Graduate Theses, Dissertations, and Problem Reports

1999

Comparative analysis of the promoters of the CyI-CyIIa-CyIIb actin gene cluster in the sea urchin Strongylocentrotus purpuratus

Ruth Lynn Dukes

West Virginia University

Follow this and additional works at: https://researchrepository.wvu.edu/etd

Comparative Analysis of the Promoters of the CyI-CyIIa-CyIIb Actin Gene Cluster

in the Sea Urchin *Strongylocentrotus purpuratus*

Ruth Lynn Dukes

Thesis submitted to the

College of Arts and Sciences

At West Virginia University

In partial fulfillment of the requirements

For the degree of

Master of Science Degree

in

Biology

James McGraw, Ph.D., Chair

Philip Keeting, Ph.D.

Jeff Price, Ph.D.

Department of Biology

Mogantown, West Virginia

1999

Table of Contents

List of Figures and Tables

Acknowledgements

Abstract

Comparative Analysis of the Promoters of the CyI-CyIIa-CyIIb Actin Gene Cluster

in the Sea Urchin *Strongylocentrotus purpuratus*

Ruth Lynn Dukes

Three of the eight actin genes in the sea urchin *Strongylocentrotus purpuratus* were used for a comparative analysis. The cytoplasmic actin genes CyI, CyIIb and CyIIa are linked in the genome and encode very similar proteins. However, their spatial and temporal expression patterns are differentially regulated. The region of CyIIa upstream of the ATG and putative transcription start site was subcloned and sequenced. Sites of protein-DNA interaction were identified using DNaseI footprinting. Comparisons of this data with that previously obtained for CyI and CyIIb showed that while CyI and CyIIb share an identical CArG (or serum response element) and extensive sequence homology in the upstream region, homology with CyIIa is essentially limited to consensus protein binding sites, including the serum response element (SRE) and an octamer motif. Along with several other sites, the SRE and octamer were found to bind protein, though binding to CyIIa was altered in comparison to CyIIb. These data indicate that the CyIIa upstream sequences have diverged extensively since its proposed duplication from CyI. These changes in sequence are likely the cause of an altered set of transcription factor binding sites, which results in a pattern of expression that is markedly different from that of CyI/CyIIb.

Introduction

The actin proteins are well characterized molecules, functioning as structural elements in eukaryotic cells. In the sea urchin *Strongylocentrotus purpuratus*, the actins are encoded by members of a multigene family, which are differentially expressed in both embryonic and adult tissues (Shott et al., 1984; Durica et al., 1988). Study of actin gene structure and function thus lends itself as a useful tool for understanding regulation of eukaryotic gene expression during development. In addition, it will serve here as a model in an investigation of the evolutionary processes regarding the origin of such multigene families.

The genome of *S. purpuratus* contains eight nonallelic actin genes. Of the eight genes, two are suspected pseudogenes (Lee et al., 1984; Scheller et al., 1981; Shott et al., 1984). Five of the genes encode cytoskeletal actins (CyI, CyIIa, CyIIb, CyIIIa and CyIIIb) and one, a muscle-specific actin (M). Hybridizations with probes of the 3' untranslated regions of the genes show no cross reaction between subgroups and at higher stringency, no cross reaction at all, even within those subgroups. The subgroups are designated by roman

numerals, which signify differences in the 3' trailer region and by small letters within the subgroups to indicate some homology (Lee et al., 1984).

The functional actin genes are separated into three linkage groups, shown in figure 1. CyI, CyIIa, and CyIIb are found within a thirty kilobase stretch of DNA and are separated from each other by approximately 7-8 kilobases. CyIIIa and CyIIIb comprise the second linkage group and are located about six kilobases apart. M actin is in a third chromosomal location. By southern blot analyses, these genes seem to account for all the actin coding genes and actin-like sequences in the *S. purpuratus* genome (Lee et al., 1984; Shott et al., 1984; Scheller et al., 1981).

Expression patterns of the actin genes have been described in the adult animal and in the embryo (Cox et al., 1986; Shott et al., 1984; Crain et al., 1981). The M gene is expressed specifically in muscle and myogenic cells. An M actin specific probe hybridizes to transcripts in muscle containing tissues of the adult and pluteus larva; no hybridization is evident in the coelomocytes, a non-muscle adult cell type (Shott et al., 1984). During embryogenesis,

FIGURE 1: ORGANIZATION AND EXPRESSION OF THE ACTIN GENE FAMILY IN _Strongylocentrotus purpuratus_. Actin genes are named M (muscle) or Cy (cytoskeletal/nonmuscle cell types). Roman numerals designate the three nonhomologous 3' nontranslated trailer sequences found in the cytoskeletal actin genes. The letters a,b, and c designate different though homologous trailer sequence variants. Linkage data are from analyses of cloned genes (J.J. Lee _et al_., 1984; Akhurst _et al_., 1986). Expression patterns in adult tissues were determined by RNA gel blot hybridizations, as reported by Shott _et al_. (1984). No expression of CyIIIa or CyIIIb has been detected in the adult animal. Expression of actin genes in embryonic cell types summarized from the _in situ_ hybridization study of Cox _et al_. (1986). Not shown are probable unlinked pseudogenes CyIIc and CyIIIc. [Adapted from E.H. Davidson, C.N. Flytzanis, J.J. Lee, J.J. Robinson, S.J. Rose, and H.M. Sucov (1985). _Cold Spring Harbor Symposium on Quantitative Biology_, **50**, 321.

```
        →                    6 kb              →                   3 kb            →
   ┌─────────┐                            ┌─────────┐                        ┌─────────┐
───┤   CyI   ├────────────────────────────┤  CyIIa  ├────────────────────────┤  CyIIb  ├───
   └─────────┘                            └─────────┘                        └─────────┘
```

embryo

early all regions of blastula   mesenchyme cells   all regions of blastula
       mesenchyme cells   mesenchyme cells

late   oral ectoderm   regions of gut   oral ectoderm
       gut   gut

adult intestine   intestine   intestine
       tubefoot   coelomocyte   tubefoot
       coelomocyte
       lantern muscle

```
                       →                    5 kb              →
                  ┌─────────┐                            ┌─────────┐
──────────────────┤  CyIIIa ├────────────────────────────┤  CyIIIb ├──────────────────
                  └─────────┘                            └─────────┘
```

embryo   aboral ectoderm only   aboral ectoderm only

```
                                  →
                             ┌─────────┐
─────────────────────────────┤    M    ├─────────────────────────
                             └─────────┘
```

embryo   pharyngeal muscle

adult   tubefoot
                       lantern muscle
                       gut muscle

```
                                          ────
                                          1 kb
```

transcripts of M appear only in those cells which are myogenic; these cells differentiate after gastrulation (Cox et al., 1986).

CyI and CyIIb show similar patterns of expression in the embryo. Transcripts of each of these genes appear at very low levels in the maternal mRNA store and both CyI and CyIIb are activated in all regions of the early embryo between 10-14 hours of development. After 18 hours, CyI and CyIIb transcripts disappear from the aboral ectoderm, and ultimately remain present only in the gut, oral ectoderm, and some mesenchyme cells (Cox et al., 1986). In the adult animal, expression of CyIIb is not significantly detectable in most tissues except the intestine by DNA dot blot analysis, whereas the CyI message appears in relative abundance in most adult tissues (Shott et al., 1984).

CyIIa is activated later during embryonic development than are CyI and CyIIb. The CyIIa transcripts are found in mesenchymal cells and in some regions of the gut and its precursors (Lee et al., 1985; Cox et al., 1986; Shott et al., 1986). CyIIIa and CyIIIb mRNAs are confined to the aboral ectoderm of the embryo and pluteus larva (Angerer and Davidson, 1984; Cox et al., 1986). Neither CyIIIa nor CyIIIb is expressed after metamorphosis (Shott et al., 1984).

Each of the sea urchin actin genes is regulated differently, with distinct levels and timing of expression within particular cell types. Current models of eukaryotic gene regulation indicate that the basis for these distinct patterns of expression lies in the number and nature of the protein binding sites in the regulatory region of each gene. These sites are ultimately a consequence of the evolutionary history of the gene family members. Hence a comparative analysis of the promoter sequences, and the sites of protein binding within the regulatory DNA of each gene, may provide important information regarding the evolution of differential gene expression.

In this research project, the promoter regions of the linked actin genes CyI, CyIIa and CyIIb were compared. These genes are linked in the genome in the same 5'-3' orientation (Scheller et al., 1981). The proteins encoded by each gene are strikingly similar. CyIIb differs from CyI by only one amino acid, and from CyIIa by three residues. CyIIa and CyI are distinguished by four amino acid residues (Durica et al., 1988; Shott et al., 1984).

At the nucleotide level CyI is more closely related to CyIIb than to CyIIa within the 5' flanking regions, within the intron following codon 121 and within silent sites (those sites at which nucleotide changes do not replace amino acids)

7

in the coding region. However, in the 100 nucleotides flanking the 3' splice junction and within the 3' flanking region, CyIIa and CyIIb show greater sequence similarity. All three genes are conserved in intron sequences within codon 204 (Durica et al., 1988; Shott et al., 1984). Hence, CyIIa and CyIIb have been included in the same subgroup. Sequence analysis of intron sequences of the three genes has indicated a single origin for the cluster which may have arisen through duplication events. Subsequent divergence of the upstream regions, by addition of new sequences or by base changes occurring at an estimated rate of 1% per 1.4 million years (the mutation rate of noncoding DNA) (Schuler et al., 1983), has resulted in differential expression of the genes of this linkage group. The protein products retain a remarkable amount of identity to each other.

By in situ hybridizations of embryos with probes to the 3' untranslated region of mRNA, no significant differences between CyI and CyIIb transcripts have been found in terms of tissue distribution, although CyI messages do appear in relatively greater abundance. CyIIa transcripts show a distinct pattern of distribution by comparison (Cox et al., 1986). Titrations of embryonic messages using RNA probes transcribed from actin gene specific clones have shown that CyI and CyIIb transcripts begin to accumulate earlier during development than those of CyIIa, with an apparent increase in abundance as early as 10-12 hours

post-fertilization. CyI and CyIIb are represented in the maternal store of mRNA, whereas CyIIa messages are not appreciably detectable until approximately 20 hours of development (Lee et al., 1986). This data suggests a correlation between expression and evolution of the CyI-CyIIa-CyIIb linkage group members. A duplication event may have yielded (essentially) two CyI actin genes, with subsequent divergence of the upstream regulatory elements into what is now CyIIa. It appears, then, that another duplication (of CyIIa) gave rise to a presumptive CyIIb. That CyIIb shares greater sequence similarity with CyI may be accounted for by a gene conversion of CyIIb with CyI (Durica et al., 1988). Comparative analysis of 5' cis-acting control regions of these three genes should offer a better basis for understanding of this evolutionary process.

Materials and Methods

Restriction Mapping

The plasmid pCyIIa-CAT.1 (Zeller, et al.) was provided by Dr. Roberta Franks of the National Institutes of Health. This plasmid was restriction mapped by digestion with the enzymes Pst I, Sma I, Acc I, Kpn I, Eco RI, Bgl II and Sal I (fig. 2). A typical restriction enzyme digestion contained about 1 μg of pCyIIa-CAT.1 DNA, 2μl of 10X universal buffer (0.33 M Tris-acetate, pH 7.9, 0.66 M potassium acetate, 0.10 M magnesium acetate, 5 mM dithiothreitol, 40 mM spermidine), 0.5 μl of the appropriate enzyme (10 - 50 units) and double distilled water (ddH$_2$O) to a final volume of 20 μl in a sterile, 1.5 ml microcentrifuge tube. The digestion was incubated in a 37°C water bath for 45 minutes. The resulting DNA fragments, along with HindIII - digested λ DNA, were electrophoresed on a 0.8% agarose gel containing 2 μl of a 20 mg/ml stock of ethidium bromide (EtBr) per 100 ml of gel solution. The distance each fragment traveled on the gel was measured from the bottom of the wells. Distances in millimeters were plotted against the known sizes in kilobases of λ DNA Hind III

FIGURE 2: SUBCLONING OF pCyIIa-CAT.1. The strategy for subcloning the plasmid pCyIIa-CAT.1 is shown. A) On the original plasmid, restriction enzymes used are abbreviated as follows: RI = Eco RI; K = Kpn I; P = Pst I; Bg = Bgl II. B) Each of the fragments generated by digestion with Eco RI, Kpn I, and Bgl II and subcloned into pUC18 are designated by the letters W, X, Y and Z for reference in the text. C) The plasmid containing fragment X was digested with Pst I, thereby generating fragments designated XI and XII, which were subcloned in a second round into pUC18.

A. Restriction map of pCyIIa-CAT.1

RI   K   P   Bg Bg   RI

                                        P   CAT
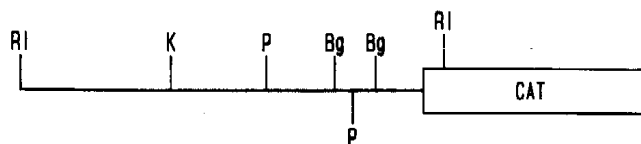
B. First round of subcloning into pUC18

W
1.4 kb

X
1.5 kb

Y
0.4 kb

Z
0.6 kb
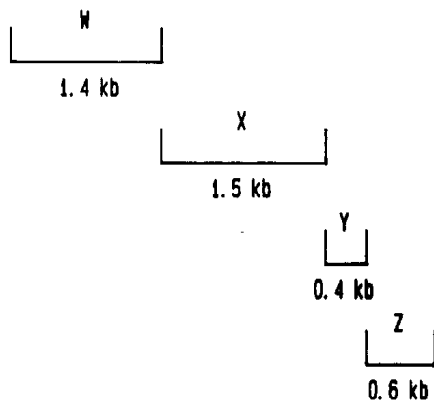
C. Second round of subcloning into pUC18

XI
0.9 kb

XII
0.6 kb   1.0 kb

12

fragments on semi-log graph paper to generate a standard curve for determining the size of each restriction fragment of pCyIIa-CAT.1.

Subcloning

Various subclones from pCyIIa-CAT.1 were generated using conventional cloning methods (Maniatis, Sambrook and Fritsch, 1989). Approximately 20 μg of pCyIIa-CAT.1 plasmid DNA was digested with 15 units each of the restriction enzymes Eco RI, Kpn I, and Bgl II in the presence of 6.5 μl of 10X U buffer and 46.5 μl ddH$_2$O for a total volume of 65 μl. The digestion mixes were incubated at 37°C for 45 minutes. After addition of another 15 units of each enzyme, the samples were incubated for an additional 45 minutes.

The cloning vector was prepared by digesting pUC18 DNA with the appropriate enzymes to accommodate a particular fragment from pCyIIa-CAT.1. pUC18 was digested with the following combinations of enzymes: Eco RI and Kpn I, Kpn I and Bam HI, Bam HI and Eco RI, and Bam HI alone. A typical digestion contained 3 μg of pUC18, 2 μl of 10X U buffer, 0.5 μl of each enzyme and ddH$_2$O to a total volume of 20 μl.

The digested pUC18 and restriction digested pCyIIa-CAT.1 DNA were electrophoresed on a 0.8% low-melting (65°C) preparative agarose gel containing EtBr. The bands were visualized under ultraviolet light and cut from the gel. The gel pieces were placed into sterile, 1.5 ml microcentrifuge tubes and incubated in a 65°C water bath until the agarose melted. One volume of ddH₂O was added and mixed by vortexing. The solution was made 50 mM Tris using 1 M Tris, pH 8.0. To remove agarose from the DNA, the sample was subjected to a standard phenol, phenol/sevag (24:1 chloroform : isoamyl alcohol), sevag extraction. The DNA was reprecipitated by addition of 1/15 volume 3 M NaOAc, pH 6.0, and 2 volumes cold 95% ethanol with incubation at -80°C for 30 minutes. The DNA was recovered by centrifugation in a cold microcentrifuge at full speed for 15 minutes. The pellet was washed with 70% ethanol, centrifuged for 5 minutes, then washed with 95% ethanol and centrifuged for another 5 minutes. The DNA pellet was dried under vacuum and resuspended in 20 μl of 10 mM Tris, pH 8.0, 1 mM EDTA (TE).

The gel-isolated fragments of pCyIIa-CAT.1 were ligated into the appropriately digested pUC18 molecules in a reaction which consisted of

~0.75 µg of the isolated pCyIIa-CAT.1 fragment, 1 µl ligase buffer (300 mM Tris-HCl, pH 7.8, 100 mM MgCl$_2$, 100 mM DTT, 10 mM ATP), 1 µl of 1 mg/ml bovine serum albumin, 0.15 µg pUC18 DNA, 1 µl ligase (400 units) and ddH$_2$O to 10 µl total volume. The ligation mixtures were incubated at 16°C, overnight (~16 hours).

The ligation mixtures were used to transform competent DH5α cells. A 300µl frozen aliquot of competent DH5α cells was thawed on ice, mixed with 5 µl of the ligation mixture, and incubated on ice for 90 minutes. The cells were then heat shocked in a 42°C water bath for 90 seconds. After a 5 minute incubation on ice, 1ml of LB broth (1 % bactotryptone, 0.5 % bactoyeast extract, 1 % NaCl) was added and the cells were grown in a 37°C shaker for 1 hour. The 1 ml culture was spread with an ethanol flame - sterilized, bent glass rod onto solid LB agar (1.5% bactoagar) containing the antibiotic ampicillin (AMP) and 5-bromo-3-indolyl b-D-galactoside (X-gal) for selection. The 150 mm plates were incubated overnight (~16 hours) in a 37°C incubator.

Those colonies which did not utilize X-gal appeared as white colonies on the plate. A sterile toothpick was used to transfer cells from individual white colonies to a second, numbered LB-Amp plate ("Master Plate"). The same toothpick was then swirled in a sterile 1.5 ml microcentrifuge tube containing 1 ml of LB broth and 10μl of 10 mg/ml ampicillin. The cells were grown overnight in a 37°C shaking incubator and analyzed using a standard mini-prep protocol.

Mini - Prep Analysis

Bacterial cells were pelleted by centrifugation in a room temperature microcentrifuge for 3 minutes. The supernatant was removed and the bacterial pellet resuspended in 20 μl lysis buffer (8% sucrose, 0.5% Triton X-100, 50 mM EDTA, pH 8.0, 10 mM Tris, pH 8.0, and 0.15 mg lysozyme). The cells were lysed by boiling for 40 seconds and the chromosomal DNA was pelleted by centrifugation in a refrigerated microcentrifuge for 15 minutes. The pellet was removed using a pasteur pipet and a 5 μl aliquot of the supernatant, plus 2 μl of 5X sample buffer (0.25% bromophenol blue, 0.25% xylene cyanol FF, 15% Ficoll) and 5 μl ddH$_2$O, was electrophoresed alongside λ Hind III marker DNA and supercoiled pUC18 on a 0.8% agarose

gel containing EtBr. DNA was visualized under ultraviolet light and samples containing plasmid DNA migrating at a slower rate than pUC18 were selected. Plasmid DNA in these samples was precipitated by addition of 1/15 volume of 3 M NaOAc (pH 6.0) and 2 volumes of isopropanol, and an incubation at -80°C for 30 minutes. The DNA was recovered by centrifugation in a refrigerated microcentrifuge for 15 minutes. The pellet was washed with 70% ethanol, centrifuged for 5 minutes, washed with 95% ethanol, and centrifuged another 5 minutes. The DNA pellet was dried under vacuum and resuspended in 20 μl TE. The desired recombinants were identified in a restriction enzyme digestion, using the appropriate enzymes to release the inserted DNA fragments.

Large Scale Plasmid Preparation

Cells from positive colonies on the Master Plate were transferred with a sterile toothpick into a flask containing 200 ml of LB broth and 60 μg of ampicillin per milliliter of culture media. Each recombinant clone was grown in the 200 ml preparations in a 37°C shaking incubator overnight. The cells were pelleted by centrifugation at 5500 rpm at 4°C for 5 minutes and resuspended in 20 ml of a buffer containing 500 mM sucrose, 10 mM EDTA,

25 mM Tris, pH 8.0, and 80 mg lysozyme. After a 10 minute incubation at room temperature, the cells were lysed by addition of 40 ml of 0.2 M NaOH, 1 % SDS with a 5 minute incubation on ice. Thirty milliliters of cold 3 M KOAc was mixed into the solution by swirling and the preparation incubated for 10 minutes on ice. Cellular debris was removed from the samples by centrifugation at 5500 rpm, at 4°C for 5 minutes. The supernatant was filtered through sterile gauze into a fresh centrifuge bottle and 0.6 volume of isopropanol was added. After a 10 minute incubation at room temperature, the now crude preparation of plasmid DNA was centrifuged again at 5500 rpm, 4°C for 5 minutes. The pellet was washed with 70 % ethanol to remove residual salt and subsequently resuspended in 8 ml TE. Cesium chloride (8.8 g) was dissolved into the sample and its density at this point was adjusted with TE or additional CsCl to 1.5 - 1.6 g/ml. The sample was transferred to a 11 ml capacity crimp-top ultracentrifuge tube and 100 μl of 20 mg/ml EtBr was added. Samples were centrifuged in a Ti-1270 (Sorvall) rotor at 50,000 rpm, 20°C for 20 hours. The band of plasmid DNA was visualized under ultraviolet light, removed with a 3 cc syringe and transferred to a 15 ml falcon tube. To remove EtBr from the DNA, 3 volumes of H_2O - saturated n-butanol was added and mixed by inversion. The phases were allowed to separate and the top/organic phase was removed. This process was generally

repeated twice more or until the pink color of the EtBr was no longer apparent. The plasmid DNA was precipitated by addition of 2 volumes of TE and 2 volumes of cold 95% ethanol and incubation at -20°C overnight. The DNA was pelleted by centrifugation in a 15 ml Corex tube at 10,000 rpm, 4°C for 20 minutes. The pellet was washed with 70% ethanol and 95% ethanol before being taken up in 250 μl TE. The plasmid DNA was quantified by reading the optical density of a diluted sample (10 μl DNA in 990 μl TE) with a Milton Roy spectrophotometer at 260 nm.

Sequencing

DNA sequencing was conducted using the Sanger - dideoxy method, specifically the Sequenase® version 2.0 protocol supplied by United States Biochemical. Approximately 3.5 μg of plasmid DNA was denatured by addition of 2 μl of 2M NaOH, 2 mM EDTA during a 30 minute incubation at 37°C. The DNA was precipitated with 3 μl of 3 M NaOAc, pH 4.6 and 75 μl of cold 95% ethanol. After ~30 minutes at -20°C, the DNA was recovered by centrifugation in a refrigerated microcentrifuge for 15 minutes at full speed. The pellet was washed with 70% ethanol, then drained and lyophilized under vacuum. The DNA was resuspended in 7 μl of ddH$_2$O. Two microliters of

5X sequenase buffer (200 mM Tris-HCl, pH 7.5, 100 mM $MgCl_2$, 250 mM NaCl) and 1 µl (0.5 pmole) of forward pUC primer (5' GTTTTCCCAGTCACGAC 3'), pUC reverse primer (5' CAGGAAACAGCTATGAC 3') or custom synthesized primer was added. The samples were heated at 65°C for 2 minutes, then cooled slowly to ~35°C for annealing of primers to the templates. Each annealing mixture was spun briefly and placed on ice. In the labeling reaction, 1 µl dithiothreitol, 2 µl of 1:5 dilution of a 5X labeling mix (7.5 mM each dGTP, dCTP, dTTP), 0.5 µl $[^{32}P]$-dATP (10 mCi/ml, 800 Ci/mmol) and 2 µl of 1:8 diluted Sequenase enzyme (T7 DNA polymerase) were added to each annealing mixture. The components of the labeling reaction were mixed and the samples were incubated for 3 minutes at room temperature. To each of four termination tubes (containing a mixture of deoxynucleotides and either ddATP, ddCTP, ddGTP, or ddTTP), 3.5 µl of the labeling reaction was added and the termination reactions incubated at 37°C for 4 minutes before 4 µl of stop solution (95% formamide, 20 mM EDTA, 0.05% bromophenol blue, 0.05% xylene cyanol) was added. The reactions (A, C, G, T) were heated at 75°C for 2 minutes immediately prior to loading onto a 6% 40:2 acrylamide:bis-acrylamide - urea gel which had been prerun for one hour at 1700 volts in 1X TBE buffer (0.089 M Tris, 0.089 M boric acid, 2 mM

EDTA). Samples were loaded 3 times, with approximately 3 hours between each successive loading. After electrophoresis, the gel was exposed to X-ray film at -80°C overnight. Gaps in nucleotide sequence data were closed using custom synthesized oligonucleotide primers (Midland Reagent Company). A primer made to the end of the sequence obtained for the BglII - EcoRI fragment (5' GTCCATCAAT-TTAACCGGG 3') generated data for a complete sequence of that region. Similarly, a primer made to the end of the sequence obtained for the Pst I - Pst I fragment (5' ATCGATTGCTGCCCTACC 3') made possible a complete sequence of that region. Finally, a primer made to the beginning of the sequence obtained for the Bgl II - Bgl II fragment (5' GATCATGCTCGAACCATCG 3') generated data which confirmed the orientation of this fragment within the subclone.

Addition of Mn buffer (0.15 M sodium isocitrate, 0.1 M $MnCl_2$) to the labeling reaction yielded sequence close to the primer for confirmation of cloning sites within each of the subclones as necessary. For obtaining sequence much farther from the primer than usual, 1.5 μl of sequence extending mixture (50 mM NaCl and 180 mM each dATP, dCTP, dGTP, dTTP) plus 1.0 μl of the appropriate termination mixture was used instead of 2.5 μl of termination mixture alone.

DNase I Footprinting

1. Preparation of end - labeled probes

A 912 base pair fragment of pCyIIa-CAT.1 was isolated from a Pst I digestion of 100 μg of the plasmid using the low melting agarose gel as described previously for subcloning. One sample of the fragment was then digested with Cla I, another sample with Dde I. Approximately 3 μg of DNA was digested in the presence of 4 μl 10X Cla I buffer (10 mM Tris, pH 7.5, 10 mM MgCl$_2$, 50 mM NaCl, 1 mM DTT) or 10X Dde I buffer (6 mM Tris, pH 7.9, 6 mM MgCl$_2$, 150 mM NaCl, 1 mM DTT), 12 units of either Cla I or Dde I, and ddH$_2$O to a final volume of 40 μl. Each digestion incubated at 37°C for 45 minutes. The DNA was then precipitated by addition of 1/15 volume 3M NaOAc, pH 6.0, and 2 volumes cold 95% EtOH. After incubating at -80°C for 15 minutes, the samples were centrifuged for 15 minutes at full speed in a refrigerated microcentrifuge. The pellets were washed with 70% EtOH and vacuum dried. Each DNA pellet was resuspended in 5 μl of TE.

The resulting fragments of the Cla I digestion were subsequently end - labeled with [α-^{32}P]dCTP, and those of the Dde I digestion were end - labeled with [α-^{32}P]dATP. Each reaction contained 5 μl of the DNA, 1 μl 10X Klenow buffer (100 mM Tris, pH 7.5, 600 mM NaCl, 70 mM MgCl$_2$), 2

μl [α-^{32}P] dCTP or dATP (20 μCi, 3000 Ci/mole), 0.5 μl each 10 mM dGTP, dTTP, and either dATP or dCTP, and 0.75 μl (5000 units/ml) Klenow. The samples incubated at 30°C for 20 minutes, then 0.5 μl of cold nucleotide was added : dCTP to the Cla I - digested DNA, dATP to the Dde I - digested DNA. The samples were incubated 15 minutes longer at 30°C. The Klenow enzyme was inactivated by heating the samples at 75°C for 10 minutes.

One microliter of each of the labeling reactions was transferred into 100 μl of TE for calculation of counts incorporated by trichloroacetic acid (TCA) precipitation. Ten microliters of this volume was spotted directly onto a glass microfiber filter. To the remaining 90 μl, 5 μl of 5 mg/ml herring sperm DNA and 1 ml cold 10% TCA were added. The sample incubated on ice 5 minutes and was passed through a glass microfiber filter which was then washed with cold 10% TCA and cold 95% EtOH. Both filters were placed in scintillation vials with 5 ml of cocktail and counted. The number of counts incorporated was determined by dividing TCA soluble counts (second filter) by total counts (first filter).

To complete preparation of necessary probes, the Cla I and Dde I - digested end - labeling reactions were subjected to a secondary digestion. To

the Cla I - digested reaction, 3 µl 10X Ava II buffer (50 mM potassium acetate, 20 mM Tris-acetate, 10 mM magnesium acetate, 1 mM DTT), 0.75 µl Ava II (4000 units/ml) and ddH$_2$O to a final volume of 30 ml were added. To the Dde I - digested reaction, 3 µl 10X Cla I buffer, 0.75 µl Cla I and ddH$_2$O to 30 µl were added. Both samples incubated 45 minutes at 37°C. The fragments of the secondary digestions were separated on a nondenaturing 5% polyacrylamide gel. The top plate of the gel apparatus was removed to expose the gel and a piece of X-ray film, wrapped with aluminum foil to be light proof, was laid over the gel. After 15 minutes of exposure at room temperature, the film was developed and used to determine the location of desired bands in the gel. These bands were removed from the gel and transferred to a small culture tube containing 1 ml TE. The DNA was eluted from the gel pieces in a 37°C shaker overnight. The liquid containing the DNA was separated from the gel piece and vacuum dried to a volume of 150 µl. The samples were passed through a spin column of Sephadex G-50 and 1 µl in 5 ml cocktail was counted for an estimate of the counts per microliter of probe.

2. Maxam - Gilbert sequencing

For the G reaction, 50,000 cpm of each probe was dried down to a 1 x 10^4 cpm/µl concentration (5 µl) and 200 µl dimethyl sulfate (DMSO) buffer

(50 mM sodium cacodylate, pH 7.0, 1 mM EDTA, pH 8.0) was added. After a 5 minute incubation on ice, 1 μl of 10% DMSO (1:9 dimethyl sulfate:ethanol) was added. The reactions were incubated at 20°C for 3.5 minutes and 50 μl DMSO stop solution (1.5 M NaOAc, pH 7.0, 1 M β-mercaptoethanol, 250 μg/ml yeast tRNA) and 750 μl 95% EtOH was added. For the A+G reaction, 100,000 cpm of probe was dried down to a 10 μl volume to which 10 μl ddH$_2$O was added. The samples were chilled on ice for 5 minutes and 50 μl formic acid was added. After a 5 minute incubation at 20°C, 180 μl hydrazine (HZ) stop solution (0.3 M NaOAc, pH 7.0, 0.1 M EDTA, pH 8.0, 100 μg/ml yeast tRNA) and 750 μl 95% EtOH were added. After a 15 minute incubation at -80°C, both the G and the A+G reactions were centrifuged for 10 minutes at full speed in a refrigerated microcentrifuge. To the pellets, 250 μl 0.3 M NaOAc and 750 μl cold 95% EtOH were added. The samples were again incubated at -80°C for 15 minutes and centrifuged for 10 minutes. The pellets were washed with cold 70% EtOH and vacuum dried. Twenty-five microliters of piperidine was added to each of the samples and they were heated at 90°C for 30 minutes. The reactions were dried under vacuum after which 10 μl ddH$_2$O was added. This step was repeated and the samples were dried under vacuum once more. The pellets were resuspended in 10 μl of sequencing stop buffer.

3. Footprinting

DNase I was used to digest unprotected regions of the end - labeled probes after incubation with nuclear extract from 24-hour sea urchin embryos. A typical reaction contained 4 µl 5X DNase I buffer (25 mM HEPES, pH 7.8, 12.5 mM KCl, 5 mM MgCl$_2$, 435 mM NaCl, 1 mM DTT), 2µl 1mg/ml poly [d(I-C)], 0.5 - 4 µl 15 mg/ml nuclear extract (or none for controls), 1-2 µl DNase I (0.05-1 µg), diluted in 10 mM HEPES, 2.5 mM CaCl$_2$) and ddH$_2$O to a final volume of 20 µl. The extract and poly [d(I-C)] incubated on ice 15 minutes in the presence of 5X DNase I buffer for nonspecific binding before addition of the probe. After an incubation of 15 minutes on ice with the probe, an appropriate dilution of DNase I was added. The samples were digested for 1-2 minutes and stopped by addition of 30 µl of DNase I stop buffer (0.6 M NaCl, 0.4% SDS, 20 mM EDTA, 3.3 mg/ml yeast tRNA, 100 µg/ml proteinase K). After a 30 minute incubation at 65°C, the samples were extracted with phenol, phenol/sevag, and sevag. Two volumes of cold 95% EtOH was added and the samples incubated at -80°C for 30 minutes before centrifugation for 15 minutes at full speed in a refrigerated microcentrifuge. The pellets were washed with 70% EtOH, vacuum dried and resuspended in 3 µl sequencing stop buffer. Samples were heated at 90°C for 4 minutes and

electrophoresed on an 8% polyacrylamide-urea gel alongside appropriate Maxam - Gilbert sequencing reactions serving as markers for each probe. The gel was exposed to X-ray film with an intensifying screen for 24 - 72 hours.

Computer Analysis of the Data

Sequencing

The base sequence of each of the subcloned fragments of pCyIIa-CAT.1 was determined directly from autoradiographs of the gels and the data was entered into the MacVector® sequence analysis program. Ultimately, the sequences of each of the fragments were integrated to yield one continuous sequence of the portion of upstream sequence of CyIIa contained on pCyIIa-CAT.1. The application "Restriction Enzyme," which was used to verify orientation and identity of the fragments of the upstream region for later integration, was also used to generate a complete restriction map of this region.

The sequences of previously characterized actin genes CyI, CyIIb, CyIIIa and CyIIIb were entered into MacVector and the application "Pustell

DNA Matrix" was used to compare their promoter regions to the portion of CyIIa upstream region that is contained on pCyIIa-CAT.1. Base sequence of the CyIIa 5' leader was also compared to pCyIIa-CAT.1 sequence.

Footprinting

The base sequence obtained for pCyIIa-CAT.1 was scanned by a program containing a database of all consensus binding sequences for known transcription factors/DNA binding proteins. Base sequence of regions along the DNA which were found to be resistant to/protected from DNase I digestion in footprinting studies were compared to data obtained from this search of the pCyIIa-CAT.1 sequence.

Results

Subcloning the CyIIa actin upstream region

The plasmid pCyIIa-CAT.1, contains approximately 4.3 kb of sequence 5' of the CyIIa start of translation plus 37 nucleotides of actin coding region ending at the junction with the CAT reporter gene coding region. The CyIIa actin coding sequence was originally obtained by isolation of the λ recombinant λSpG2-8 from a *Strongylocentrotus purpuratus* genomic library (Scheller et al., 1981). To create pCyIIa-CAT.1, an Eco RI fragment of λSpG2-8 was first subcloned into pBluescript. Sequences 3' of the Bam HI site were deleted by digestion with Bam HI (which also cut within the pBluescript polylinker). The approximately 1.6 kb pUC CAT Bgl II - Bam HI fragment was inserted into the Bam HI site to yield pCyIIa-CAT.1.

Subclones of the CyIIa upstream region were generated as shown in fig. 2. The plasmid pCyIIa-CAT.1 was restriction digested with Eco RI, Kpn I, and Bgl II and resulting fragments W (~1.6 kb), X (~1.8 kb), Y (~0.5 kb) and Z (~1.0 kb) were gel purified. These fragments were subsequently

cloned into the plasmid vector pUC18. The subclone pUCX was restriction mapped further to reveal suitable cloning sites. Plasmid pX was restriction digested with Pst I, generating fragments XI (~1.0 kb) and XII (0.8 kb), which were subcloned into pUC18. Subclones were verified by restriction analysis and agarose gel electrophoresis.

Sequence Analysis

The strategy for sequencing pCyIIa-CAT.1 is shown in figure 3. In the first round, plasmids containing the fragments W, X, Y, and Z were sequenced using pUC forward or pUC reverse primers provided in the Sequenase® kit, which primed on regions flanking the multiple cloning site on either side of the insert. A complete sequence was obtained for fragments Y and Z by using these primers and by varying the sequencing reactions to clarify sequences further away from the priming site and to facilitate detection of overlap between forward and reverse-primed reactions. Also, the sequencing reactions were modified to clarify sequences very close to the priming site for verification of the identity and orientation of each fragment. A complete sequence was also obtained for fragment XII using a primer made to the end of the data obtained in the first round of sequencing. Since

FIGURE 3: STRATEGY FOR SEQUENCING pCyIIa-CAT.1. The restriction map of pCyIIa-CAT.1 is shown for reference. Primers used are indicated with arrows. A primer made to the end of sequence obtained for the Pst I - Bgl II fragment generated complete data for that portion of pCyIIa-CAT.1. Another primer made to the "beginning" of sequence obtained for the Bgl II - Bgl II fragment verified its orientation on the original plasmid.
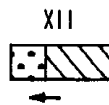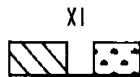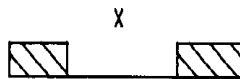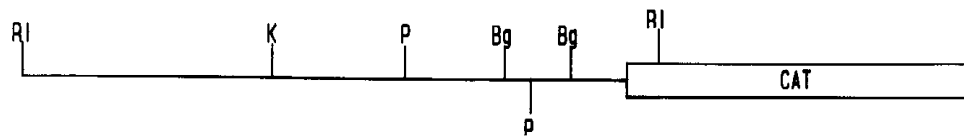
	= First round of sequencing using pUC primers
	= Second round of sequencing using internal primers
	= Completed sequence

1 kb

32

fragment Y utilized the same restriction site (Bgl II) in subcloning, another primer was made to the end of fragment Y sequence. When pCyIIa-CAT.1 was used as a template for sequencing with this primer, the resulting sequence data matched sequence already obtained for fragment XII and the correct orientation of the sequence data for fragment Y was determined. The MacVector® sequence analysis program application "Pustell DNA Matrix" was utilized to compare forward and reverse sequence data.

Computer Analysis

"Pustell DNA Matrix" yielded important information on the sequence when used to compare forward and reverse sequencing reaction data for detection of overlap. The data was ultimately compiled then into one continuous sequence of approximately 1.72 kb, shown in fig. 4.

The CyIIa upstream sequences were identified by comparing the sequence of the CyIIa 5' leader, isolated from a coelomocyte cDNA library (Smith, 1992), to the 1.72 kb genomic sequence using the Pustell DNA Matrix application. Homology was found in two locations along the sequence. The first 47 bp of the leader matched an area roughly halfway

FIGURE 4: SEQUENCE OF THE CYIIa UPSTREAM REGION. The sequence data obtained for the upstream region of CyIIa as it is represented on pCyIIa-CAT.1 is shown. Restriction enzyme sites used in subcloning are included, as well as homologies to the 5' leader sequence, distinguished by underlining. The start of translation (ATG) is shown in bold italics. The BamHI/BglII site, at which CAT reporter gene sequence was cloned in, is also indicated.

```
        Pst I
   1    CTGCAGCAAGTAAGACTACCAAATGCACACGATCATTAATTTCATCGCTTCCTCATTTTATCAATAGACATT

  73    CCTAATGTATTACAATGTATATGCCTTTACGAACGCCGAGCTATTTAATATACATTCACCTCTTATGAATAC

 145    TGTTATTTCAGTAGTCACAATGGGCATTCGTTAATCGATTGCTGCCCTACCTAAACAACAGTCATTCATGCA

 217    AATGAAATACAACTATTCGTGACAGTTGAATGAAATAATTTGTTACGATGCACTTTTAAACTTTACATAGGC

 289    TAGGCTTACGTATCAAATTTCAATATTACGAAAATGGACCATTTTCTGAATACGCAACTTTTATGGTCTCAA

 361    TGGAGTAACTCATACGGAACGAGGTGGTACATGGGCAAACGAACTTCCCTGTAATTGGGATTACAGTCCAGG

 433    GTAAATAGTCCGGAGTATCAACAGGAAGTAGGTCCTAGCACCCTGGTGCTTCAGAAGGAAGTCGCCGACAAA

 505    ATTTGTTATCTGCCACCGAAATGTTTTACACAGTGACTCAGGGTAAAAACAGGTCATGAATATTCATAAGTT

 577    TACTTTACTAAACTCGTACTGCCTTATTAGTGTAACACAAACACGTTTTATGTGTGTAAAGCACCTCGATGC

 649    ACGCTAAGGCCCGCCCCCGTTATACACCATGTACAAACAAAAGCGAAAATGACCATAAAAGGGAATGAGAAT
            Bgl II
 721    GTAAAAGATCTATCACGGAACTAGCATAAATATCGATGGTTCGAGCATGATCTGTCACTTGATTTGTCTGCC
            homologous to the 5' leader
 793    CCTTAGGATTACATAGGCTCCTGCTTTCCTTTATTCTTCTCAAGGTAAGTCGTGAAATACTTCGCTATACTA
                                                   Pst I
 865    CTTCAAATCAGAGTCAAAGGATCTTTTTATTATATTGGCAGTACTGCAGACAGTACAGACTGGTGGGCTTCA

 937    GGTTGGAGTTTAGTTAACTAGCTTTCGAAAGTCAAAGAACTAATTGTTTGAGTATTAATTACAGTCATATTT

1009    ACAATCTAATGGTTTAAAGTCGACGTGACCATGATACGACACTGTCATGGCGATTGCTTTTGTTACAGGAGG
                                                               Bgl II
1081    AAAGACTGTTGGCGTGAACGGTTGCAATTGAAATTCCATGCGTAGTTTCTGTTCGTAGATCCAGATCTACAA

1153    ATAAAATATCTTTTCTTATTAAGATAATGCAATCTTAATTGCCTCCGTGTTCTTTCGAATTGTCACTCATTC

1225    TTCAAATAAAGATTGTGAGATCACGCGTTTTCTGTACCCTACCCTACAAATACGTAGGACACCTGGGTATGT

1297    AGTGAACCTAAAGTTTATAAATGATGTTCTTGTTTGTCCATCAATTTAACCGGGAAAAAAATTTATCTGTCT
                                   homologous to the 5' leader
1369    AATATCATTATCTATTTTCACACTTTTAGATCAAACTAGATTAAACAAATCATCATGGTGACGACGATGTT
            BamHI/BglII
1441    GCCGCTCTTGTCGTCGACAACGGATCTAGCTTGGCGAGATTTTCAGGAGCTAAGGAAGCTAAAATGGAGAAA

1513    AAAATCACTGGATATACCACCGTTGATATATCCAATGGCATCGTAAAGAACATTTTGAGGCATTTCAGTCAG

1585    TTGCTCAATGTACCTATAACCAGACCGTTCAGCTGGATATTACGGCCTTTTTAAAGACCGTAAAGAAAAATA
                                                                Eco RI
1657    AGCACAAGTTTTATCCGGCCTTTATTCACATTCTTGCCCGCCTGATGAATGCTCATCCGGAATTC
```

35

between the first Bgl II and the second Pst I sites. The last 25 bp of the leader matched the area on the sequence, which was just 5' of the ATG, shown in fig. 4.

This data indicates that an intron of 562 bp is located within the CyIIa 5' leader sequence. The end of the cDNA homology is presumed to represent or be near the actual start of transcription, approximately 633 bp from the ATG.

In light of the observed amino acid sequence homologies among CyI, CyIIb and CyIIa (Durica et al., 1988; Schott et al., 1984), the MacVector® software package was also used for comparative analyses of DNA sequence obtained for pCyIIa-CAT.1 with CyI and CyIIb upstream sequence. Only limited similarities were found between CyIIa and CyI/CyIIb (see figs. 5 through 7), though CyI and CyIIb show significant homologies to each other. Notably, when CyIIa was examined for protein binding sequences, a serum response element (SRE) and some homology to the protein binding site D of CyI were found along a region in which CyI and CyIIb share extensive homology and SRE sites which are identical to each other. The CyIIa SRE,

FIGURE 5: SEQUENCE COMPARISON OF CYIIA WITH CYIIB AND CYI. The sequence of CyIIa is shown. Homologous regions of CyI appear beneath the appropriate locations on CyIIa in bold type. Homologous region of CyIIb appears beneath an appropriate location in italics. Mismatched bases of the CyI and CyIIb sequences are in lower case letters. The start of translation (ATG) is shown in bold italics.

```
   1   CTGCAGCAAGTAAGACTACCAAATGCACACGATCATTAATTTCATCGCTTCCTCATTTTATCAATAGACATT

  73   CCTAATGTATTACAATGTATATGCCTTTACGAACGCCGAGCTATTTAATATACATTCACCTCTTATGAATAC
       CaTcATtTAgcAtAATGTATATaaCTTgAgG

 145   TGTTATTTCAGTAGTCACAATGGGCATTCGTTAATCGATTGCTGCCCTACCTAAACAACAGTCATTCATGCA

 217   AATGAAATACAACTATTCGTGACAGTTGAATGAAATAATTTGTTACGATGCACTTTTAAACTTTACATAGGC

 289   TAGGCTTACGTATCAAATTTCAATATTACGAAAATGGACCATTTTCTGAATACGCAACTTTTATGGTCTCAA

 361   TGGAGTAACTCATACGGAACGAGGTGGTACATGGGCAAACGAACTTCCCTGTAATTGGGATTACAGTCCAGG

 433   GTAAATAGTCCGGAGTATCAACAGGAAGTAGGTCCTAGCACCCTGGTGCTTCAGAAGGAAGTCGCCGACAAA

 505   ATTTGTTATCTGCCACCGAAATGTTTTACACAGTGACTCAGGGTAAAAACAGGTCATGAATATTCATAAGTT

 577   TACTTTACTAAACTCGTACTGCCTTATTAGTGTAACACAAACACGTTTTATGTGTGTAAAGCACCTCGATGC

 649   ACGCTAAGGCCCGCCCCCGTTATACACCATGTACAAACAAAAGCGAAAATGACCATAAAAGGGAATGAGAAT
                                           AAAATagCCAaAAAAGGGAtaaAaAga

 721   GTAAAAGATCTATCACGGAACTAGCATAAATATCGATGGTTCGAGCATGATCTGTCACTTGATTTGTCTGCC
       Ga

 793   CCTTAGGATTACATAGGCTCCTGCTTTCCTTTATTCTTCTCAAGGTAAGTCGTGAAATACTTCGCTATACTA

 865   CTTCAAATCAGAGTCAAAGGATCTTTTTATTATATTGGCAGTACTGCAGACAGTACAGACTGGTGGGCTTCA

 937   GGTTGGAGTTTAGTTAACTAGCTTTCGAAAGTCAAAGAACTAATTGTTTGAGTATTAATTACAGTCATATTT

1009   ACAATCTAATGGTTTAAAGTCGACGTGACCATGATACGACACTGTCATGGCGATTGCTTTTGTTACAGGAGG

1081   AAAGACTGTTGGCGTGAACGGTTGCAATTGAAATTCCATGCGTAGTTTCTGTTCGTAGATCCAGATCTACAA

1153   ATAAAATATCTTTTCTTATTAAGATAATGCAATCTTAATTGCCTCCGTGTTCTTTCGAATTGTCACTCATTC
       TgTCTTaTCTTAgTAAGATAATatcAtgtTTAA

1225   TTCAAATAAAGATTGTGAGATCACGCGTTTTCTGTACCCTACCCTACAAATACGTAGGACACCTGGGTATGT

1297   AGTGAACCTAAAGTTTATAAATGATGTTCTTGTTTGTCCATCAATTTAACCGGGAAAAAAATTTATCTGTCT
                                            AtggTgAgCGGGAAAAAAAgaaATCTGaC

1369   AATATCATTATCTATTTTCACACTTTTAGATCAAACTAGATTAAACAAATCATCATGTGTGACGACGATGTT

1441   GCCGCTCTTGTCGTCGACAACGGATCTAGCTTGGCGAGATTTTCAGGAGCTAAGGAAGCTAAAATGGAGAAA

1513   AAAATCACTGGATATACCACCGTTGATATATCCAATGGCATCGTAAAGAACATTTTGAGGCATTTCAGTCAG

1585   TTGCTCAATGTACCTATAACCAGACCGTTCAGCTGGATATTACGGCCTTTTTAAAGACCGTAAAGAAAAATA

1657   AGCACAAGTTTTATCCGGCCTTTATTCACATTCTTGCCCGCCTGATGAATGCTCATCCGGAATTC
```

38

FIGURE 6: SEQUENCE COMPARISON OF CYI WITH CYIIA AND CYIIB. The sequence of CyI is shown. Homologous regions of CyIIa appear beneath the appropriate locations on CyI in bold type. Homologous regions of CyIIb appear beneath appropriate locations on CyI in italics. Mismatched bases of CyIIa and CyIIb sequences are in lower case letters. (SRE is underlined.)

```
-953  TTTCAAGTTACTGTTATAGATGTTCATCCATGATCTATCATCATTTAGCATAATGTATATAACTTGAGGTATA
                                      -CcTaATgTAttAcAATGTATATgoCTTtAcG-

-880  AACAATTATGTTAACTTTATTTTTTATTATTTTTATTGTACATTTATATAATCATTCCTGTTTTTATGCAAAT
                      -TAcCTcgATgaTaTaTTATAaTTTtTTGTA-

-807  ATTATGTAAACATCGTATGATTGGTCACTGGACTACTGATTGTACATTGTGTTGGTATTAATAAACCGAAATT

-734  GAATTGAATTGAATTGAATTGAAATGTGCACATTTTATGCTGTGTATCATTCAGCTTGAAAAAAATTGTGTTT
                -ATTGAAaacAAgaGAATTGtAATGTaC-

-661  GTTTATTATGATAGATATGCCTATAATCCTGAATACCTTAGATCATCTTGCTATTGCTCTACATCTGNAGGTA

-588  TAGCAGAATTTCTGAAAAATACATCTACATTTTGGCTATTTTCCTATAACTTCTCTGCAAGGATTTCGCATAT

-515  ATAGGAAACACATAGCATGAGAATTTCGTTTTGCAATTCCTTTCTTACATTTTAGCATAGAATTTGAAATCAA

-442  ATGGGTACTGCAGCAAAAGCATCACATCATCCATGTTACATTTATACTAGTTATATTCCATGCGAAACCCATG

-369  GAGAGAAGTTAAAGGCAAAATAGCCAAAAAAGGGATAAAAAGAGAAGAAATGGGGTTTGTAAAGTGCTTGATC
                      -AAAATgaCCAtAAAAGGGAatgAgAatG-

-296  TATATGTAATAATTCGTAAACACTCTGCTATAAGGAACTGAGCTCATATCATCATGGAGAGAGCATGCCTCAC
                                      -GAGAGatctCCTCAC-

-223  GTGTATTTGTGACCTTTAAAATCAGCCACTTGTGCGAATCAGAACCTGCGTGGAGTAAAACAAAAAAACGCCGG
      GTGTATTcGTGtaC-                        -CaaGaAaaAAAAaAAAAAAACGCtcG

-150  GTGGTTTTCTGAGGCGATCCCCGATAAGGATATTAATGAATGCGCCGAGGCATACGATGCGTCTAAATTTCTC
      GgtGgTaTtcGAGGCGATCCCCtATAAGGATATTAATGAgTG-  -CGAGGCgTgCGgTGCGTCTAAAaTTCTC

-77   CTTATATGGTCAAATATGTTTCACGTACAACGTCCGGTCATTATTATAATACCTGCTTGCTCCTGCTCCTCAG
      CTTATATGGTCAAatatgccTTCt-                        -TGCgCCTtCatCTCAG

-4    CCTCATTCCAACCTCAACCAAGCAAACTTGCAGGTGTGATCATTCTCTCTCTCTTTCGGTAAGTCTTTAAGAA
      CtTCATTCaAgttTCAAC-                        -TcCTCTCTtTCTTTCGGTAAGTCTaTAtttc

+69   TTTCCTACGATTCTCCTACGATTTCCTTAGCTCSTCTACAGGATCATAGCTCCATAGCAGCGATTCGCATGCA
      TT-

+142  AAGCTAAGGGCAGTTTGAAATCAGCTGCAGTCAGTGGTTGTTCATATTAGGATATTGTACTTCGATCGTTACT

+215  GAGTCTCTGTGATGCTTTTTTCGACCCGAGTAAATATCGTCTCTAAATTATCTTCATTTTATTAATCTGTATG

+288  TATTTTGTAATTGATGTATGATTGTATGATTGTAACTCTATGTCTTATCTTAGTAAGATAATATATGTTTAAG
                                      -TaTCTTtTCTTAtTAAGATAATgcAatcTTAA-

+361  TTGTAGTTACTTGCTAATTGAATGCTTATCGAGTATTGCTGAT
```

40

FIGURE 7: SEQUENCE COMPARISON OF CYIIB WITH CYI AND CYIIA. The sequence of CyIIb is shown. Homologous regions of CyI appear beneath the appropriate locations on CyIIb in bold type. Homologous region of CyIIa appears beneath an appropriate location in italics. Mismatched bases of the CyI and CyIIa sequences are in lower case letters.

```
-1161   CGCAAACCACCTGGTGAAATGAGTCAGCGCTCGCCAATAAGCGGTAATGGAAAGCTAGCTTTTTCTTCTTCCG

-1088   ACTCGCATGTTGACATGTGGTATCAACGATAACTCTTTTCACAATCGATCTGATTTCCTAAAAAAAATACCCAA

-1015   ATCATCTTACAAATTTCATTGCTTTGTGGTTGAATATTACGTACATTACAATTCTCTTGTTTTCAATGTTTTA
                                          GcACATTtCAATTCaaTTcaaTTCAAT

 -942   ACTCTGAATTAATGCAAAAGCGATACCATGATGAGGTGAAAATTATTTTGCATTTTATCAGGGGCTGATATAA

 -869   ATATCCAGAAGGGAATGGTGAGCGGGAAAAAAAGAAATCTGACATTAAGCAAAGCGACATGATTTTTCTAGAG
                AattTaAcCGGGAAAAAAAAtttATCTGtC

 -796   AACGCGGTTCCTTTTACAAATACAAACATGCTTTGAAATTATGCCTGAAAAGAAATGCTCTTTAAATTATTCG

 -723   ACTTCTGATAAGAAATCGCTGAAAATTAGGGCCGGACAAAACAATTATGTGTGTATAGTCCCATATACGGATT

 -650   TCTTTGGCAGTTCATGTTATTTTTTTAGCAAAGTAAGGTCAATTTTTTTGGGGGAAAAAACTCATTCCAATTT

 -577   TCTCAATCAGTATTAATTAGGCGCTTCTTAACTACTATATAATCAATCTTTCCAAAAAGTAATTAACATTTAA

 -504   AGAAATTTTTGAGTCTTGTAAAGATTTACAAAAAATATAATATATCATCGAGGTATTTGAGCCTTTTCATCCT
                                         TACAAtAAAaATAATAaAaaATaaAGtTA

 -431   TCTTTTTCGTGTCGGTGTTATTAGTTTCAGAAGAATATAATTGTATATCAAACAGAAAAAATTGAACAGAATA

 -358   GATAGCAAAAGAAATTAGGTATGTAGAGTAGTANNCTGAACTGTATAAGAACCTGTATATAGGAAGTGAGCCC

 -285   TTGATTATACATATATGGTATGGAGAGATCTCCTCACGTGTATTCGTGTACAGCCAACTTGTGTGAAATCAAG
                 TggaGAGAGcatgCCTCACGTGTATTtGTGacC

 -212   CAGAAACAAAATGCATGCGCCTGCGTGTGGGTGGAGCAAGAAAAAAAAAAAAAAAAAACGCTCGGGTGGTTATAT
                                         CgtGgAgtAAAAcAAAAAAACGCcgGGtgGtT        T

 -139   TCGAGGCGATCCCCTATAAGGATATTAATGAGTGGCGAGGCGTGCGGTGCGTCTAAAATTCTCCTTATATGGT
            ctGAGGCGATCCCCgATAAGGATATTAATGAaTG CGAGGCaTaCGaTGCGTCTAAAtTTCTCCTTATATGGT

 -66    CAATATGCCTTCTCGTACAACTGTGGATAGTAGATATACCCTGTGCGCCTTCATCTCAGCTTCATTCAAGTTT
            CAA                                TGCtCCTgCtcCTCAGCcTCATTCcAaccT

  +7    CAACCTCACATCTCCTCTCTTTCTTTCGGTAAGTCTATATTTCTTGAATTC
            CAAC        .       TtCTCTCTGTCTTTCGGTAAGTCTtTAagaaTT
```

42

although consensus, differs at three nucleotides from that of CyI/CyIIb.

Footprinting

A 912 bp PstI-PstI fragment of pCyIIa-CAT.1 was gel purified and used for generating footprinting data. The purified fragment was digested with ClaI and end-labeled using ^{32}P-dCTP. Following labeling the fragment was digested with AvaII. Four footprinting probes were thus generated (fig.8). These probes encompassed 774 bp of sequence including 73 bp downstream of the splice junction and 910 bp of sequence 5' of the start of transcription (minus the 138 bp AvaII-AvaII fragment that was excluded from the footprints). Each of the probes was allowed to bind proteins present in a crude nuclear extract from *S. purpuratus* embryos (prepared at 24 hours post-fertilization). After footprinting, as described in the Materials and Methods, certain regions along the DNA probes were found to be protected from digestion with DNase I by bound protein. Also identified were sites of enhanced Dnase I sensitivity, which are marked with arrows and appear in the footprint as bands with increased intensity in comparison to the control (fig. 9). Analysis of the protected sequences with an online computerized database of DNA-binding protein recognition sites (Altschul et al., 1990)

FIGURE 8: MAP OF FOOTPRINTING CONSTRUCTS: The 912 base pair region of pCyIIa-CAT.1 chosen for footprinting studies is shown. A restriction map of this region as well as the sizes, locations, and designations of the restriction fragments used for footprinting appear below the map of the original plasmid. An asterisk indicates the radiolabeled end of each probe. C = Cla I; A = Ava II; D = Dde I.
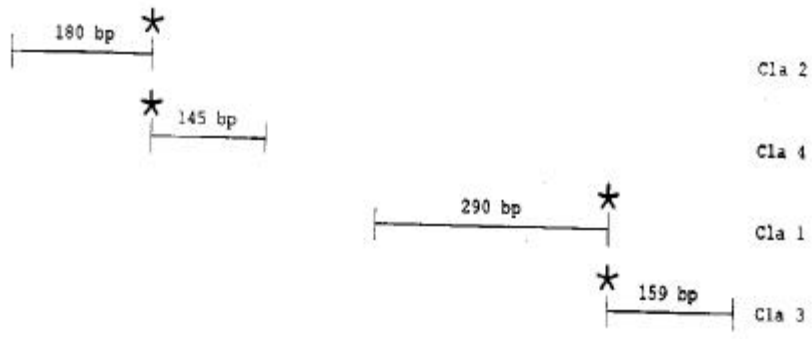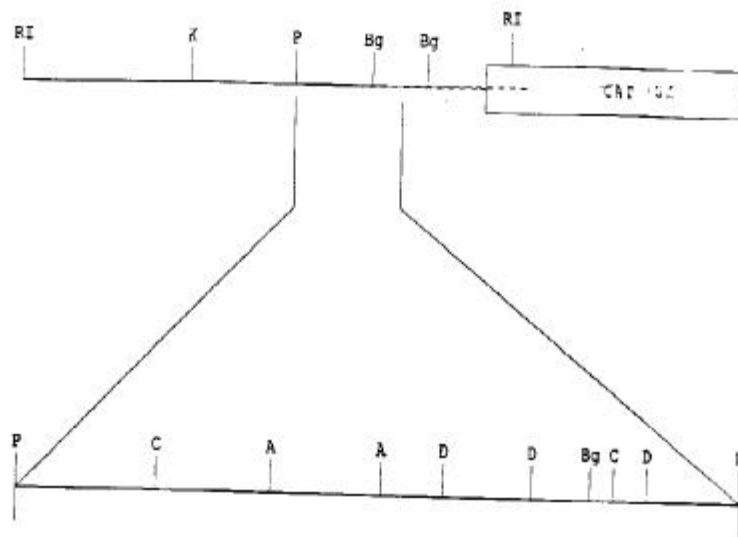
FIGURE 9: DNase I FOOTPRINT ANALYSIS OF CyIIa ACTIN UPSTREAM REGION USING CLA 1 PROBE. CyIIa DNA was subjected to digestion by DNase I without the addition of nuclear protein extract prepared from 24 hour S. purpuratus embryos as controls, or with the indicated amount of protein mixture. All were incubated with poly [d(I-C)] prior to addition of nuclear extract for nonspecific binding. The amount of DNAse I, time of digestion and amount of nuclear protein extract used for each assay is indicated above each lane. The boxed regions indicate sites of protein binding. Arrows at left indicate regions of hypersensitivity.

47
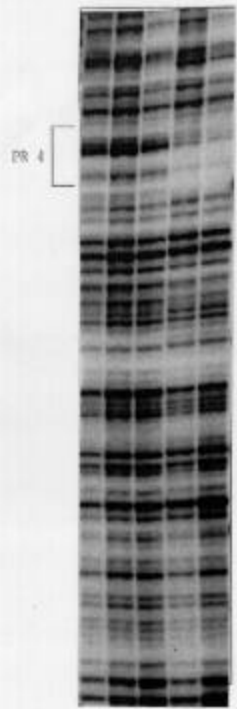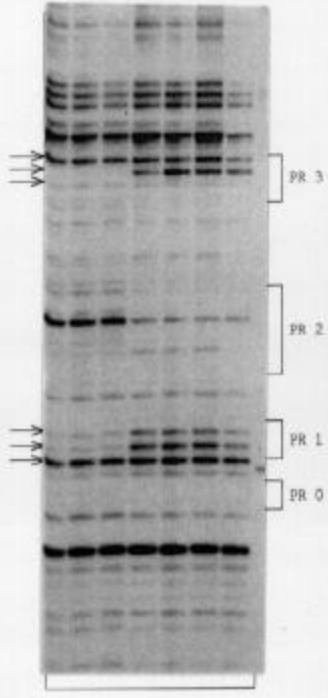
revealed some homologies to well-characterized transcription factors (Table 1).

Cla 2, a 180 bp PstI-ClaI fragment, represents the beginning of the larger (912 bp) PstI-PstI restriction fragment for which continuous sequence had been obtained. Two protected regions, one of which contains a TATA box, were detected with Cla 2 (fig. 10). Cla 4, 145 bp in length, is a Cla I - Ava II fragment immediately adjacent to Cla 2. It was found to contain an octamer consensus sequence and an additional protected region (fig. 11). The 290 bp Cla 1 sequence, delineated by an Ava II site and a Cla I site, covers the ligation junction and may include the start of transcription. Among the four regions of protein binding along this probe (fig. 9) is the consensus sequence for the serum response element (SRE). The last of the four probes, Cla 3, is a 159 bp Cla I - Pst I fragment representing the 3' boundary of continuous sequence data obtained. Also showing four protected areas, it includes a CCAAT box (fig. 12).

FIGURE 10: DNaseI FOOTPRINT ANALYSIS OF CyIIa ACTIN UPSTREAM REGION USING CLA 2 PROBE. CyIIa DNA was subjected to digestion by DNase I without the addition of nuclear protein extract prepared from 24 hour S. purpuratus embryos as controls, or with the indicated amount of protein mixture. All were incubated with poly [d(I-C)] prior to addition of nuclear extract for nonspecific binding. The amount of DNase I, time of digestion and amount of nuclear protein extract used for each assay is indicated above each lane. The boxed regions indicate sites of protein binding.
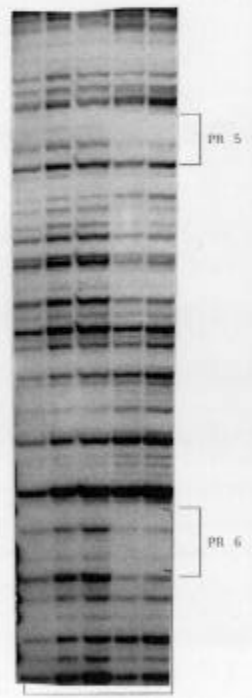
FIGURE 11: DNase I FOOTPRINT ANALYSIS OF CyIIa ACTIN UPSTREAM REGION USING CLA 4 PROBE. CyIIa DNA was subjected to digestion by DNase I without the addition of nuclear protein extract prepared from 24 hour S. purpuratus embryos as controls, or with the indicated amount of protein mixture. All were incubated with poly [d(I-C)] prior to addition of nuclear extract for nonspecific binding. The amount of DNase I, time of digestion and amount of nuclear protein extract used for each assay is indicated above each lane. The boxed regions indicate sites of protein binding.

TIME (MIN)   1.5   1   1   1   1

DNase I (ul)   1   1   1   1   1
                 1   1 (1:1) 1   1

Extract (ug)   --   --   -- 10 20
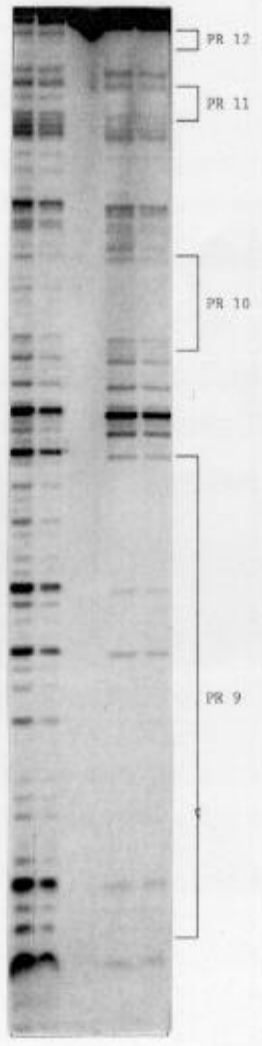
PR 12

PR 11

PR 10

PR 9

52

FIGURE 12: DNaseI FOOTPRINT ANALYSIS OF CyIIa ACTIN UPSTREAM REGION USING CLA 3 PROBE. CyIIa DNA was subjected to digestion by DNase I without the addition of nuclear protein extract prepared from 24 hour S. purpuratus embryos as controls, or with the indicated amount of protein mixture. All were incubated with poly [d(I-C)] prior to addition of nuclear extract for nonspecific binding. The amount of DNase I, time of digestion and amount of nuclear protein extract used for each assay is indicated above each lane. The boxed regions indicate sites of protein binding.

TIME (MIN)	1.5	1	1	1	2	2	2
DNase I (ul)	1	1	0.5	1	1	1	2
Extract (ug)	--	--	--	10	30	60	60
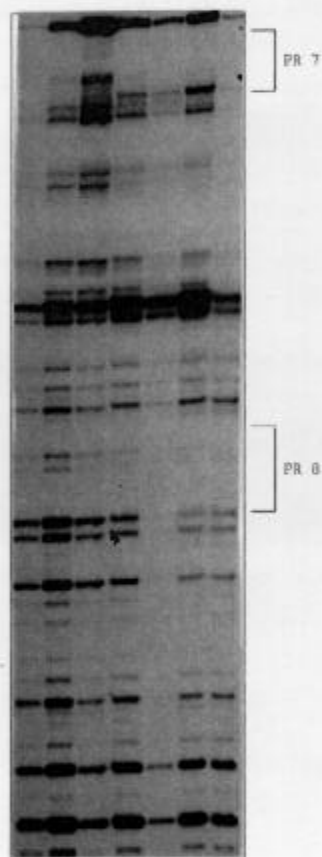
Discussion

That the protein products of CyI, CyIIa, and CyIIb show remarkable amino acid sequence homology to one another (Durica et al., 1988; Schott et al., 1984) lends support to the genes having arisen from a common progenitor. That their patterns of expression vary more widely is at least an indication that their regulatory regions have evolved separately. The precise mechanism by which they reached their present forms was explored in this study. By sequence comparisons of introns and 3' and 5' flanking untranslated regions (Lee et al., 1984; Akhurst et al., 1987; Durica et al., 1988; Fang and Brandhorst, 1994), the CyI-CyIIa-CyIIb genes are more closely related to one another than to the unlinked CyIIIa-CyIIIb genes. The finding that the 3' untranslated regions of CyIIa and CyIIb are homologous, while the 5' untranslated regions of CyIIb are nearly identical to the same region on CyI leads to the additional conclusion that the original gene cluster arose from two gene duplications (CyI to CyII and CyII to CyIIa and CyIIb') followed by a gene conversion between the 5' regions of CyI and CyIIb'(giving rise to CyIIb) (Durica et al., 1988).

Analysis in this study of promoter sequences reveal that CyIIa shows only limited homology to CyI and CyIIb (figs. 5, 6, 7), while Lee et al. (1984) found intron sequences to be similar enough to cross hybridize. This is perhaps an indicator of some selection pressure on the production/function of product, though not on the timing and degree of expression.

Like all the *S. purpuratus* cytoplasmic actin genes (Akhurst et al., 1987; Katula et al., 1987; Flytzanis, 1989), the 5' leader sequence of CyIIa was found to have an intron within it. Like both CyI and CyIIb, CyIIa promoter sequences were found to contain the consensus binding site for the serum response element (SRE) as well as an octomer motif, TATA box and CCAAT box (table 1). The CyIIa octomer sequence is the same as that found in both CyI and CyIIb. It should be noted here that though the SRE found in CyIIa is the consensus, it differs in the A/T region by 3 bp from those of CyI and CyIIb, which are identical to each other.

Footprinting analysis of the CyIIa promoter sequence obtained in this project was used to determine if features apparent in the DNA sequence data actually bound protein. Twelve areas on the four probes generated were protected from digestion by DNase I by protein present in a nuclear extract

TABLE 1: Sequences and Features of Protected Regions. For each of those regions which were shown to be protected from DNaseI digestion in footprinting studies, the probe used, location of the protected region, its length in base pairs, and its sequence is given along with a listing of putative protein binding sites and other notable features (underlined unless otherwise noted).

Table 1: Sequences and Features of Protected Regions

protected region	probe	location	sequence	features
PR1	Cla 1	711-715	GAATG	DNaseI hypersensitivity (bold),
PR2	Cla 1	701-710	CCATAAAAGG	serum response element
PR3	Cla 1	682-687	CAAACA	
PR4	Cla 1	559-580	CATGAATATTCAT-AAGTTTACT	
PR5	Cla 2	114-123	TATTTAATAT	TATA box
PR6	Cla 2	54-66	CATTTTATCAATA	
PR7	Cla 4	317-305	GTAATATTGAAAT	
PR8	Cla 4	226-218	GTATTTCAT	octamer motif
PR9	Cla 3	803-774	TAATCCTAAGGG-GCAGACAAATCA-AGTGAC	cDNA homology
PR10	Cla 3	828-816	GAATAAAGGAAAG	
PR11	Cla 3	873-863	GATTTGAAGTA	
PR12	Cla 3	902-893	CCAATATAAT	CCAAT box

from 24-hour *S. purpuratus* embryos (figs. 9 through 12). Though there was binding of protein to the SRE, it was not complete (fig. 9), as it is in both CyI and CyIIb. The octamer site also showed evidence of protein binding with additional protection of the areas flanking the octamer consensus (fig. 11). A large protected region was found to be centered around the putative start of transcription (fig. 12). The location and size of this footprint suggests that this may represent binding of the transcription initiation complex.

Though some sequence similarities were found between CyIIa and CyI/CyIIb, including the SRE, TATA box and some smaller regions within site D (Ganster et al., 1992), CyI and CyIIb share a much more similar set of protein binding sites within the promoter region. Site D is conserved in CyI and CyIIb and when deleted from a CyI-CAT reporter construct, *cat* expression was decreased by 10-20%, indicating that a positive regulatory sequence is contained within this region (H. Paul, R.J. Winfrey, K.S. Katula, unpublished). The finding by Collura and Katula (1992) was that at least 198 bp of upstream sequence, including the SRE and site D is required for correct spatial expression of CyI. That the area on CyIIa believed to be homologous to site D in CyI/CyIIb does not bind protein accentuates just one of the divergences within what would become CyIIa following gene duplication.

Over evolutionary time, protein binding to site D has been eliminated, while binding to the SRE has been modified. It is likely that these changes alone contributed significantly to the altered pattern of expression for CyIIa compared to CyI/CyIIb.

When more of the upstream region of CyIIa is analyzed, a more complete picture of the types of transcription factors that bind, at what levels, in what combinations and at what times can be assembled. The specific points at which these differ from CyI and CyIIb will more effectively retrace the genetic modifications that over time produce a perpetuated change.

References

Akhurst, R.J., Calzone, F.J., Lee, J.J., Britten, R.J. and Davidson, E.H., 1987. Structure and organization of the CyIII actin gene subfamily of the sea urchin, *Strongylocentrotus purpuratus*. **J. Mol. Biol.** 194: 193-203.

Altschul, S.F., Gish, W., Miller, W., Myers, E.W. and Lipman, D.J., 1990. Basic local alignment search tool. **J. Mol. Biol.** 215: 403-410.

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

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

Crain, W.R., Durica, D.S., and VanDoren, K., 1981. Actin gene expression in developing sea urchin embryos. **Molec. Cell Biol.** 1: 711-720.

Davidson, E.H., 1986. **Gene Activity in Early Development**. Academic Press.

Durica, D.S., Garza, D., Restrepo, M.A. and Hrynicewicz, M.M., 1988. DNA sequence analysis and structural relationships among the cytoskeletal actin genes of the sea urchin *Strongylocentrotus purpuratus*. **J. Mol. Evol.** 28: 72-86.

Fang, H. and Brandhorst, P., 1994. Evolution of actin gene families of sea urchins. **J. Mol. Evol.** 39: 347-356.

Flytzanis, C.N., Bogosian, E.A. and Niemeyer, C.C., 1989. Expression and structure of the CyIIIb actin gene of the sea urchin *Strongylocentrotus purpuratus*. **Mol. Reprod. Dev.** 1: 208-218.

Ganster, R., Paul, H. and Katula, K.S., 1992. Analysis of the DNA binding proteins interacting with specific upstream sequences of the *Strongylocentrotus purpuratus* CyI actin gene. **Mol. Reprod. Dev.** 33: 392-406.

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

Lee, J.J., Shott, R.J., Rose, S.J. III, Thomas, T.L., Britten, R.J. and Davidson, E.H., 1984. Sea urchin actin gene subtypes: gene number, linkage and evolution. **J. Mol. Evol.** 172: 149-176.

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

Maniatis, T., Sambrook, J. and Fritsch, E.F., 1989. **Molecular Cloning : a Laboratory Manual**, 2nd edition. Cold Spring Harbor Laboratory Press, New York.

Scheller, R.H., McAllister, L.B., Crain, W.R. Jr., Durica, D.S., Posakony, J.W., Thomas, T.L., Britten, R.J. and Davidson, E.H., 1981. Organization

and expression of multiple actin genes in the sea urchin. **Molecular Cell Biology**. 7: 609-628.

Schuler, M.A., McOsker, P. and Keller, E.B., 1983. DNA Sequence of Two Linked Actin Genes of Sea Urchin. **Mol. Cell Biol.** 3: 448-456.

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

Smith, L.C., Britten, R.J. and Davidson, E.H.: SpCoel1, 1992. A sea urchin profilin gene expressed specifically in coelomocytes in response to injury. **Mol. Biol. Cell.** 3: 403-414.

Zeller, Robert W., Cameron, Andrew R., Franks, Roberta R., Britten, Roy J. and Davidson, Eric H., 1992. Territorial expression of three different trans-genes in early sea urchin embryos detected by a whole-mount fluorescence procedure. **Dev. Biol.** 151: 382 - 390.