R11: a cis-regulatory node of the sea urchin embryo gene network that controls early expression of SpDelta in micromeres

Roger Revilla-i-Domingo, Takuya Minokawa, Eric H. Davidson*

Division of Biology 156-29, California Institute of Technology, Pasadena, CA 91125, USA
Received for publication 24 June 2004, accepted 9 July 2004
Available online 3 September 2004

Abstract

A gene regulatory network (GRN) controls the process by which the endomesoderm of the sea urchin embryo is specified. In this GRN, the program of gene expression unique to the skeletogenic micromere lineage is set in train by activation of the pmar1 gene. Through a double repression system, this gene is responsible for localization of expression of downstream regulatory and signaling genes to cells of this lineage. One of these genes, delta, encodes a Notch ligand, and its expression in the right place and time is crucial to the specification of the endomesoderm. Here we report a cis-regulatory element R11 that is responsible for localizing the expression of delta by means of its response to the pmar1 repression system. R11 was identified as an evolutionarily conserved genomic sequence located about 13 kb downstream of the last exon of the delta gene. We demonstrate here that this cis-regulatory element is able to drive the expression of a reporter gene in the same cells and at the same time that the endogenous delta gene is expressed, and that temporally, spatially, and quantitatively it responds to the pmar1 repression system just as predicted for the delta gene in the endomesoderm GRN. This work illustrates the application of cis-regulatory analysis to the validation of predictions of the GRN model. In addition, we introduce new methodological tools for quantitative measurement of the output of expression constructs that promise to be of general value for cis-regulatory analysis in sea urchin embryos.

Keywords: cis-Regulatory element; Gene regulatory network; delta; Endomesoderm specification; Sea urchin

Introduction

In the process of development, a network of gene regulatory interactions underlies each specification event (Davidson et al., 2002a). These interactions occur at genomic cis-regulatory elements that respond to the set of inputs (i.e., transcription factors) presented in each cell, and which control the expression of each gene, in each domain of the embryo. The properties of the set of all relevant cis-regulatory elements ultimately determine the architecture of the gene regulatory network (GRN) that underlies embryonic specification.

An explicit model of the GRN directing the specification of the distinct endodermal and mesodermal cell types of the sea urchin embryo has been published (Davidson et al., 2002a,b; reviewed by Oliveri and Davidson, 2004). This model predicts inputs to the cis-regulatory elements of the many genes involved, based on an extensive experimental perturbation analysis. The full explanatory power of the model, however, can only be achieved when we have in our hands the key fragments of genomic DNA that execute the cis-regulatory interactions predicted by the model. These cis-regulatory elements will serve to provide the ultimate tests for the correctness of the model. Also, their identification will eventually make possible experiments in which chosen parts of the network of cis-regulatory interactions can be deliberately modified, thereby highlighting the roles of specific portions of the circuitry.

Oliveri et al. (2002) demonstrated that the program of gene expression specific to the skeletogenic primary
mesenchyme cell (PMC) lineage is set in train by the *pmar1* gene, acting through a double repression system. Two developmental functions that are specific to the PMC lineage are set in action as a direct consequence of the operation of this repression system. The first of these is the emission of the Delta signal, which serves as a spatial cue that triggers the specification of mesodermal cell types from common endomesodermal progenitor cells. Expression of the ligand Delta between 7th and 9th cleavages in the micromere lineage, the precursors of PMCs, activates a Notch receptor in adjacent endomesodermal (veg2) cells, and this is required for normal specification of mesodermal fate in these cells (McClay et al., 2000; Sweet et al., 1999, 2002). Thus, the cells of the veg2 territory immediately adjacent to the micromere descendants are specified as mesoderm; the rest of the cells of the veg2 territory will become endoderm. The GRN model predicts that expression of *delta* in the micromere lineage depends on activating factors that are ubiquitously present (Fig. 1). The normally exclusive expression of this gene in the micromere lineage depends on a repressor (“Repressor of mic” in Fig. 1) that is also active everywhere, except in this lineage. There, the *pmar1* gene product in turn represses the gene encoding the otherwise ubiquitous repressor. The second developmental function executed specifically by cells of the PMC lineage is to give rise to the skeletogenic mesenchyme of the postgastrular embryo. The regulatory genes *tbr*, *alx1*, and *ets1* are all known to contribute to the activation of several biomineralization genes that are responsible for the skeletogenic differentiation of the micromere lineage (Ettensohn et al., 2003; Fuchikami et al., 2002; Kurokawa et al., 1999; Oliveri et al., 2002). The GRN model predicts that these three regulatory genes are expressed specifically in micromere descendants due to *cis*-regulatory interactions that include the same mechanism used to localize the expression of *delta*, that is, the *pmar1* repression system summarized in Fig. 1. In particular, this prediction rules out the possibility that any of these three genes is upstream of *delta* or of each other, in agreement with the fact that none of these three genes affects the expression of *delta* or of each other (Oliveri et al., 2002).

The goal of the present study was to test the GRN model by identifying a fragment of genomic DNA from the Strongylocentrotus purpuratus *delta* gene, here called *delta*, that executes the predicted *cis*-regulatory interactions. We first set ourselves to recover the *cis*-regulatory element that drives the expression of *delta* in the micromere descendants at the right time. We were then able to ask whether it responds to the *pmar1* repression system as in the GRN model prediction.

### Materials and methods

**Isolation and analysis of BAC clones containing Spdelta and Lvdelta genes**

A BAC clone named 046A16 containing the *delta* gene had been obtained earlier. BAC clones named 020B17 and 071J09 containing the *Lvdelta* gene were recovered by cross-species hybridization of a *Lytechinus variegatus* BAC genomic library (Cameron et al., 2000). The partial sequence of a *delta* cDNA clone, obtained by Zhu et al. (2001), was used to design the probe for the cross-species hybridization. This probe was obtained by PCR amplification from the cDNA clone (left primer: 5'-aacaacagtcagg-gacact-3'; right primer: 5'-acatggtcagacacatgtg-3').

BAC clones of both species were sequenced by DOE’s Joint Genome Institute. These sequences are available at www.sugp.caltech.edu (under Resources/Annotation). The exons of the *delta* gene in *S. purpuratus* and *L. variegatus* BAC clones were identified using the sequence of both partial *S. purpuratus* and complete *L. variegatus* cDNA clones (Sweet et al., 2002). The sequences were annotated using the SUGAR software package (Brown et al., 2002). This software was used to identify coding sequences of genes neighboring *delta* in the BAC clones.

**Comparison of the genomic sequence around the delta genes of *S. purpuratus* and *L. variegatus**

The FamilyRelations software package (Brown et al., 2002) was used to compare the BAC sequences of *S. purpuratus* and *L. variegatus*. Window sizes used in the comparison ranged from 10 to 200 bp. The pairwise view of

---

**Fig. 1.** Network interactions predicted to be responsible for expression of *delta* in micromere lineage cells (modified from Davidson et al., 2002b; Oliveri et al., 2002). Thick horizontal lines from which bent arrows extend represent *cis*-regulatory elements responsible for expression of the genes named beneath the lines. *cis*-Regulatory elements represented in dimmed color indicate that the gene they control is silent. *cis*-Regulatory elements represented in full color indicate that the gene they control is active. The arrows and barred lines indicate the inferred normal function of the input (activation or repression). (A) In the micromere lineage, the *pmar1* gene is active, and it represses a gene encoding a yet unknown, otherwise globally expressed repressor (“Repressor of mic”), resulting in the activation of *delta* exclusively in these cells. (B) In the rest of the embryo, *delta* is kept silent by repressor of mic. Ub, ubiquitous activator.
the software was used to identify conserved regions. The Dot Plot view was used in some cases to identify the boundaries of the conserved regions found.

Preparation of reporter constructs

Selected regions R1 through R12 of the BAC clone 046A16 of *S. purpuratus* were amplified by means of PCR. The relevant sequences were amplified from the BAC clone by using the “Expand High Fidelity PCR System” (Roche). Primers used for the amplifications were equipped with restriction digest anchors. The sequences of the primers used for the amplification of region R11 were left primer, 5′-aagtaggtaccatgccaacatgaagatgc-3′; right primer, 5′-taagtgaagacctatcagtcctgt-3′.

Reporter constructs R1-GFP through R12-GFP were prepared by cloning the amplified regions R1 through R12, respectively, into the multiple cloning site of the universal *S. purpuratus* expression vector EpGFPII (Cameron et al., 2004). That the correct sequences had been cloned was confirmed by restriction mapping. The vector EpGFPII contains the region around the start of transcription of the *endo16* gene (from −117 to +20). The activity of this basal promoter element has been described in detail elsewhere (Yuh and Davidson, 1996; Yuh et al., 1996, 1998). The EpGFPII expression vector also contains the coding sequence of the GFP protein. All reporter constructs were linearized by restriction digestion upstream of the cloned fragment.

Animals and microinjection of reporter constructs

Microinjection solutions were prepared containing 350–1000 molecules/pl of the reporter construct to be microinjected together with 4- to 9-fold molar excess of *Hind*III-digested carrier sea urchin DNA and 0.12 M KCl (Franks et al., 1990).

Gametes from *S. purpuratus* maintained in our year-round culture system were obtained and microinjected as described by Rast (2000). This protocol is essentially based on the original protocol by McMahon et al. (1985) with significant modifications. The volume of solution microinjected into the embryos was estimated by observing the size of the disturbance produced in the egg cytoplasm. We aimed at microinjecting a volume of 2 or 5 pl of solution depending on the experiment. Experiments were carried out in that nominally 700, 1200, 2500, or 4000 molecules of the reporter construct were microinjected into the eggs.

Microinjected embryos were reared at 14°C to various developmental stages. Some embryos were reared that had not been microinjected, and that had been obtained from the same female and prepared in a similar way as the microinjected embryos. These uninjected embryos were used to control for possible developmental anomalies caused by microinjection.

Simultaneous microinjection of R11-GFP reporter construct and *pmar1* mRNA

The preparation of *pmar1* mRNA by plasmid transcription was performed as described (Oliveri et al., 2002). Microinjection solutions were prepared containing 400 molecules/pl of R11-GFP reporter construct and 22 ng/pl of *pmar1* mRNA together with 7-fold molar excess of *Hind*III-digested carrier sea urchin DNA and 0.12 M KCl. Nuclease-free water was used to prepare the microinjection solutions. Approximately 5 pl of the microinjection solution was microinjected into embryos using the same method as described above for the microinjection of reporter constructs.

Determination of GFP expression in microinjected embryos

Microinjected embryos were visualized on an epifluorescence Axioskop 2 Plus microscope (Zeiss, Hallbergmoos, Germany) equipped with the recording device AxioCam MRm (Zeiss). Expression of GFP in each embryo was determined by the presence of cells fluorescing at a level significantly higher than background. For each GFP-expressing embryo, the location of the GFP-expressing cells was determined according to the morphology of the embryo.

Images were collected and processed in Adobe Photoshop.

Quantification of R11-GFP DNA in microinjected embryos

The Sigma “GenElute Mammalian Total RNA Miniprep Kit” is designed to isolate total RNA. Along with RNA, however, small amounts of DNA are also recovered. This was exploited to quantify the R11-GFP DNA in microinjected embryos in which the GFP expression level was also to be quantified. RNA and DNA were isolated, as described in the manufacturer’s manual, from samples of 100–150 embryos that had been microinjected with the R11-GFP reporter construct or *pmar1* mRNA. Samples were not digested with DNase I, so that the extracted DNA remained in the samples for quantification. Quantitative PCR (QPCR) was conducted using primer sets designed to amplify products of 125–150 bp of the coding sequence of GFP (GFP primer set) and the coding sequence of the *foxb* gene (*foxb* primer set). For sequences of primers, see http://sugg.caltech.edu/resources/methods/q-pcr.psp. Amplification reactions were analyzed on an ABI 5700 sequence detection system using SYBR Green chemistry (PE Biosystems, Foster City, CA). Reactions were run in triplicate with samples from two embryos. Thermal cycling parameters were 95°C for 30 s, 60°C for 1 min, 40 cycles. The number of molecules of R11-GFP DNA per embryo was estimated by using the *foxb* gene as an internal standard; we know that there are two copies of the *foxb* gene per cell (Luke et al., 1997).
Quantification of GFP, delta, and pmar1 mRNA in microinjected embryos

Samples for which the amount of R11-GFP construct DNA had been measured were then treated with DNase I using the DNA-free kit (Ambion, Austin, TX), as described in the manufacturer’s manual, to remove all existing DNA. QPCR was conducted as described above to confirm that no DNA remained in the samples.

cDNA was prepared from the samples by reverse transcription-PCR (RT-PCR). The TaqMan Reverse Transcription Reagents Kit (Applied Biosystems, Foster City, CA) was used for this purpose. The RNA preparation (38.5 μl) was used in a 100-μl reverse transcription reaction (note, though, that in more recent experiments 30 μl was used instead of 38.5 μl, and this seems to improve the efficiency of the RT-PCR).

QPCR was conducted as described in the previous section using primer sets designed to amplify products of 125–150 bp of the cDNA generated from 18S ribosomal RNA, GFP mRNA, ubiquitin mRNA, Spz12-1 mRNA, delta mRNA, and pmar1 mRNA (for primer sequences, see http://sugp.caltech.edu/resources/methods/q-pcr.psp). Amplification reactions were analyzed as described above. Reactions were run in triplicate with cDNA from four to six embryos. For all QPCR experiments, the data from each cDNA sample were normalized against the ubiquitin mRNA and 18s rRNA levels, which are known to remain relatively constant during the developmental stages used (Nemer et al., 1991; Ransick et al., 2002). Absolute quantification of the number of ubiquitin or 18s rRNA transcripts in uninjected embryos was obtained by using Spz12-1 as an internal standard. The number of Spz12-1 transcripts in embryos of the relevant stages had been measured earlier by RNA titration (Wang et al., 1995). The number of ubiquitin or 18s rRNA transcripts was then used for absolute quantification of the number of GFP and delta mRNA transcripts in microinjected embryos.

Results

Spatial and temporal expression pattern of delta during endomesoderm specification

Sweet et al. (2002) showed that in L. variegatus the delta gene is expressed starting at around 7th cleavage in micromere descendants. As the PMCs ingress into the blastocoeol, the expression of Lvdelta in micromere descendants disappears, and expression starts in presumptive secondary mesenchyme cells (SMCs). Whole mount in situ hybridization (WMISH) experiments carried out by Oliveri et al. (2002) indicated a similar pattern of expression in S. purpuratus. The delta gene is expressed in the micromeres starting no later than 8 h after fertilization, and the transcripts remain in their descendants at 18 h. We carried out further WMISH experiments that show that in S. purpuratus, delta transcripts remain present in the micromere lineage until these cells ingress into the blastocoeol at 20 h (data not shown). At this time, expression of delta ceases in the micromere lineage, and as reported for the Lvdelta gene (Sweet et al., 2002), expression is then activated in presumptive SMCs (data not shown). By 24 h, expression of delta is seen only in presumptive SMCs.

To further refine the time at which delta expression starts, we measured the levels of delta mRNA at several stages of development by means of QPCR. As shown in Fig. 2A, these experiments indicate that delta is first expressed between 6 and 8 h after fertilization. Our objective was then to identify the genomic element(s) that is responsible for the specific expression of delta in the micromere lineage, from 6–8 to 20 h after fertilization.

Fig. 2. Temporal expression pattern of endogenous delta gene compared to the temporal expression pattern of GFP mRNA from the R11-GFP reporter construct. (A) QPCR data indicating levels of delta mRNA at different developmental stages. Experimental data are indicated by dots. The line joining these dots is inferred. The error bars represent one standard deviation. Note: for the sake of accuracy in the comparison, the levels of delta mRNA vary extensively between different experiments, depending on the number of R11-GFP DNA molecules incorporated in the genome of the microinjected embryos in each case. The timing at which GFP mRNA expression starts, nevertheless, is accurately reproduced in each experiment.
Genomic sequence surrounding the delta gene

The sequence near the delta gene was annotated to determine the regions where its cis-regulatory system might likely be found. A BAC clone containing the delta gene was sequenced, and the positions of the delta exons in this clone are indicated in Fig. 3A. This BAC clone contains the complete 2394 bp of Delta coding sequence, divided into 11 exons, which together extend over almost 15 kb of the genome. Application of the SUGAR annotation package (Brown et al., 2002) revealed the presence of the coding sequence of an unnamed gene about 37 kb upstream of the

![Fig. 3. Comparative interspecific sequence analysis. (A) Map of the S. purpuratus genomic sequence around the delta gene (top) and pairwise view of a FamilyRelations comparison of S. purpuratus (Sp) and L. variegatus (Lv) orthologous genomic sequences around the delta gene. Horizontal black lines represent these BAC sequences. Coordinate positions in the respective BAC clones are indicated. Pink blocks indicate the position of other genes immediately upstream and downstream of delta in the S. purpuratus genome. Orange blocks indicate the positions of the coding sequence of delta, as obtained from sequenced cDNA clones and by comparison to the coding sequence of Lvdelta. START indicates start of translation, STOP, the coding sequence termination. The two blue dashed lines indicate the limits of the S. purpuratus genomic sequence that was compared to the orthologous L. variegatus genomic sequence. The shaded area indicates the region of the genome of S. purpuratus between the start of translation and the coding sequence termination. Each tic on the sequences demarcates 1 kb from the previous tic. The red lines connecting the two BAC sequences indicate interspecific sequence similarities, here consisting of 100% identity for a sliding window of 20 bp. Yellow stars indicate sequence similarities that contain simple sequences, for example, microsatellites. Numbered green boxes indicate the sequence regions that were selected to be tested experimentally. (B) Dot plot view of part of the FamilyRelations comparison in A but using a different criterion. In this case, each dot indicates interspecific similarities, consisting of 90% identity in the sequence of the two species, for a sliding window of 10 bp. Here the S. purpuratus sequence is on the horizontal axis, and the L. variegatus sequence is on the vertical axis.](image-url)
**delta** start of translation and another gene is predicted about 33 kb downstream of the termination of the **delta** coding sequence (Fig. 3A). Therefore, the cis-regulatory regions that control the expression of **delta** are likely to reside within the 85 kb of genomic sequence between the genes identified upstream and downstream of **delta**.

Identification of conserved genomic sequences as putative cis-regulatory elements

We compared the relevant genomic region of *S. purpuratus* with the orthologous region of the *L. variegatus* genome to identify conserved sequence patches. *S. purpuratus* and *L. variegatus* diverged about 50 million years ago, and this distance is useful for the identification of putative cis-regulatory elements, which are recognized as significantly conserved sequence elements (Yuh et al., 2002, 2004). To this end, *L. variegatus* BAC clones containing the coding sequence of *Lvdelta* were obtained by cross-species hybridization of a *L. variegatus* BAC genomic library and sequenced. Analysis of the *S. purpuratus* and *L. variegatus* genomic sequences with the SUGAR software revealed 70 kb of *L. variegatus* BAC sequence that is orthologous to the *S. purpuratus* genomic sequence around **delta**. This 70 kb of genomic sequence extends, in the *S. purpuratus* genome, from the next gene upstream of **delta** to about 18 kb downstream of the termination of the **delta** coding sequence (Fig. 3A).

The 70 kb of orthologous genomic sequence was scanned computationally for short conserved sequence regions using the FamilyRelations software package (Brown et al., 2002). This tool allows for the detection of sequence similarities above a chosen criterion within sliding windows set at chosen window sizes. Fig. 3A shows a pairwise view of this comparison. In this view, every red line connecting the *S. purpuratus* and *L. variegatus* sequences indicates an interspecific sequence similarity at the chosen criterion; in the case of Fig. 3A, it represents the presence of a sequence stretch of 20 bp that is identical in the two species. Given the stringency of the criterion chosen, only regions with very high similarity are detected.

The comparison of the two orthologous sequences was also visualized using a dot plot view. Fig. 3B shows a small portion of such a view. Each dot indicates the presence of a sequence of 10 bp in which at least 9 bp is identical in the two sequences. The low stringency of the criterion used in Fig. 3B results in a high level of noise due to random matches. These random matches appear as isolated dots, while sequence similarities corresponding to “true” conservation can be distinguished by their diagonal continuity. The dot plot view offers an important advantage with respect to the pairwise view, in that it better shows the structure of the sequence similarities. Thus we see that most of the conserved stretches in Fig. 3A consist of isolated blocks of very well-conserved sequence, with sharp boundaries, surrounded by very poorly conserved sequence. Fig. 3B shows one of these blocks.

Conserved blocks with significant similarity were chosen and analyzed in detail using the Mapping Closup function of FamilyRelations. Regions consisting of simple sequence (e.g., microsatellites; yellow stars in Fig. 3A), regions consisting of coding sequence (orange blocks in Fig. 3A), and conserved regions shorter than 100 bp were excluded from further analysis. The remaining conserved patches were considered putative cis-regulatory elements of the **delta** gene. A total of 12 such regions, named R1 through R12 (green blocks in Fig. 3A), were selected for experimental test of cis-regulatory function during the relevant developmental stages.

The R11 DNA fragment accurately generates the early expression pattern of the **delta** gene

To test the cis-regulatory function of the selected conserved regions R1–R12, we prepared constructs R1-GFP–R12-GFP. Each construct was microinjected into embryos, and expression of GFP was monitored at several stages between fertilization and mesenchyme blastula stage. In the present report, we focus exclusively on region R11. As the following work shows R11 generates the early expression pattern of **delta**. The cis-regulatory activities of the remaining conserved regions and the overall organization of the **delta** gene will be discussed elsewhere, because while some of these constructs are active they do not generate the phase of expression we are interested in the present report.

Table 1(“R11” column) indicates the locations where GFP expression was observed at three different stages of development in embryos that had been microinjected with the R11-GFP reporter construct. Images of some representative embryos are shown in Figs. 4A–F. In interpreting these data, we have to bear in mind two technical points: first, due to the time it takes for the GFP to be translated and for the chromophore to form, there is a delay of about 4 h from the time the mRNA accumulates to when fluorescence becomes detectable; second, that exogenous DNA is incorporated in mosaic fashion in microinjected sea urchin embryos. Within minutes after injection into the egg cytoplasm, linear DNA molecules are ligated together to form one or a few very large, end-to-end concatenates (McMahon et al., 1985). Then, early in cleavage, an exogenous DNA concatenate is incorporated randomly into the genome of usually one blastomere (Flytzanis et al., 1985; Hough-Evans et al., 1988; Livant et al., 1991). Once incorporated, the exogenous DNA replicates together with the endogenous DNA and is inherited by the progeny of the host cells (Flytzanis et al., 1987; Franks et al., 1988; Livant et al., 1991). As a consequence, each of the microinjected embryos will have one or a few clones of cells that contain exogenous DNA, and that therefore have the possibility to express the reporter gene.

As shown in Table 1, GFP fluorescence was observed at blastula stage (15–17 h after fertilization) in a significant number of embryos (29%), indicating that R11 had driven
the expression of GFP mRNA at least as early as 11 h. At the 15–17 h blastula stage, it is often impossible to identify distinct cell types in the embryos by morphological observation alone. The small micromeres can sometimes be distinguished, however, thereby indicating the vegetal pole of the embryo. Those embryos expressing GFP in which this identification was possible showed that GFP fluorescence is always localized to cells immediately next to the small micromeres (Fig. 4A). Whenever expression of GFP was observed, it was confined to a small region of the embryo (Figs. 4A and D).

At early mesenchyme blastula stage (20–22 h after fertilization), as the micromere descendants begin their ingestion into the blastocoel, the embryos expressed GFP either in the ingressing cells (Fig. 4B) or in underlying cells which from their position appear about to ingress (Fig. 4E; Table 1). At later mesenchyme blastula stage (24–26 h after fertilization), when all cells of the micromere lineage have completed ingestion into the blastocoel, the PMCs, vegetal plate cells, and ectodermal cells of the embryo can be clearly distinguished. At this stage, expression of GFP was seen almost exclusively in the PMCs (Table 1, Figs. 4C and F).

These results indicated that the R11-GFP construct drives expression of GFP in cells of the micromere lineage beginning sometime before 11 h postfertilization. No cells other than the micromere descendants accumulate significant levels of GFP mRNA, even transiently. Arnone and Davidson (1997) showed that GFP is very stable in these embryos, and therefore the fluorescence seen at any given stage of development is the sum of all prior episodes of expression. Had GFP transcripts been transiently accumulated to significant levels in cells other than those of the micromere lineage any time before 20 h postfertilization, we would have seen fluorescence in the descendants of those cells at mesenchyme blastula stage. To check for this, embryos expressing during earlier blastula stages were kept alive and individually monitored for GFP expression until they reached mesenchyme blastula stage. All these embryos expressed in PMCs at mesenchyme blastula stage (data not shown). Importantly in no case did expression disappear between these two stages, demonstrating that indeed GFP fluorescence is stable and not transient, and most importantly, that the only cells showing GFP fluorescence throughout the blastula stage are precursors of the PMCs.

Hough-Evans et al. (1988) and Livant et al. (1991) showed that incorporation of exogenous DNA happens most often at the 3rd or 4th cleavage stages. According to this, we would expect from the lineage map (Cameron et al., 1987; Davidson, 1986) that in only 35–40% of the microinjected embryos would exogenous DNA be incorporated in the cells of the micromere lineage. Consistent with this, previous cis-regulatory studies on a gene encoding a biomineralization protein specific to PMCs yielded exactly this frequency of expressing embryos (Makabe et al., 1995). Similarly, Table 1 shows that the fraction of embryos expressing R11-GFP between 20 and 26 h was 35–38%.

It remained to be seen whether the developmental time course of GFP mRNA expression driven by R11 accurately mimics the temporal expression pattern of delta. To resolve this we compared the levels of GFP mRNA to those measured for delta at several stages of development.

<table>
<thead>
<tr>
<th>Stage</th>
<th>Injection</th>
<th>R11 % TOTAL</th>
<th>R11 % Expr</th>
<th>R11 + pmar1 % TOTAL</th>
<th>R11 + pmar1 % Expr</th>
</tr>
</thead>
<tbody>
<tr>
<td>Blastula (15–17 h)</td>
<td>TOTAL</td>
<td>231</td>
<td>67</td>
<td>117</td>
<td>54</td>
</tr>
<tr>
<td></td>
<td>Expressing</td>
<td>67</td>
<td>29</td>
<td>54</td>
<td>46</td>
</tr>
<tr>
<td>Early mesenchyme blastula (20–22 h)</td>
<td>TOTAL</td>
<td>515</td>
<td>107</td>
<td>93</td>
<td>87</td>
</tr>
<tr>
<td></td>
<td>Expressing</td>
<td>182</td>
<td>35</td>
<td>93</td>
<td>87</td>
</tr>
<tr>
<td></td>
<td>Ingressing PMCs</td>
<td>179</td>
<td>35</td>
<td>98</td>
<td>80</td>
</tr>
<tr>
<td></td>
<td>Blastula wall cells</td>
<td>6</td>
<td>1</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>Mesenchyme blastula (24–26 h)</td>
<td>TOTAL</td>
<td>897</td>
<td>125</td>
<td>120</td>
<td>96</td>
</tr>
<tr>
<td></td>
<td>Expressing</td>
<td>344</td>
<td>38</td>
<td>120</td>
<td>96</td>
</tr>
<tr>
<td></td>
<td>PMCs</td>
<td>316</td>
<td>35</td>
<td>92</td>
<td>78</td>
</tr>
<tr>
<td></td>
<td>Vegetal plate</td>
<td>29</td>
<td>3</td>
<td>8</td>
<td>7</td>
</tr>
<tr>
<td></td>
<td>Presumptive ectoderm</td>
<td>20</td>
<td>2</td>
<td>6</td>
<td>6</td>
</tr>
</tbody>
</table>

('% TOTAL' means percentage of embryos with respect to the 'TOTAL' number of embryos observed; '% Expr' means percentage of embryos with respect to the number of 'Expressing' embryos.

Notes.

(1) At early mesenchyme blastula stage, cells were scored as ingressing PMCs if they were inside the blastocoel, ingressing into the blastocoel, or immediately next to the cells ingressing into the blastocoel.

(2) Values shown have been obtained by summing over all the experiments carried out in which no anomalies were observed in the development of the microinjected embryos.

(3) Each value in the table derives from experiments carried out using eggs from at least three different females. In some cases, eggs from as many as 20 different females were used.
in embryos that had been microinjected with the R11 construct (Fig. 2). These data show that transcription of GFP mRNA begins between 6 and 8 h after fertilization, and continues to increase up to at least 20 h after fertilization (Fig. 2B). The time course almost exactly resembles the temporal expression pattern of the endogenous delta gene from fertilization to 20 h postfertilization. Temporally as well as spatially, the expression pattern driven by R11 accurately recapitulates the early expression pattern of delta.

The R11 expression pattern depends on operation of the pmar1 repression system

If the sequence element R11 is responsible for localizing the expression of delta to the micromere descendants, we would expect that it should contain binding sites for those transcription factors that control the expression of delta. Therefore, if the predictions of the network model are correct, we would expect that R11 should contain binding sites for activating factors that are ubiquitously present in the embryo, and that it should respond to a repressor, expression of which is prevented in micromere descendants by the pmar1 gene product (Fig. 1; Oliveri et al., 2002). Thus, we would expect that ectopic expression of pmar1 in cells other than micromere descendants should result in R11-driven expression of GFP in those cells; global expression of pmar1 should result in GFP expression everywhere.

Global expression can be effected by microinjection of pmar1 mRNA into fertilized eggs (Oliveri et al., 2002). To examine the effect on the expression pattern generated by
R11, we analyzed GFP expression in embryos that had been microinjected simultaneously with the R11-GFP reporter construct and with pmar1 mRNA. Results are shown in Table 1 and Fig. 4, which compare the expression of GFP driven by the R11-GFP reporter construct in normal embryos (Table 1, column “R11”; Figs. 4A–F) and in embryos globally expressing pmar1 mRNA (Table 1, column “R11 + pmar1”; Figs. 4G–L). At the 15–17 h blastula stage, the number of embryos expressing GFP is significantly higher in embryos with ectopic pmar1 expression (46%) than in normal embryos (29%). Most strikingly, some of these embryos displayed expression in several patches of cells, located on opposite sides of the embryo (Figs. 4G and J). This was never observed in normal embryos (Figs. 4A and D), and it indicated that expression of R11-GFP is no longer localized to the micromere lineage in embryos that ectopically express pmar1 mRNA.

At early mesenchyme blastula stage, the majority of embryos bearing ectopic pmar1 mRNA expressed GFP in cells other than the micromere lineage: 78% of the expressing embryos display ectopic GFP expression when pmar1 is expressed ectopically, whereas only 3% do so normally (Table 1). Figs. 4H and K clearly illustrate this effect. In these embryos, expression of R11-GFP is observed in cells of the blastula wall in addition to the ingressing cells that normally express the construct.

At 24–26 h after fertilization, the morphology of embryos undergoing global pmar1 expression (Figs. 4I and L) is no longer normal (Figs. 4C and F), probably because all cells in the embryo have been transformed to PMC fate (Oliveri et al., 2002, 2003). It is now impossible to distinguish the cells that would have normally become PMCs from the rest of the cells. Note that at this stage almost all (96%) the embryos globally expressing pmar1 display GFP (Table 1). So high a percentage of embryos expressing GFP can be expected only if R11 activates expression of GFP in any cell of the embryo where the exogenous R11-GFP DNA happens to be integrated. Also, the large size of the clones expressing GFP in these embryos (Figs. 4I and L) is also consistent with the conclusion that R11 drives expression of GFP in all cells that also express pmar1 mRNA.

The results shown in Table 1 and Fig. 4 are consistent with the hypothesis that wherever pmar1 is active, the regulatory activity of R11 is derepressed, as predicted by the GRN model (Fig. 1). Because pmar1 is normally transcribed only in the micromere descendants, R11 normally drives expression of GFP only in these cells, but when pmar1 is expressed ectopically in all cells of the embryo, expression of GFP driven by R11 is also expanded to the whole embryo. The response of R11 to global expression of pmar1 thus accurately recapitulates the response of the endogenous delta gene to the same perturbation, which causes expression of delta to expand to all cells (Oliveri et al., 2002). This equivalence provides strong support for the claim that the R11 element suffices to generate the control functions that govern delta expression in the cells of the micromere lineage.

**Measurement of incorporated exogenous DNA and its transcriptional activity**

To measure quantitatively the derepression of R11-GFP caused by global expression of pmar1 mRNA, we developed what is essentially a new method of assessing expression of exogenous constructs in vivo. This relies on use of QPCR to assess both the amount of incorporated DNA and the amount of transcript generated from it in the experimental embryos.

An important preliminary consideration is that the amount of GFP mRNA that will be transcribed in a sample of embryos microinjected with the R11-GFP reporter construct will depend on the overall number of DNA molecules that happen to be incorporated. Furthermore, as shown earlier (Livant et al., 1990), we may assume that for a small amount of incorporated R11-GFP DNA, the amount of GFP mRNA transcribed will be linearly dependent on the number of incorporated R11-GFP DNA molecules. By “small” here is meant much smaller than the number of molecules of R11-GFP DNA required to saturate the transcription of GFP mRNA due to titration of the regulatory factors. Under these conditions, the activity of R11-GFP reporter constructs in the different samples can be compared by normalizing the absolute number of GFP mRNA molecules in each sample to the number of R11-GFP DNA molecules incorporated per embryo in that sample.

The diagram in Fig. 5A describes how the method was carried out. Total RNA was isolated along with small amounts of DNA from samples of embryos microinjected with R11-GFP. The number of molecules of R11-GFP DNA per embryo in these samples was estimated using QPCR. The single copy foxB gene was used as an internal standard to assess the number of genomes recovered, as described in Materials and methods. To quantify the levels of GFP mRNA, the samples were treated with DNase I to remove all existing DNA. cDNA was then prepared from the sample and the levels of GFP mRNA were measured by QPCR. To confirm that the method we used consistently recovers genomic DNA as well as RNA, we tested the nucleic acids isolated from over 50 samples of embryos microinjected with R11-GFP reporter construct, and from more than 10 samples of embryos that had not been microinjected. In all samples, the foxB sequence was amplified to detectable levels, indicating that sufficient genomic DNA had always been recovered (data not shown). GFP DNA was detected in all samples of embryos microinjected with R11-GFP reporter construct, but not in any samples of uninjected embryos (data not shown).

To confirm that the R11-GFP DNA detected by this method consisted mainly of DNA that had been incorpo-
rated into the genomes of the microinjected embryos, measurements of the number of molecules of R11-GFP DNA were made at several developmental stages. Fig. 5B shows the relative amounts of DNA detected by the GFP primer set to the number of copies detected by the foxb primer set at the indicated times postfertilization. Measurements were made on samples from which total RNA and small amounts of DNA had been isolated using the “GenElute Mammalian Total RNA Miniprep Kit” (Sigma).

Embryo, and at 24 h there are approximately 60,000 molecules of R11-GFP DNA. The amount of DNA estimated at the 10-h stage represents approximately 15 times the amount of DNA microinjected in each embryo (approximately 700 molecules of DNA were microinjected); and the amount estimated at the 24-h stage represents approximately 85 times the number of DNA molecules microinjected. Therefore, if any unincorporated exogenous DNA is detected at all, it represents an insignificant proportion of the detected DNA. The amount of exogenous DNA measured as incorporated into the genomes of the embryos of Fig. 5B is in fact the amount that would be present if most of the microinjected DNA had been incorporated into a single blastomere genome between 3rd and 4th cleavage, as expected (Hough-Evans et al., 1988).

We have not calculated the efficiency of the RNA kit used in isolating genomic DNA (see Materials and methods). It is possible that the efficiency of this kit in recovering genomic DNA varies from sample to sample. But it is important to note that even if that was the case, it would not affect these measurements, because the use of the internal foxb DNA standard renders the results independent of the absolute fraction of genomic DNA recovered. We need only assume that no part of the genome is isolated with a different systematic efficiency than any other part.

Timing and magnitude of the effect of ectopic pmar1 on R11 expression

The effect of global pmar1 mRNA expression on the activity of R11-GFP normalized to the amount of incorporated DNA is shown in Fig. 6A and on the level of delta mRNA in Fig. 6C. In Fig. 6B, the normalized activities of R11-GFP of Fig. 6A have all been multiplied by the number of R11-GFP DNA molecules incorporated in the control sample expressing the endogenous pmar1 gene normally. This gives a direct comparison of the amounts of transcript that would have been produced had all the samples contained the same amount of exogenous DNA. The values in Fig. 6B still reflect normalized activities, and the advantage of this representation is that it allows us to compare the relative levels of GFP mRNA at different stages. More importantly, this representation is equivalent to that of Fig. 6C, and therefore Fig. 6B can be directly compared to Fig. 6C.

Up to 5 h after fertilization, the delta gene is silent, and ectopic expression of pmar1 mRNA has no effect on the very low observed levels of delta mRNA. But at the 10- and 15-h stages, it results, respectively, in greater than 3- and 5-fold increases in endogenous delta mRNA (Fig. 6C). Similarly, the activity of R11-GFP is significantly increased (more than 2-fold) by ectopic pmar1 expression at the 10- and 15-h stages, but it is not affected at the 5-h stage (Figs. 6A and B). Even though Fig. 6A seems to indicate that R11-GFP is active at 5 h after fertilization, Fig. 6B clearly shows that the amount of GFP mRNA at this stage is insignificant; less than five molecules per
embryo are detected at the 5-h stage. These results confirm that the derepression of the R11 regulatory element caused by ectopic pmal1 mRNA can be detected as a quantitative increase in the activity of R11-GFP; and they also indicate that this happens at the same stages at which expression of endogenous delta is observed to increase in the same embryos.

It is important to note that in experiments in which the amount of incorporated R11-GFP DNA was approximately 10 times larger than in the experiment of Fig. 6, the amount of measured GFP mRNA was also approximately 10 times larger (data not shown). Therefore, in the experiment of Fig. 6, the amount of R11-GFP DNA incorporated was far from the amount of DNA required.

Discussion

We show here that the R11 DNA fragment contains cis-regulatory information sufficient to recreate the exact spatial and temporal pattern of the delta gene in its initial phase of expression, when it is transcribed exclusively in the micromere lineage early in development. As we shall report elsewhere, a different cis-regulatory module of the delta gene reproduces the next phase of its expression in secondary mesenchyme precursors. The properties of R11 bear directly on the GRN model for endomesoderm specification, as we discuss briefly below. But before this there are two methodological aspects of this work that bear consideration. These are the means by which R11 was found and the means by which its response to experimental perturbation was quantitatively determined.

Identification of R11 by interspecific genomic sequence comparison

In the comparison of the orthologous genomic sequences of S. purpuratus and L. variegatus surrounding the delta gene, the R11 cis-regulatory element appears as a 3 kb-long block of very well conserved sequence surrounded by very poorly conserved sequence (Fig. 3B). Conservation at the level of 90–100% identity covers almost the entire block. Previous studies from this laboratory have already shown that the evolutionary distance between S. purpuratus and L. variegatus is very useful for identification of functional cis-regulatory elements (Brown et al., 2002; Yuh et al., 2002, 2004). The immediate identification of R11 by the same method adds further supporting evidence. R11 is located more than 13-kb downstream of the termination of the delta gene coding sequence (Fig. 3A), and finding this element by conventional mapping or deletion methods would have been extremely laborious. The “FamilyRelations” software (Brown et al., 2002) was used for this interspecific sequence comparison, and other more or less equivalent sequence comparison methodologies have also been successful in identifying cis-regulatory elements in many different genes and species pairs (e.g., Aparicio et al., 1995; Brickner et al., 1999; Hardison, 2000; Loots et al., 2000; Manzanares et al., 2000; Muller et al., 2002; Nonchev et al., 1996; Oeltjen et al., 1997). Many additional examples could be cited. Given the appropriate species distance for the gene in question, interspecific sequence comparison can be an extremely effective method for locating the control machinery of the genome; at the right distance, cis-regulatory elements stand out very clearly as conserved sequence patches. In the case we illustrate in Fig. 3, the signal to noise ratio is so high that the element is unmistakably distinguished from the surrounding sequence.

What is most impressive is how sharply defined the boundaries of the element are. These boundaries are revealed explicitly by the dot plot of Fig. 3B at the 9 out of 10 identity criterion applied here. This represents in
principle a significant augmentation of methodologies for cis-regulatory analysis: experimental procedures generally provide either a convenient but much larger fragment than the actually functional regulatory module, or a “minimal element” that gives some function. We see that there is available an additional independent criterion, the computational definition of the natural boundaries of the conserved regulatory sequence patch.

Quantification of exogenous incorporated DNA and reporter mRNA

This work has included an augmentation of experimental cis-regulatory analysis methods as well. There are many applications when it is necessary to measure the output of an exogenous cis-regulatory expression construct in quantitative terms. Chief among these is to determine the effects of various mutations; and to determine the response of the element to perturbation of a trans input that affects its activity, positively or negatively. So far, quantification of the level of expression of exogenous constructs in sea urchin embryos has been achieved by use of a reporter gene encoding chloramphenicol acetyltransferase (CAT), the enzymatic activity of which can be measured in lysates encoding chloramphenicol acetyltransferase (CAT), the enzymatic activity of which can be measured in lysates (e.g., Flytzanis et al., 1987; Livant et al., 1988; Kirchhammer et al., 1996; Yuh and Davidson, 1996; Yuh et al., 1996, 1998, 2001, 2004).

In general, the amount of transcribed reporter mRNA depends on the numbers of molecules of the expression construct that are incorporated into the genomes of microinjected embryos (Livant et al., 1988). Flytzanis et al. (1987) used filter hybridization with radioactively labeled probes to measure incorporated DNA, and concluded that if enough construct DNA is incorporated, the levels of transcribed reporter mRNA are independent of the amount of this DNA; in other words, the amounts of expression describe a saturation function with respect to the number of incorporated DNA molecules (cf. Livant et al., 1988, 1991). This fact has been exploited in several studies to analyze the quantitative effects of mutations on the kinetics of cis-regulatory expression (e.g., Yuh et al., 1998, 2001).

Here we describe a new method, based on QPCR measurements, for the simultaneous quantification of transcribed reporter mRNA and incorporated reporter DNA. This method provides certain advantages with respect to measurement of CAT activity. First, because the level of transcription is obtained by directly measuring the amount of reporter mRNA at given times, the result depends only on the rates of construct transcription and of reporter mRNA turnover, rather than on these rates plus the rates of reporter protein synthesis and protein turnover. The last is particularly difficult to measure or estimate. Second, the QPCR method is compatible with the use of any reporter gene, including GFP, rather than limited to the use of the CAT reporter. Thus, for example, measurements can be carried out on samples of embryos that have previously also been scored for spatial GFP fluorescence. Third, the amount of incorporated reporter DNA is very easily quantified at the same time and in the same sample of embryos in which the reporter mRNA is measured. This provides a very efficient way of normalizing the levels of reporter mRNA to the amount of incorporated reporter DNA, which is a major source of variation in the activity of different batches of embryos. The major advantage of this normalization is that it is no longer required to microinject enough DNA so that transcription of the reporter gene reaches saturation. Finally, the QPCR method allows for measurement of the expression of any endogenous gene(s) in the same sample of embryos in which the levels of reporter mRNA and DNA are quantified. This can be particularly useful for analysis of the effects of perturbations on an incorporated cis-regulatory element.

cis-Regulatory analysis of R11 expression and the network model for endomesoderm specification

The GRN model predicts genomically encoded cis-regulatory interactions that would explain the expression of its constituent genes at the right places and times to serve their developmental functions in the specification process (Davidson et al., 2002a,b; Oliveri and Davidson, 2004; for current version of this model, see http://sugp.caltech.edu/endomes/). The cis-regulatory element controlling early delta gene expression in the micromere lineage is a particularly important node of the GRN: it accounts for transcriptional expression of the spatial information that sets in train the specification of the secondary mesenchyme domain of the embryo. The specific prediction is that the expression of delta in the micromere lineage under control of this cis-regulatory element depends on activating factors that are ubiquitously present and on a repressor (”Repressor of mic” in Fig. 1) that is in turn repressed exclusively in the cells of the micromere lineage in consequence of pmar1 expression (Fig. 1; Oliveri et al., 2002). The isolation and experimental analysis of the R11 delta cis-regulatory element reported here proves that there indeed exists a genomic DNA fragment that executes exactly the predicted interactions.

In untreated embryos, R11 accurately drives expression of the reporter construct, exclusively in the micromere lineage, while in embryos globally expressing pmar1 mRNA, R11 becomes capable of causing expression in any cell of the embryo. This behavior perfectly reproduces the response of the endogenous delta gene to the same perturbation. R11 may contain target sites for activating factors that are ubiquitously present, and it may also contain the sites for the repressor controlled by the pmar1 gene product. However, until such sites are identified by mutation, it remains possible that this repression is mediated indirectly, and that R11 (and the delta gene) is controlled by a localized activator that is under pmar1 system control. But the kinetics of delta gene expression, which very shortly
follows pmr1 activation (Oliveri et al., 2002; Fig. 2 of this paper), suggest that the repression is likely to be exerted directly.

This work illustrates one of the major useful aspects of the GRN model, viz, that the model specifies experimentally testable candidate inputs into each of its cis-regulatory elements. In turn experimental cis-regulatory analysis feeds back into the network model by validating these predictions. As such analysis is extended to the key nodes of the GRN, there emerges an explanatory structure that will directly represent the genomic regulatory code underlying specification and development.

Acknowledgments

We are very grateful to Dr. Ellen Rothenberg and Dr. Paola Oliveri for very useful criticism and fruitful discussions. We also thank Pat Leahy for taking care of the sea urchins used for this work, and Jane Wyllie and Dr. Veronica Hinman for technical guidance. The first author also thanks Dr. Albert Erives for his instruction on molecular biology, and he would like to apologize to the sea urchins sacrificed for this work. This research was supported by NIH Grants HD37105 and GM61005 and by NASA/Ames Grant NAG2-1587.

References


McCay, D.R., Peterson, R.E., Range, R.C., Winter-Vann, A.M., Ferkowicz,


