### Table 2. Spatial analysis of GFP expression observed at various stages for EDCBA-, BA-, B2B2bEp-GFP and B2B1A-GFP constructs

<table>
<thead>
<tr>
<th>Construct</th>
<th>Stage*</th>
<th>Scored embryos†</th>
<th>Positive:negative‡</th>
<th>% Vegetal expression §</th>
<th>% PMC expression ¶</th>
<th>% SMC expression ¶</th>
<th>% Coelomic pouch expression ††</th>
<th>% Gut expression</th>
<th>% Ectoderm expression ‡‡</th>
</tr>
</thead>
<tbody>
<tr>
<td>EDCBA-GFP</td>
<td>B</td>
<td>100</td>
<td>83:17 (83.0)</td>
<td>66.3</td>
<td>71.1</td>
<td>13.2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>G</td>
<td>60</td>
<td>41:19 (68.3)</td>
<td>43.9</td>
<td>87.8</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Pr</td>
<td>48</td>
<td>39:9 (81.2)</td>
<td>94.9</td>
<td>7.7</td>
<td>10.2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Pl</td>
<td>71</td>
<td>59:12 (83.0)</td>
<td>64.4</td>
<td>3.4</td>
<td>10.2</td>
<td>78.8</td>
<td></td>
<td></td>
</tr>
<tr>
<td>BA-GFP</td>
<td>B</td>
<td>218</td>
<td>171:47 (78.4)</td>
<td>52.0</td>
<td>59.1</td>
<td>11.7</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>G</td>
<td>143</td>
<td>100:43 (69.9)</td>
<td>35.0</td>
<td>67.0</td>
<td>2.0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Pr</td>
<td>31</td>
<td>23:8 (74.2)</td>
<td>95.6</td>
<td>13.0</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Pl</td>
<td>170</td>
<td>111:59 (65.3)</td>
<td>60.3</td>
<td>9.0</td>
<td>18.0</td>
<td>55.0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>B-GFP</td>
<td>B</td>
<td>78</td>
<td>19:59 (24.4)</td>
<td>42.0</td>
<td>79.0</td>
<td>15.8</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>G</td>
<td>98</td>
<td>33:65 (33.7)</td>
<td>72.7</td>
<td>63.6</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Pr</td>
<td>143</td>
<td>25:118 (17.5)</td>
<td>72.0</td>
<td>24.0</td>
<td>8.0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Pl</td>
<td>87</td>
<td>22:65 (25.3)</td>
<td>50.0</td>
<td>0</td>
<td>4.5</td>
<td>50.0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>B2B1A-GFP</td>
<td>B</td>
<td>180</td>
<td>85:95 (47.5)</td>
<td>68.2</td>
<td>56.5</td>
<td>14.4</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>G</td>
<td>107</td>
<td>29:78 (27.1)</td>
<td>41.4</td>
<td>58.6</td>
<td>3.4</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Pr</td>
<td>194</td>
<td>128:66 (66.0)</td>
<td>31.2</td>
<td>3.1</td>
<td>6.2</td>
<td>78.1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>B3B2bEp-GFP</td>
<td>B</td>
<td>95</td>
<td>54:41 (56.8)</td>
<td>53.7</td>
<td>55.5</td>
<td>11.1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>G</td>
<td>85</td>
<td>42:43 (4.9)</td>
<td>76.2</td>
<td>57.1</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Pr</td>
<td>107</td>
<td>44:63 (41.1)</td>
<td>68.1</td>
<td>34.1</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Pl</td>
<td>119</td>
<td>60:59 (50.4)</td>
<td>58.3</td>
<td>8.3</td>
<td>3.3</td>
<td>41.7</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* Blastulae were collected at 26 to 30 hours, gastrulae at 46 to 50 hours, prisms at 56 to 60 hours and plutei at 72 to 76 hours postfertilization. Blanks denote regions that have not yet formed or have been supplanted by other structures.

† Values represent data from at least two separate experiments, carried out on independent batches of eggs.

‡ Embryos with more than two fluorescent cells were scored as positive.

§ % expression = (Σ positive embryos/Σ scored embryos)\times100.

¶ % territorial expression = [Σ positive embryos in a given territory/Σ positive embryos]\times100.

** Vegetal expression is probably expression in ingressing skeletogenic mesenchyme cells, rather than endodermal cells, though this has not been directly determined; see text.

†† These values are minimal, because coelomic pouches had not yet formed in some of the embryos examined, which as noted in text, are slightly delayed due to initial effects of the DNA injection.

‡‡ Ectodermal expression is probably artifactual as it is confined to single cells and disappears with advancing age of the embryos; these are characteristics of unstably incorporated DNA (see Yuh and Davidson, 1996).

Abbreviations: B, blastula; G, gastrula; PMC, primary or skeletogenic mesenchyme; Pr, prism; Pl, pluteus; SMC, secondary mesenchyme cell.
the time course of endogenous regulatory sequence that (1) reproduces gene unequivocally includes a CyIIa coelomic pouches, as expected (Fig. 1).

Expression in skeletogenic cells in late tip, (2) that it properly turns off SMCs on the oral side of the archenteron, just as for the endogenous CyIIa gene. All results for this stage were consistent with the observations in Fig. 2 showing that expression is confined to one side of the invaginating archenteron, i.e., the oral side. Almost all labeled embryos at early-mid gastrula stage (88%) display expression in some SMCs and many in skeletogenic mesenchyme as well (Fig. 5 (A2), Table 2; cf. Table 1). Skeletogenic mesenchyme labeling attains its peak values at prism stage (95% of labeled embryos; Table 2), which is at the end of the period when transcripts of the endogenous CyIIa gene can be seen in these cells. As discussed by Arnone et al. (1997), this is due to the long half-life of GFP in skeletogenic mesenchyme cells. Indeed, these retain this fluorescence even at pluteus stages, as indicated quantitatively in Table 2 and illustrated in Fig. 5 (A3, A4). The diffusion of GFP through the syncytial skeletogenic cables accounts for the complete patterns of labeling seen in the skeletogenic structures (Arnone et al., 1996), which is often observed in pregastrular-stage embryos and is then lost.

Fig. 5 (B2, B3) illustrates EDCBA-GFP expression in early-mid gastrula-stage embryos: labeling is now observed at the tip of the archenteron, just as for the endogenous CyIIa gene. All results for this stage were consistent with the observations in Fig. 2 showing that expression is confined to one side of the invaginating archenteron, i.e., the oral side. Almost all labeled embryos at early-mid gastrula stage (88%) display expression in some SMCs and many in skeletogenic mesenchyme as well (Fig. 5 (A2), Table 2; cf. Table 1). Skeletogenic mesenchyme labeling attains its peak values at prism stage (95% of labeled embryos; Table 2), which is at the end of the period when transcripts of the endogenous CyIIa gene can be seen in these cells. As discussed by Arnone et al. (1997), this is due to the long half-life of GFP in skeletogenic mesenchyme cells. Indeed, these retain this fluorescence even at pluteus stages, as indicated quantitatively in Table 2 and illustrated in Fig. 5 (A3, A4). The diffusion of GFP through the syncytial skeletogenic cables accounts for the complete patterns of labeling seen in the skeletogenic structures (Arnone et al., 1996), which is often observed in pregastrular-stage embryos, though the late decline in skeletogenic expression at pluteus stage shown in Table 2 suggests that it does, or (3) that it contains all the elements required for the transient though...

Fig. 3. Map of the CyIIa cytoskeletal actin gene cis-regulatory region, and reporter gene constructs used in this study. A restriction map of the 4.4 kb upstream of the CyIIa translational start codon is shown at the top, including the restriction sites used in assembly of the reporter gene constructs (see Materials and Methods). Colored boxes indicate DNA sequence elements, the functions of which were directly assayed in this work. Major regions are designated A-E below the restriction map and relevant subregions A1, A2, B1-B3 and c are as indicated above the map. The position of a 560 bp leader intron is shown as a bracket below the map, extending from region B1 to near the 3' end of A1. Below the map the locations of binding sites for two different factors are reported. Blue ovals indicate SpCGF1 target sites, the location of which have been established by gel-shift analysis (unpublished results of the authors) or sequence homology (Zeller et al., 1995a,b). The red box labeled Ep stands for the CyIIa gene (the A TG start codons from the translation initiation site and including regions A and B was obtained by K. Katula (GenBank Accession no. AF032880); the remainder of the upstream sequence was obtained by us (GenBank Accession no. AF034254).
strong coelomic pouch expression seen with the endogenous gene.

**Initial localization of CyIIa cis-regulatory elements**

The five regions of the CyIIa sequence included in the correctly expressed EDCBA constructs are defined only by conveniently spaced restriction sites (Fig. 3) and, in an initial series of experiments, the key components of the cis-regulatory apparatus were located within these regions. Briefly, we discovered in these experiments that the complete spatial and temporal control system for developmental expression of CyIIa is located within the BA subregion and that components located elsewhere have the function only of stepping up the amplitude of the expression mediated by BA.

In Fig. 4A the time courses of CAT enzyme output generated by the progressively shorter constructs DCBA-CAT, CBA-CAT, BA-CAT and A-CAT (constructs 2, 3, 7 and 13 of Fig. 3) are compared with that of EDCBA-CAT. Loss of regions E and D successively decreases the amplitude of expression, without much affecting the kinetics. Thus BA-CAT essentially reproduces these kinetics, though at only 40-50% of the amplitude of EDCBA-CAT. Fig. 6 (A1-A4) illustrates the spatial expression mediated by BA-CAT and this is summarized quantitatively in Table 2. BA-GFP accurately reproduces the dynamic spatial pattern of expression of EDCBA-GFP in every embryonic domain and at the expected developmental stages: compare, for example, the embryos at mesenchyme blastula stage (Fig. 6, A1), gastrula (A2), prism (A3) and pluteus (A4) to their equivalents in Fig. 5. It is important to note that no ectopic expression whatsoever is seen at any stage in embryos bearing BA-GFP, beyond the low levels also observed in the control EDCBA-GFP embryos (Table 2). The measurements shown in Fig. 4B provide some further information regarding the amplification functions mediated by elements residing within subregions E, D and C. The main result is that essentially the same levels of expression as generated by EDCBA-CAT are obtained if only region E plus a 250 bp subelement of region C (indicated as lower case ‘c’) are combined with the BA subregions (construct EcBA-CAT, Fig. 3, construct 3). Again, spatial expression mediated by this construct is the same as that of EDCBA-CAT and no additional expression is observed (data not shown). Subregion E has some amplifying activity when combined with BA, accounting for about half the difference between EcBA-CAT and BA-CAT (construct EBA-CAT, Fig. 3, construct 5). However the amplifying elements in subregion C cannot function in this way, since the output of cBA-CAT (Fig. 3, construct 6) is the same as that of BA-CAT (Fig. 4B), and the same is true of CBA-CAT versus BA-CAT (Fig. 4A). The additional activity of the C element is observed in this series only when E is also present.

**Fig. 4.** Temporal and quantitative expression profiles of various reporter gene constructs. The indicated constructs (see Fig. 3) were injected into fertilized eggs and CAT activity was measured as described by Yuh and Davidson (1996) in samples of 50 embryos for each time point. The quantity of CAT enzyme molecules per embryo was determined by comparison of the samples to standards of known CAT concentration, using the conversion 2.6 x 10^7 molecules of CAT enzyme per 10^-8 units of activity (McMahon et al., 1984). Data in A and B were obtained from different experiments on different batches of eggs. Not all the constructs were always tested at once, but the reference plasmid, EDCBA-CAT, was always included. (A) Expression time courses for complete and deleted CAT fusion constructs, compared to time course of accumulation of endogenous CyIIa transcripts. The profiles reported are averages of 3 to 8 independent measurements at each time point. For sake of graph legibility, the standard deviations are not shown; standard deviations were on the average ±12.7% of the values shown (range ±8% to ±25% at the early end of the curves). CyIIa transcripts per embryo (Lee et al., 1986) are indicated for three time points as gray bars (right ordinate). (B) Average CAT enzyme production per embryo for combinations of subregions E and C with subregion BA. Data for each time point were normalized to the value of CAT activity per embryo obtained for the reference plasmid EDCBA-CAT, the value of which was set to 1.0. Barred vertical lines represent standard deviations, which were calculated after the normalization (n varies from 3 to 5 except for the construct cBA-CAT, the activity of which was measured in only two experiments).
a synergistic behavior that indicates interaction between elements of subregions E and C.

### Cis-regulatory functions within the BA subregion

The BA subregion is about 1500 bp in length. As shown in Fig. 4, A-CAT is entirely inactive and subregion A is not required for correct spatial expression. Table 2 shows that construct B-GFP (Fig. 3, construct 9) generates normal spatial patterns of expression; this is illustrated in Fig. 6 (B1-B4). However, the level of expression is depressed, as shown by the smaller fractions of embryos that produce sufficient GFP to permit detection (i.e., about 20-35% of injected embryos instead of the 65-80% seen with BA-GFP; Table 2). While this result demonstrates that all necessary spatial control elements are present in subregion B, it also indicates that subregion A contains an important amplifier of expression located downstream of the transcription start site, probably within the intron. Further experiments placed this amplifier in the A1 subregion: thus BA1-CAT and BA1-GFP (Fig. 3, construct 8) behave exactly as do BA-CAT and BA-GFP, in regard to both quantitative output and spatial expression (data not shown).

The B subregion was divided into three fragments, B1-B3, by means of the restriction sites indicated in Fig. 3. A construct in which the most proximal of these fragments, the 181 bp B1 element, is alone associated with A to form the construct B1A-CAT (Fig. 3, construct 12), is almost completely inactive. As shown in Fig. 7, addition of either the 261 bp B2 subregion or the 465 bp B3 subregion produced constructs (B2B1A-CAT, Fig. 3, construct 10, and B3B1A-CAT, construct 11) which display essentially the same quantitative activity over developmental time as does BA-CAT, with one exception. This is that late in development there is a sharp decrease in expression of B3B1A-CAT, whereas B2B1A-CAT functions in a way that is essentially indistinguishable from BA-CAT. Since late in development expression is largely confined to the endoderm, this result implies that positive endoderm regulators are present in B2, a supposition that we confirm below. Further experiments placed the amplification in the A1 subregion: thus BA1-CAT and BA1-GFP (Fig. 3, construct 8) behave exactly as do BA-CAT and BA-GFP, in regard to both quantitative output and spatial expression (data not shown).

### Table 3. Expression of BA-GFP and mEnBA-GFP in pluteus-stage embryos

<table>
<thead>
<tr>
<th>Construct</th>
<th>Expts†</th>
<th>Scored embryos‡</th>
<th>Positive/negative‡</th>
<th>% PMC expression</th>
<th>% SMC expression</th>
<th>% Coelomic pouch expression</th>
<th>% Gut expression</th>
<th>% Ectoderm expression</th>
</tr>
</thead>
<tbody>
<tr>
<td>BA-GFP</td>
<td>A+B+C</td>
<td>170</td>
<td>111:59 (65.3)</td>
<td>60.3</td>
<td>9.0</td>
<td>18.0</td>
<td>55.0</td>
<td>1.8</td>
</tr>
<tr>
<td></td>
<td>E</td>
<td>135</td>
<td>67:68 (49.6)</td>
<td>67.2</td>
<td>3.0</td>
<td>4.5</td>
<td>61.2</td>
<td>3.0</td>
</tr>
<tr>
<td>mEnBA-GFP</td>
<td>D</td>
<td>92</td>
<td>46:46 (50.0)</td>
<td>52.2</td>
<td>17.4</td>
<td>2.2</td>
<td>8.7</td>
<td>15.2***</td>
</tr>
<tr>
<td></td>
<td>E</td>
<td>97</td>
<td>37:60 (38.1)</td>
<td>75.7</td>
<td>16.2</td>
<td>5.4</td>
<td>10.8</td>
<td>0</td>
</tr>
</tbody>
</table>

*Plutei were collected at 72 to 76 hours postfertilization.
†Individual experiments are denoted by capital letters. Each experiment was carried out on a single batch of eggs. A+B+C represent data reproduced from the three BA-GFP experiments pooled in Table 2, for comparison with Experiment E. In the latter mEnBA-GFP and BA-GFP were compared on the same batch of eggs; Experiment D serves as a separate and independent comparison for the mEnBA-GFP results of Experiment E.
‡Embryos with more than two fluorescent cells were scored as positive.
§% expression = [Σ positive embryos/Σ scored embryos]×100.
¶% territorial expression = [Σ embryos positive in a given territory/Σ positive embryos]×100.
**This batch of embryos produced a remarkably high level of ectopic expression.
Abbreviations: PMC, primary or skeletogenic mesenchyme cell; SMC, secondary mesenchyme cell.

In summary, the results of the experiments summarized in Fig. 7 and Table 2, and illustrated in Fig. 6, produce the following further conclusions regarding the CyIla cis-regulatory system. (1) The spatial control elements included within region B, excluding the basal promoter region, suffice to reproduce the entire spatial pattern of expression. (2) The B3 subregion includes sites whose function is required to set the normal level of expression in skeletogenic cells, and that are particularly important late in development. (3) The combined B1B2 subregion also includes sites that promote skeletal expression, as well as sites which at a quantitative

Therefore, the B3 subelement must contain sites that mediate late skeletogenic expression. Since there are only a few skeletogenic cells that express these constructs even though the GFP product spreads to all of them (Arnone et al., 1997), the absence of these sites does not greatly affect the overall output of CAT enzyme from B2B1A-CAT. It does, however, materially affect the fraction of total injected embryos displaying detectable GFP. Table 2 shows that the level of detectable expression is normal, i.e., the same for B2B1A-GFP embryos as for BA-GFP embryos, late in development when gut is the major locus of expression (66% versus 65%; Table 2), but it is lower earlier when skeletogenic expression is more important. Fig. 6 (C1-C4) illustrates B2B1A-GFP expression in embryos of various stages. It is clear that, even though fewer pluteus-stage embryos express detectable levels of GFP this construct than from BA-GFP, B2B1A-GFP retains the ability to generate skeletogenic mesenchyme, as well as gut and SMC expression.

The active constructs all contain the B1 element, and indeed if this is removed they do not function. This is not solely because B1 contains the transcription start site, where the B1A assembles, since substitution of either the SV40 or Endo16 basal promoters for B1 failed to rescue expression (not shown). The reason is that essential sequences are present in the 20 bp of the B1 element immediately upstream of the start site. Thus, inclusion of these 20 bp but no other region of B1, together with the Endo16 basal promoter, in construct B3B2bEp-GFP (Fig. 3, construct 16) yields a vector that in spatial terms expresses almost identically with B-GFP, as shown in Table 2.
level account entirely for SMC expression and late endodermal expression. (4) There is an essential target site just upstream of the transcription start in region B1. (5) The A1 element contains a strong amplifier of B subregion function, while the A2 element plays no detectable role. Attention is thus focused on B2B1, since it is capable of mediating expression in every embryonic domain where *CyIIa* is transcribed and, except for skeletogenic cells, at the correct relative levels. As above, no ectopic expression was observed even when the test constructs contained only the core B2B1A sequence. A *CyIIa cis*-regulatory element required for endoderm expression: interactions with the same factor that controls midgut expression of the *Endo16* gene

Inspection of the sequence of the B2 subregion revealed a striking similarity with a target site sequence in ‘module B’ of the *Endo16 cis*-regulatory system. This system has been extensively characterized (Yuh and Davidson, 1996; Yuh et al., 1996). Late in development *Endo16* is expressed only in the midgut, and only the *cis*-regulatory element denoted as module B is required for this expression. Though four different proteins bind specifically within module B of *Endo16* (Yuh et al., 1994), only one of these binds uniquely (i.e., there are no other target sites for this protein elsewhere in the 2300 bp *Endo16 cis*-regulatory system). The target site for this factor, here referred to as the ‘En site,’ has been shown to be necessary and sufficient to generate midgut expression in injected *Endo16* fusion constructs: a synthetic oligonucleotide containing this site produces midgut expression, and mutation of a few base pairs in the core of this sequence abolishes it (Yuh and Davidson, unpublished data). The En sites, as they appear in both *Endo16* and the *CyIIa* B2 subregion, are shown in Fig. 8A. This site, CATGAAAT, occurs exactly in both genes and, in *CyIIa*, a second occurrence of ATGAAAT follows immediately, in inverted orientation. A further similarity in the sequences downstream of the En site that is present in the same region of both genes is also indicated in Fig. 8A.

If the inverted repeat of the En site in the *CyIIa* B2 subregion is significant, the *CyIIa* sequence might be expected to bind the protein that interacts at the En site more tightly than does the *Endo16* sequence. This protein can be partially purified by affinity chromatography from 24 hour nuclear extracts using either the *Endo16* oligonucleotide or the *CyIIa* version shown in Fig. 8A (Yuh et al., 1994; and unpublished data). In Fig. 8B is shown a gel-shift experiment in which the partially purified preparation was reacted with the *CyIIa* probe, in the presence of the various oligonucleotide competitors shown in Fig. 8A. The experiment demonstrates that indeed the *Endo16* sequence competes only about 1/20th as well as does the *CyIIa* oligonucleotide. However, it does compete, and as the mutations tested in the *Endo16m3/4* and *Endo16m1/2* lanes of Fig. 8B show, the competition is indeed due to the En site, rather than to the downstream region of sequence similarity. The *Endo16* and *CyIIa cis*-regulatory sequence elements thus do indeed interact with the same DNA-binding protein and, as expected from the fact that it includes two adjacent copies, the *CyIIa* element interacts with the protein much more strongly.

To explore directly the role of the double En site in the *CyIIa cis*-regulatory system, we generated mutations that destroy this site and tested them in vivo. One of these mutations (CyBmEn; Fig. 8A) completely aboliishes the ability of the sequence to form complexes with the protein factor. When introduced into embryos, the constructs *mEnBA-GFP* and *mEnB2B1A-GFP* (Fig. 3, constructs 14 and 15) indeed fail to produce significant GFP expression in endoderm at the pluteus stage of development. Data for *BA-GFP* and *mEnB-BA-GFP* are reproduced in Table 3. As observed earlier (see Table 2), over 55% of *BA-GFP* embryos displaying detectable activity express in the gut at this stage, while only 10-15% of active *mEnB-BA-GFP* embryos express in the gut. The total level of expression is somewhat reduced as well in the *mEnBA-GFP* groups. However, expression in SMC and skeletogenic mesenchyme seems completely unaffected by the En site mutation. We thus conclude that the same *cis*-regulatory interaction as is required for *Endo16* expression in the midgut is also required for *CyIIa* expression in the midgut, and also in the hindgut.

**Expression of the CyIIa gene in sea urchin embryos**

A surprising amount of change has occurred in the regulatory systems controlling spatial expression of cytoskeletal actin genes since divergence of the sea urchin orders Temnopleuroida and Echinoida, in that none of the four *Lytechinus pictus* (Temnopleuroida) cytoskeletal actin genes has a pattern of expression identical to any one of the five *S. purpuratus* (Echinoida) actin genes (Fang and Brandhorst, 1996). However, the embryonic expression of one of the *L. pictus* genes, *LpCl*, resembles that of the *CyIIa* gene of *S. purpuratus* in a number of respects: its transcripts appear in certain cells in the middle of the vegetal plate at the mesenchyme blastula stage, in cells at the tip of the invaginating archenteron in early gastrula stages, in blastocoelar SMCs in the late gastrula as well as in ‘a few’ skeletogenic mesenchyme cells and, at the end of

---

**Table 4. Expression of EDCBA-GFP in *L. pictus* embryos**

<table>
<thead>
<tr>
<th>Stage*</th>
<th>Scored embryos</th>
<th>Positive:negative</th>
<th>% Vegetal expression</th>
<th>% PMC expression</th>
<th>% SMC expression</th>
<th>% Coelomic pouch expression</th>
<th>% Gut expression</th>
<th>% Ectoderm expression</th>
</tr>
</thead>
<tbody>
<tr>
<td>B</td>
<td>21</td>
<td>11:10 (52.5)</td>
<td>52.4</td>
<td>19.0</td>
<td></td>
<td></td>
<td></td>
<td>14.3</td>
</tr>
<tr>
<td>G</td>
<td>35</td>
<td>13:22 (37.1)</td>
<td>23.1</td>
<td>84.6</td>
<td></td>
<td></td>
<td></td>
<td>15.4</td>
</tr>
<tr>
<td>Pr</td>
<td>34</td>
<td>11:23 (32.4)</td>
<td>27.3</td>
<td>90.9</td>
<td></td>
<td></td>
<td></td>
<td>9.1</td>
</tr>
<tr>
<td>P1</td>
<td>41</td>
<td>16:25 (64.0)</td>
<td>62.5</td>
<td>62.5</td>
<td>50.0</td>
<td></td>
<td>37.5</td>
<td>0</td>
</tr>
</tbody>
</table>

* *Blastulae were collected at 24 hours, gastrulae at 44 hours, prisms at 56 hours and plutei at 68 hours postfertilization.
†% expression = [Σ positive embryos/Σ scored embryos]×100.
‡% territorial expression = [Σ embryos positive in a given territory/Σ positive embryos]×100.
embryogenesis, there is intense expression of LpC1 in midgut and hindgut (Fang and Brandhorst, 1996). The main differences are quantitative and temporal. Expression of LpC1 in skeletogenic mesenchyme is slight in L. pictus compared to the intense gastrula-stage expression of CyIIa in the skeletogenic mesenchyme of S. purpuratus and, in L. pictus, hindgut expression of LpC1 begins earlier than does that of CyIIa in S. purpuratus, i.e., at the gastrula rather than the pluteus stage.

The EDCBA-GFP construct was introduced into L. pictus zygotes and its spatial expression monitored as above. The results are illustrated in Fig. 9, and summarized quantitatively.

**Fig. 5.** Representative spatial patterns of expression generated by EDCBA fusion constructs. (A1-A4) Expression of EDCBA-GFP. GFP fluorescence is shown in green in these digital false color images, in embryos examined at the indicated times pf. (A1) Mesenchyme blastula. Expression is seen in newly expressed skeletogenic mesenchyme cells, and in cells of the vegetal region which are likely to be skeletogenic mesenchyme in process of ingression (see text). (A2) Late gastrula. The embryo has been rotated in the plane of the animal-vegetal axis so that the oral side faces down, in order to show the skeletogenic mesenchyme cells arranged as a ring around the archenteron. As discussed by Arnone et al. (1997), expression in all of the 32 skeletogenic cells present at this stage probably results from the diffusion of the fluorescent protein through the skeletogenic syncytium. Several fluorescent secondary mesenchyme cells arising from the tip of the archenteron are also visible (arrowheads). (A3,A4) Anal views of two pluteus-stage larvae, which show expression in skeletogenic cells and also in extended regions of the gut (arrows). The pluteus in A4 also displays GFP fluorescence in one of the already formed coelomic pouches (arrowhead). Note in both plutei the absence of expression in the foregut, which is not a domain of CyIIa expression, and is distinguishable in this view because it bends with respect to the anus-midgut axis towards the oral side of the embryo, facing to the right in these images. (B1-B4) Expression of EDCBA-CAT. CAT transcripts are displayed by WMISH, as a purple stain; digital images made on a Roche ProgRes imaging system are reproduced. (B1) Mesenchyme blastula; staining is confined in this example to ingressed skeletogenic mesenchyme. (B2) Early gastrula, viewed along the animal-vegetal axis. Staining is evident in this optical section on one side of the tip of the archenteron and in some (probably) skeletogenic mesenchyme cells to one side of the archenteron. (B3) Early gastrula, viewed from the side. Staining is again seen at the tip of the archenteron and in some mesenchyme cells (the dark region observed at the base of the archenteron is an optical shadow rather than bona fide staining). (B4) Pluteus-stage embryo displaying staining in midgut and hindgut (arrow) and one coelomic pouch (arrowhead).

**Fig. 6.** Spatial expression of various B subregion CyIIa expression constructs. Composite false color video images of bright-field (gray) and fluorescent (green) exposures are shown superimposed. Embryos were imaged at the indicated developmental times after zygotes were injected. A(1-4) BA-GFP: (A1) Lateral view of mesenchyme blastula embryos, vegetal pole downward. Fluorescent cells are localized in the vegetal region and in ingressing mesenchyme cells. (A2) Mid-gastrula, frontal view, showing expression in several secondary mesenchyme cells emigrating from the tip of the archenteron (arrowheads). Filopodia that also contain GFP are visible in some of these cells. (A3) Lateral view of prism-stage embryo expressing GFP in skeletogenic cells. The right side of the embryo is shown in this image, which includes an entire anal skeletal rod (the other is out of focus). (A4) Pluteus with oral face to the right, viewed from the anal surface. A large patch of fluorescent cells is observed on one side of the tripartite gut, extending from the midgut or the stomach to the hindgut, or intestine. Fluorescence is also observed in skeletogenic cells located along the spicules. B(1-4) B-GFP: (B1) Mesenchyme blastula, as in A1. (B2,B3) Two gastrulae viewed along the animal-vegetal axis. The optical section in B2 shows fluorescent secondary mesenchyme cells at the tip of the archenteron. (B3) The optical section passes through the ring of skeletogenic mesenchyme cells and a single fluorescent SMC near the archenteron is also visible. (B4) Early pluteus-stage embryo oriented so that the oral surface faces left, viewed from the anal side. Expression is observed in both skeletogenic and gut cells. C(1-4) B2B1A-GFP: (C1) Mesenchyme blastula, as in A1. (C2) Late gastrula viewed from the animal pole, with the oral face down. Fluorescent cells are located at the tip of the archenteron and all skeletogenic cells are labeled as well. These are seen in the typical ring-shaped configuration connecting the two clusters at either side of the archenteron, where skeletogenesis has already begun. A triradiate spicule is also visible on the left side of the image. (C3) Anal view of a late prism-stage embryo. Fluorescence is confined to skeletogenic cells (the faint green tinge seen over the gut is a background fluorescent artifact). (C4) Pluteus-stage embryo oriented as in A4 and displaying the same domains of expression.
Fig. 7. Quantitative temporal output of B subregion CyIIa expression constructs. CAT enzyme was measured in embryos bearing B2B1A-CAT, B3B1A-CAT and B1A-CAT expression constructs (see Materials and Methods and legend to Fig. 4) at the indicated times pf (open symbols). The output of BA-CAT, measured on the same batches of eggs, served as a standard of comparison (closed symbols; see also Fig. 4A). Data from three comparable experiments are averaged; standard deviations were on the average ±7.8% of the values shown (range ±3% to ±20%).

in Table 4. Overall, the expression pattern is very similar to that illustrated in Fig. 5A for the same construct in S. purpuratus embryos. Thus EDCBA-GFP is again first expressed in the center of the vegetal plate and in ingressing skeletogenic mesenchyme cells at late blastula stage (Fig. 9A), it is then expressed at the tip of the archenteron during gastrulation (Fig. 9B,C), in scattered secondary mesenchyme cells in the blastocoel of gastrula-stage embryos (Fig. 9D) and in skeletogenic mesenchyme cells of early prism-stage embryos (Fig. 9E). Fluorescence in skeletogenic cells is also observed at later prism and pluteus stages (Fig. 9F,G), together with a strong expression in coelomic pouches (Fig. 9F,G). In some embryos, the gut also expresses (Fig. 9G,H). At late stages, some SMCs that express the transgene are always present, scattered in the blastocoel. When compared with Table 2, the results summarized in Table 4 reveal several quantitative differences. (1) The construct was less active from gastrula to prism stages, though equivalently active at pluteus stage (~35% versus ~75% of embryos display detectable fluorescence). (2) Only about a quarter of expressing L. pictus embryos show fluorescence in skeletogenic mesenchyme, compared to >90% in S. purpuratus. (3) Blastocoeal SMC expression is much stronger in L. pictus at prism and pluteus stages, when 90% and 62% of active embryos display SMC fluorescence while, at these stages, in S. purpuratus only 3-8% of expressing embryos retain SMC fluorescence. (4) Only about half as many L. pictus embryos display gut expression at pluteus stage compared to S. purpuratus. (5) No ectopic expression was observed; i.e., expression in any domain other than those in which both LpC1 and CyIIa are transcribed. Overall, the data suggest that some L. pictus factors do not recognize the target sites for gut and skeletogenic mesenchyme expression in the EDCBA construct as well as do the S. purpuratus factors or they are missing or present at lower concentrations in L. pictus (points 1, 2 and 4). However, the major difference is in the much more extensive expression

Fig. 8. Target site for a positively acting transcriptional regulator required for endoderm expression in Endo16 and CyIIa genes. (A) Comparisons of normal and mutated sequences. The wild-type sequence of the relevant region of module B of the Endo16 cis-regulatory sequence is shown at the top (Yuh et al., 1994) and beneath it the relevant portion of the B2 subregion of the CyIIa cis-regulatory sequence (see text). The mutated version of this sequence that was included in the mEndBA-GFP and mEndB3B1A-GFP constructs (see Materials and Methods) is shown below. The Endo16 3/4 and Endo16 1/2 oligonucleotides were used as competitors in the experiment shown in B. The En site is indicated by the heavily outlined boxes and an additional sequence shared between these regions of the Endo16 and CyIIa regulatory sequence by lightly outlined boxes. Numbers indicate the positions of the sequences shown with respect to the transcription initiation sites, and lower case letters denote base pairs that were altered in the mutations. Arrows indicate the inverted ATGAAAT motif present in the CyIIa gene. (B) Gel-shift reactions displaying cross competition among En sites. Oligonucleotides are shown along the top. In each reaction, 0.1 ng of 32P oligonucleotide probe (CyIIaB2 sequence) was incubated with 1 ng of the DNA-binding factor, which had been partially purified by affinity chromatography. Each reaction contained 200 ng of poly(dI-dC)/poly(dI-dC) and the indicated amount (ng) of oligonucleotide competitor (the dash indicates absence of any competitor). Protein purification and gel-shift analysis were performed as described by Yuh et al. (1994).
of the S. purpuratus construct in SMCs. In this respect, the outcome of the experiment is that expression of EDCBA-GFP is similar to that of the Lpc1 gene in L. pictus rather than that of the CyIIa gene in S. purpuratus. Positive regulators for SMC expression must thus be present longer and/or at higher concentrations or in more active form in these cells in L. pictus than in S. purpuratus embryos.

DISCUSSION

Organization and character of the CyIIa cis-regulatory system

The CyIIa cytoskeletal actin gene is expressed in an unusually complex spatial pattern, as indicated in earlier studies (Cox et al., 1986; Miller et al., 1996) and illustrated in greater detail in Figs 1 and 2 of this paper. We found that, so far as could be determined, the 4.4 kb of genomic sequence included in the initial expression constructs EDCBA-CAT (Zeller et al., 1992) and EDCBA-GFP suffice to reproduce completely the endogenous pattern of CyIIa expression that is summarized in Table 1, in both spatial and temporal respects. Two different kinds of regulatory elements were uncovered. Outside of the B region of the CyIIa cis-regulatory sequence are elements that are required to maintain the amplitude of expression, but for which there is no evidence of spatial control functions. The B region contains elements that are necessary and sufficient to determine in which cells the gene will be transcribed, and when.

Quantitative control of the expression level is effected in part by a dispersed set of regulatory sites scattered about the upstream region of the sequence, in regions E to C. Thus constructs that contain only the BA region operate at only about 40% of the amplitude of those that include the complete EDCBA fragment. The experiments summarized in Fig. 4 show that the normal level of expression can be generated by inclusion either of E plus a small subregion of C, or of region E plus region D. The properties of this dispersed amplification system lead to the prediction that its function may depend on DNA looping caused by interactions between bound SpGCF1 molecules. This factor is able to loop cis-regulatory DNA that contains multiple target sites for it (Zeller et al., 1995a). DNA looping might accelerate transcription initiation by increasing local concentration and promoting interactions with other positively acting basal factors. Fig. 3 (EDCBA-CAT, construct 1) shows that subregions E and c of the CyIIa cis-regulatory domain contain SpGCF1 sites, and one speculation is that these indicate an E-c interaction. SpGCF1 sites are present as well in subregion D of CyIIa upstream sequence and this region as well has a small amplifying activity. Thus the level of expression decreases when region D is removed (Fig. 4A, construct CBA-CAT versus DCBA-CAT). The implication of these sites is based on functional analysis of the SpGCF1 sites in the CyIIa cytoskeletal actin gene (Zeller et al., 1995b; Kirchhammer et al., 1996; Coffman et al., 1997), in which they are arranged in a parallel fashion. In CyIIa, a distal cluster of SpGCF1 sites was shown to mediate a similar 2- to 3-fold amplification of the level of expression without affecting its spatial localization, and this effect requires the participation of other SpGCF1 sites in the proximal region of the cis-regulatory domain close to the start of transcription, perhaps analogous in function to the site in the B1 subregion of the CyIIa gene. There is an apparent redundancy between the E-c and D+E amplifications, since either combination generates the control (i.e., EDCBA-CAT) amount of expression over time, as shown in Fig. 4. This behavior is consistent with the prediction that the multiple SpGCF1 sites in these regions are involved in the observed amplification of expression. That is, only alternative SpGCF1:SpGCF1 complexes could exist at any one time, each involving a minority of the sites in each subregion, so that the output of the whole set of sites and that of spaced subsets will be the same.

A much stronger amplifier of a different kind exists within the leader intron in CyIIa subregion A1. There are no SpGCF1 sites in this area, which lies several hundred bp downstream of the transcription initiation site. Its excision decreases expression to the point where only about half the normal number of embryos bearing GFP expression constructs display detectable levels of expression (compare constructs B-GFP and BA-GFP in Table 2). GFP is at least as sensitive a cytological reporter as is CAT RNA expression monitored by WMISH (Arnone et al., 1997) and CAT RNA is easily detected at levels that produce 1-5x10^6 molecules of CAT enzyme/embryo (Yuh et al., 1996). A many-fold reduction in
activity would thus be required to render expression of CyIIa fusion constructs undetectable (cf. Figs 4, 7). Powerful as it is, the intron amplifier in subregion A1 has no qualitative effect whatsoever on spatial expression and neither do subregions E, D or C.

The total spatial and temporal control ‘package’ in the CyIIa cis-regulatory system appears to lie within region B, and most of it within the 440 bp B2B1 subregion. This is shown in the experiments of Fig. 6 and Table 2. Within the 440 bp B2B1 sequence, we may infer the presence of target sites that mediate the initial activation of the gene in ingressing skeletogenic cells, and its expression in these cells during their climax phase of spicule secretion in gastrula and prism stages, in nascent SMCs at the archenteron tip, in some blastocoelear SMCs late in development, in coelomic pouches, and in midgut and hindgut at pluteus stages. The spatial control apparatus of the S. purpuratus CyIIa cis-regulatory system operates in all these same embryonic domains when introduced into Lytechinus pictus embryos, as shown in Fig. 9 and Table 4. The presentation of the specific factors that mediate expression in these several differentiated and differentiating cell types is therefore a conserved feature in sea urchin embryos. The B2B1 element is structured so as to respond to this set of factors. It is interesting, however, that, in L. pictus, the SMC-specific factors evidently remain active or are extant at higher concentrations further into development than in S. purpuratus. This is not an artifact, since the endogenous C1 cytoskeletal actin gene of L. pictus, which otherwise displays much the same pattern of expression as does the CyIIa gene in S. purpuratus (Fang and Brandhorst, 1996; Cox et al., 1986; Miller et al., 1996; Fig. 1 of this paper) is also expressed later in development in blastocoelear SMCs in L. pictus than is CyIIa in S. purpuratus.

The late endoderm regulator of CyIIa and Endo16

The 2300 bp cis-regulatory system controlling expression of the Endo16 gene in the vegetal plate and gut of S. purpuratus embryos consists of six cis-regulatory modules, in each of which several different factors bind (Yuh et al., 1994, 1996; Yuh and Davidson, 1996). However, as the archenteron is formed and the gut begins to differentiate most of this complex apparatus becomes unnecessary, and for the terminal and permanent phase of Endo16 expression only module B is required. Constructs containing this element alone, associated with a basal promoter and reporter, are expressed exclusively in midgut in late embryos, exactly as is the endogenous Endo16 gene. The key regulator in Endo16 module B is a factor binding at a target site the sequence of which is CATGAATA (the ‘En’ site). We show in the experiments of Fig. 8 that this site is also present twice in the B2 subregion of the CyIIa cis-regulatory system, and that the double En site binds the same protein factor as does the single Endo16 En site, though with at least 20-fold enhanced affinity. Furthermore, mutation of the double CyIIa En site to a sequence that has no binding activity causes specific failure of either B2B1A-GFP or BA-GFP expression in the endoderm, though expression in other domains is unaffected. The shared En target site, and the factor that binds to it, thus define a late embryo, endoderm-specific gene battery, to which CyIIa and Endo16 both belong. Unpublished experiments of Yuh and Davidson prove that an oligonucleotide including a single copy of this site suffices to initiate midgut-specific expression, and thus no additional spatial controls, and more specifically, no negative interactions, are required for midgut expression.

It remains to be explained why the CyIIa gene and the B2B1A-GFP construct are expressed in both midgut and hindgut, rather than midgut alone, as is the Endo16 gene. Two possible explanations are that either the much higher affinity of the double site found in CyIIa permits this gene to respond to lower concentrations of the factor in the hindgut, to which the Endo16 element would be blind, or there is a second hindgut-specific element in the CyIIa B2B1 subregion. This would have to work in concert with the shared En site since, in the absence of this site, little expression in either hindgut or midgut is observed. Resolution of this issue awaits cloning of the factor, now in progress.

The position of CyIIa in the embryonic gene regulatory network

Nowhere in this study did we encounter evidence for negative spatial regulation of expression. The involvement of negative regulators is indicated experimentally by the advent of ectopic expression when their target sites are deleted or mutated. By this criterion, it can be excluded that CyIIa subregions E, D, C, A1, A2 and B3 function as spatial repressor modules, as, for example, do modules E, F and DC of the Endo16 gene (Yuh and Davidson, 1996). It remains formally possible that negative spatial regulation is indeed a feature of the CyIIa system if it is mediated entirely by target sites that are closely linked to those binding the positive regulators within the B2B1 control sequence, or elsewhere. Nonetheless, as so far defined, the CyIIa cis-regulatory system is strikingly different from those of many genes that function during embryonic specification, in that these genes include prominent and easily identified negative regulatory components (see Arnone and Davidson, 1997 for review of many other developmental cis-regulatory systems that include negative controls of spatial expression). The structure of the CyIIa control system can be considered to reflect the role this gene plays in the embryonic regulatory process: CyIIa in fact does not function during embryonic specification but rather after that process has been completed, since this gene is expressed in already well-defined cell types and in already formed morphological components of the embryo.

Comparison with CyIIIa illustrates this argument. This gene encodes a very similar protein but its mode of embryonic expression is very different. CyIIIa is transcribed zygotically only in aboral ectoderm lineages, beginning within a few cell divisions after the segregation of aboral ectoderm founder cells from other pluripotential blastomeres during cleavage (reviewed in Kirchhamer and Davidson, 1996). Thus its initial phase of expression occurs in the context of early specification processes, long in advance of the differentiation of the squamous aboral ectoderm cell type. CyIIa expression is controlled by an extended cis-regulatory system consisting essentially of two complex modules that together occupy over 2 kb of upstream sequence, both of which include essential target sites for negative as well as positive spatial control functions. As we have seen, the CyIIa system is by contrast compact; it causes expression in many different cell types and territories, and there is at the very least no obvious negative regulatory function evident in it. When CyIIa fusion
Regulatory functions that initially define the spatial domains assemblages of target sites for positively acting factors are depend directly on regionalized maternal components. These are the conserved positive regulatory effects that are displayed in the transgenic hatching embryos of Fig. 9 are not in general very robust since the efficiency of expression in *Lytechinus* is quite low (Table 4). Were this same degree of inefficiency to have affected any putative negative spatial interactions, extensive ectopic expression would surely have been observed.

The general implication is that the structure of a cis-regulatory system betrays its position in the normal developmental regulatory network. A series of cis-regulatory modules from mammalian genes that operate downstream of embryonic specification events were collated by Arnone and Davidson (1997). Each consists of relatively compact assemblage of target sites, exclusively for positively acting factors. Three other examples are known from very early sea urchin embryos, all lacking negative spatial regulators and all consisting of single compact modules as well. All three genes are autonomously activated and begin to be transcribed in specific regions of cleavage-stage embryos defined along the preformed animal/vegetal axis by localized or locally active maternal factors. These are the SM50 skeletal matrix gene, which is activated in skeletogenic lineages that arise at the vegetal pole of the egg (Makabe et al., 1995), the SpHE hatching enzyme gene (Wei et al., 1995) and the SpAN protease gene (Kozlowski et al., 1996). The latter two are both activated in non-vegetal blastomeres. In a sense, the autonomous activation of these early embryonic genes is also a downstream function, not of prior zygotic gene regulatory operations, but rather of cytoplasmic regulatory processes that depend directly on regionalized maternal components.

In sum, this argument suggests that genes utilized in development that are spatially controlled by compact assemblages of target sites for positively acting factors are likely to act in already specified cells, downstream of those regulatory functions that initially define the spatial domains where they are to be transcribed.

We are particularly grateful to Dr Chiou-Hwa Yuh for her advice and assistance throughout this work. This paper owes much to the perspicacious and critical insights of Drs James Coffman and Chiou-Hwa Yuh and Professor Ellen Rothenberg of this Institute, who were kind enough to review it in detail for us. The research was supported by the Stowers Institute for Medical Research, and by a grant from NIH (HD-05753). E. L. M. was partially supported by a Caltech SURF fellowship.

**REFERENCES**


Franks, R. R., Hough-Evans, B. R., Britten, R. J. and Davidson, E. H. (1988). Spatially deranged though temporally correct expression of a *CyIII* is quite low (Table 4). Were this same degree of inefficiency to have affected any putative negative spatial interactions, extensive ectopic expression would surely have been observed.

We are particularly grateful to Dr Chiou-Hwa Yuh for her advice and assistance throughout this work. This paper owes much to the perspicacious and critical insights of Drs James Coffman and Chiou-Hwa Yuh and Professor Ellen Rothenberg of this Institute, who were kind enough to review it in detail for us. The research was supported by the Stowers Institute for Medical Research, and by a grant from NIH (HD-05753). E. L. M. was partially supported by a Caltech SURF fellowship.
Expression of the *CyIIa* gene in sea urchin embryos


