Transcriptional Control of Spatially Regulated Genes in the Early Sea Urchin Embryo

Thesis by

Robert Werner Zeller, Jr.

In Partial Fulfillment of the Requirements

for the Degree of

Doctor of Philosophy

California Institute of Technology

Pasadena, California

1995

(Submitted November 30, 1994)

This thesis is dedicated to my family, whose support has allowed me to pursue my interests in nature since a very early age.

Acknowledgments

First, I would like to thank Eric for allowing me the opportunity to explore sea urchin embryo development. I would also like to thank my committee members, Scott Fraser, Lee Hood, Howard Lipshitz and Paul Sternberg, whose time and encouragement was greatly appreciated. A special thanks goes to Scott, who brought me yogurt after a serious bout with *Salmonella* in Woods Hole and started me on the road to recovery.

I would also like to thank past and present members of the Davidson lab who contributed reagents, techniques and assistance over the years; especially Roger Anderson, Andy Cameron, Jim Coffman, Joe Minor and Andy Ransick. Frank Calzone, now at Amgen, instructed me in the fine arts of gel shift assays and massive embryo culture (about 10^{11} or so, give or take a few billion). I will never forget the site of 20 billion embryos filling numerous garbage cans or, for that matter, the **CLIFF** at Point Arenas. Jim Moore taught me the finer aspects of non-linear least squares data analysis and Jane Rigg provided assistance above and beyond the call of duty. I promise Stephanie Canada that I will never again give her so many manuscripts-in-preparation at the same time.

To my roommate Charles Spence, I owe a great deal of thanks. He provided an outstanding laser printer which produced this thesis. I believe the several days of computer network installation we invested was well worth the effort and learning experience. I promise to do my part in turning the living room back into a living room.

My family provided a tremendous amount of love, support and encouragement both in the best and worst of times. Without their help, the job of completing this thesis would have been much more difficult.

Finally, I would like to thank Christina Niemeyer for her understanding. A couple of months turned into six or eight, and her support helped to make this time go faster. She has reminded me that there is more to life than lab work.

Abstract

The regulatory domains of several *Strongylocentrotus purpuratus* embryo genes appear to be organized into "modules" of closely associated target sites for DNA-binding proteins. The Sp(G/C)F1 protein is a novel DNA-binding protein which binds specifically to G/C rich *cis*-elements. The protein was purified from embryo nuclear extracts by DNA affinity chromatography and the corresponding cDNA isolated. Five different Sp(G/C)F1 polypeptides, produced from a nested set of AUG initiation codons, are synthesized from a single mRNA template. Each protein shares a common C-terminus and a centrally located DNA-binding domain but incorporates variable amounts of an N-terminal proline-rich region. Since proline-rich regions often serve as transcriptional activation domains, the five Sp(G/C)F1 proteins are likely to possess different transcriptional "activation potentials." Sp(G/C)F1 proteins bind DNA cooperatively, most likely as homodimers, and multiple protein-DNA complexes are formed at high protein to DNA ratios *in vitro*. When bound to two or more target sites in the *CyIIIa* regulatory domain, the Sp(G/C)F1 polypeptides form protein-protein contacts, and "loop out" intervening regions of DNA which could allow the interaction of distant regulatory modules. The Sp(G/C)F1 protein is thus likely to serve as an intermodule communicator in the regulatory domains of many sea urchin embryo genes.

Antibodies were used to quantitate the nuclear concentrations of two different DNA-binding proteins, P3A1 and P3A2, which bind specifically to P3A *cis*-elements found in several different regulatory domains. Both proteins are present maternally, but by late gastrula, P3A1 protein has disappeared from the nucleus leaving P3A2 protein levels relatively unchanged at about 10^4 molecules per nucleus. Since the relative binding affinities of P3A1 protein for different DNA target sites are up to 50X lower than the affinities of P3A2 protein for the same sites, P3A1 protein is therefore likely to be of

functional significance only in early to mid-cleavage before genes known to contain P3A target sites, such as *CyIIIa*, *SM50* and *Spec1a*, are expressed. The P3A1 and P3A2 proteins can thus serve as an antagonistic regulatory switch at P3A sites in cleavage stage sea urchin embryos.

Table of Contents

Chapter 1

INTRODUCTION

Purification of Sea Urchin DNA-Binding Proteins and Cloning of the Corresponding Genes

Gene expression is, in general, a function of transcriptional regulation, which results from the combinatorial binding and interaction of sequence-specific DNA binding proteins with the regulatory region of a gene. A number of reviews have been written (e.g., Ptashne, 1988; Mitchell and Tjian, 1989; Clark and Doherty, 1993; Kingston and Green, 1994) which describe in detail how transcription factors bind to DNA and either activate or repress gene expression. In an embryo, a gene is regulated both spatially and temporally, determining whether or not transcription should occur and at what levels. A thorough examination of the molecular mechanisms governing gene expression will help us to understand a fundamental question in developmental biology: how a complex, multicellular organism arises from a single, fertilized egg.

The sea urchin embryo is an excellent system in which to examine the regulatory control of gene expression by the interaction of *cis*-regulatory elements and DNA-binding proteins. The regulatory domains for several spatially regulated genes have been defined, and target sites for DNA-protein interactions have been localized to specific DNA sequences within these regions. A number of proteins which bind specifically to these sites have been purified and the corresponding cDNAs isolated. This thesis discusses the roles of three different DNA-binding proteins in the transcriptional control of spatially regulated sea urchin genes. These DNA-binding proteins were isolated by their specific affinities for *cis*-regulatory target sites present in the promoter of the CyIIIa cytoskeletal actin gene. This introductory chapter will describe the identification of *cis*-regulatory target sites in sea urchin promoters and discuss the advantages of using a biochemical approach to isolate DNA-binding proteins which regulate genes expressed in the embryo.

Territorial markers of the sea urchin embryo

The sea urchin embryo develops by an invariant, canonical cleavage pattern (Cameron *et al.*, 1987; Cameron and Davidson, 1992). The same regions of egg cytoplasm are inherited by cells occupying similar spatial positions in all embryos. By the end of cleavage, five territories have been established - the aboral ectoderm, the oral ectoderm, the vegetal plate, the skeletogenic mesenchyme and the small micromeres. Each territory is defined by its cell lineage, a unique pattern of gene expression, and the presence of one or more distinct cell types (Davidson, 1989, 1990). As cleavage proceeds, the fate of the blastomeres becomes progressively restricted until a given blastomere will only produce cell types found in a single territory. At this point, the blastomere is said to become a "founder" cell for that territory. A territory is therefore composed of the descendants of a series of different founder cells within which the molecular interactions which regulate territory-specific gene expression are established (Cameron and Davidson, 1992). The key to understanding the establishment of territory-specific patterns of gene expression at a molecular level lies in the interactions between specific DNA-binding proteins and *cis*-regulatory sequences controlling genes expressed in the early embryo.

A number of genes which are expressed in a territory-specific pattern have now been identified and are listed in Table 1 (Coffman and Davidson, 1992). By examining the regulation of genes expressed in various spatial and temporal patterns, we can begin to elucidate how regions of differential gene expression are initiated in the early embryo. In *Strongylocentrotus purpuratus*, genes expressed exclusively in the aboral ectoderm include the cytoskeletal actin genes CyIIIa and CyIIIb (Cox *et al.*, 1986) as well as the genes encoding the calcium binding proteins Spec1 and Spec2 (Lynn *et al.*, 1983; Hardin *et al.*, 1988). Two other genes are expressed in this territory - the gene encoding

arylsulfatase (Yang *et al.*, 1989) and a metallothionein gene (Nemer *et al.*, 1991). Several genes have been characterized from the skeletogenic mesenchyme territory. Among these are the genes encoding the spicule matrix proteins SM30 (George *et al.*, 1991) and SM50 (Benson *et al.*, 1987), and the gene encoding the cell surface glycoprotein MSP130 (Parr *et al.*, 1989; Harkey *et al.*, 1992). The spicule matrix proteins form the organic matrix of the embryonic skeleton, and MSP130 is involved in calcium uptake (Farach *et al.*, 1987). The Endo16 gene encodes an extracellular protein initially expressed in the vegetal plate; later its expression is restricted to the midgut (Nocente-McGrath *et al.*, 1989; Ransick *et al.*, 1993). Some genes are expressed in multiple territories such as the cytoskeletal actins CyI and CyIIb which are initially expressed throughout the embryo and are later restricted to the oral ectoderm and vegetal territories (Cox *et al.*, 1986).

Identification of sea urchin DNA-binding protein target sites in cis*-regulatory domains*

The identification of *cis*-regulatory regions depends on the ability of embryonic cells to express exogenous reporter genes. McMahon *et al.* (1985) and Flytzanis *et al.* (1985) developed a technique for microinjecting solutions containing reporter genes into sea urchin eggs or zygotes. Reporter genes are composed of fragments of DNA containing *cis*-regulatory elements linked to DNA sequences encoding a protein which is not normally expressed in the sea urchin embryo and is easily detected. The occupation of *cis*-regulatory sites by *trans*-acting factors activates the reporter gene resulting in the production of the foreign protein. Quantitation of enzyme activity from reporter genes expressing chloramphenicol acetyl transferase (CAT) or luciferase may be used to examine the level of *trans*-gene expression in the embryo. I*n-situ* hybridization examination of *trans*-gene mRNA or the enzymatic detection of β-galactosidase reporter genes reveals the spatial patterns of *trans*-gene expression. By comparing *trans*-gene

expression to endogenous gene expression, it is possible to determine the promoter sequences which are necessary and sufficient for proper temporal and spatial regulatory control.

The regulatory regions of genes expressed in three different embryonic territories have now been characterized using the gene transfer methods described above. Zygotic injections of *CyIIIa-CAT* or *CyIIIb-CAT* have identified the regulatory domains for two different cytoskeletal actin genes, CyIIIa and CyIIIb, expressed exclusively in the aboral ectoderm territory. The CyIIIa regulatory domain occupies 2.3 kb of sequence upstream from the start of transcription and contains regulatory elements necessary and sufficient for proper temporal and spatial gene expression (Flytzanis *et al.*, 1987; Hough-Evans *et al.*, 1987, 1988; Zeller *et al.*, 1992, Appendix C). A 2.2 kb region is sufficient for correct temporal expression of the *CyIIIb-CAT* gene (Niemeyer and Flytzanis, 1993). Spec *trans*-genes employing a CAT reporter, containing ~5.6 kb, 1.5 kb and 5.6 kb of upstream region from the Spec1, Spec2a and Spec2c genes respectively, were found to have proper temporal expression after microinjection into eggs (Gan *et al.*, 1990b). When these same promoter regions were fused to β-galactosidase and injected into eggs, proper spatial expression was only observed for the Spec2a-*lacZ* construct (Gan *et al.*, 1990a) suggesting that Spec1 and Spec2c *trans*-genes are missing *cis*-regulatory elements which control proper aboral ectoderm expression. At any rate, these *trans*-gene experiments have identified several regulatory domains which contain the necessary *cis*-elements for aboral ectoderm expression.

In the early embryo, Endo16 is first expressed in the vegetal plate, then in the gut, and later restricted to only the mid-gut (Ransick *et al.*, 1993). A 2.3 kb region of the Endo16 promoter was found to confer proper temporal and spatial expression when fused to a CAT reporter (Yuh *et al.*, 1994). As we will see, the regulatory domain of the

Endo16 gene contains a large number of DNA-binding protein target sites which most likely control the complicated temporal/spatial expression of this gene. The regulatory domains of two spicule matrix protein genes have also been identified. An *SM50-CAT* reporter gene containing about 450 bp of regulatory region is properly expressed in the sea urchin embryo (Sucov *et al.*, 1988). A 2.6 kb of SM30 upstream region directs proper mesenchyme-specific expression of a CAT reporter (Akasaka *et al.*, 1994).

The identification of sequences which are specifically recognized by DNA proteins is relatively straightforward. Here two different territory-specific genes serve as examples, CyIIIa and Endo16. The procedure used to localize DNA-binding protein target sites is diagrammed in Figure 1 (Thézé *et al.*, 1990; Yuh *et al.*, 1994). About 20 sites of specific DNA-protein interactions are found in the CyIIIa regulatory domain (Thézé *et al.*, 1990; Fig. 1A), and 38 sites in the Endo16 promoter (Yuh *et al.*, 1994; Fig 1B). The basic strategy is to subdivide the regulatory domain into small fragments, 300-400 bp, which are then used as probes and combined with nuclear extracts and analyzed in gel mobility shift assays (Fried and Crothers, 1981; Calzone *et al.*, 1988; Thézé *et al.*, 1990; Yuh *et al.*, 1994). When fragments containing regulatory target sites are bound by a DNA-binding protein, the migration of the DNA-protein complex is retarded relative to uncomplexed probe. Positive fragments are then further subdivided and tested in the same fashion; eventually regulatory target sites are narrowed down to a 20-40 bp region. Further examination of target sites with oligonucleotide probes defines the DNA sequence best suited for the isolation of the DNA-binding protein.

This method is very rapid and detects proteins with high affinities for DNA target sites. Specific target site affinities are often described by the dimensionless term K_r, which is simply the ratio of the equilibrium constant for the interaction of the protein with its specific target site, divided by the equilibrium constant for the interaction of the

protein with non-specific DNA (Calzone *et al.*, 1988). Measurements of K_r are often made by challenging DNA-protein complexes with increasing amounts of unlabeled, specific target sites ("cold probe"). As the concentration of specific target sites increases, the amount of labeled complex will decrease as more protein begins to specifically interact with the cold probe. Large K_r values are thus indicative of high affinity DNA-protein interactions; the CyIIIa factors have K_r values of $\geq 10^4$. Since transcriptional control is mediated by high affinity protein interactions at *cis*-regulatory sites, the rapid mapping procedure described above identifies those target sites which are likely to have significant regulatory functions. All of the CyIIIa target sites mapped in this manner have such functions (Franks *et al.*, 1990; Hough-Evans *et al.*, 1990; C. Kirchhamer and E. Davidson, unpublished).

Methodologies for site-specific DNA-binding protein identification

There are two basic methods which allow the identification of proteins which bind specifically to a DNA target site of known sequence. The first method, often referred to as "ligand-based expression screening," depends on the functional expression of recombinant DNA binding proteins and their specific recognition of DNA target sites (reviewed by Singh, 1993). In practice, a library is constructed in which proteins encoded by cDNA inserts are expressed in an active form that will bind DNA. Duplicate filters of protein-expressing clones are then hybridized with a labeled DNA-target site probe. Some positive clones contain cDNA inserts encoding DNA binding proteins which recognize the DNA target site used as a probe; however, there are many false-positive signals. Since first described in the isolation of C/EBP (Vinson *et al.*, 1988) and MBP-1 (Singh *et al.*, 1988), this ligand screening method has been used to identify a number of different DNA-binding proteins.

The ligand-based screening method described above has several advantages over the second method for DNA-binding protein isolation, affinity chromatography. DNA-binding proteins are generally present in very low levels, typically a few hundred to a few thousand molecules per cell in the sea urchin embryo (Calzone *et al.*, 1988; Ch. 4). Methods such as affinity chromatography require tremendous amounts of material for the purification of sufficient quantities of DNA-binding proteins for subsequent micro-sequence analysis. Since the ligand-screening utilizes recombinant protein, the large amounts of nuclear material are not required for factor purification. For organisms in which it is impractical to obtain large quantities of nuclear material, either from cell tissue culture, the isolation of specific tissues or the collection of large numbers of embryos, the ligand-screening method is often the only feasible means to identify transcription factors. In addition, cDNA clones which encode the DNA-binding protein are identified during the screening process; additional screening steps are not required except for the isolation of additional clones.

Despite these advantages, the ligand-based screening method has a number of disadvantages. The first disadvantage is the requirement of obtaining properly refolded protein from the plaques expressing the recombinant cDNA clones. In many instances, overexpression of eukaryotic proteins in *E. coli* results in the formation of insoluble inclusion bodies in which the recombinant polypeptides are deposited. In order for DNA recognition and binding to occur, the recombinant protein must be dissolved and appropriately renatured. If the recombinant protein does not properly refold, the ligand-based screening method will fail to identify the appropriate cDNA clone. In our experiences of CyIIIa factor purification, the majority of sea urchin DNA binding proteins were not identified by the ligand-screening method. About 10 CyIIIa target sites were used as probes to screen an early embryo expression library, but only the genes encoding two proteins, SpZ12-1 (Wang *et al.*, 1994) and SpP3A1 (Höög *et al.*, 1991),

were identified (G.-W. Wang and F. Calzone, unpublished). The SpZ12-1 and SpP3A1 proteins have high affinities for their DNA targets; proteins with low target affinities are often not detected with ligand screening. A second disadvantage of the method is the failure to identify transcription factors which require heterodimer formation for DNA-binding activity since a given cDNA clone will only express one recombinant protein species. Thus proteins such as fos (reviewed by Johnson and McKnight, 1989) and myogenin (reviewed by Li and Olson, 1992), which require heterodimer formation for high affinity DNA binding activity, will not be easy to identify by ligand-based screening.

Affinity chromatography is a general, all purpose method for isolating DNA-binding proteins (reviewed by Kadonaga, 1991; Jarrett, 1993) and is the method we use for the isolation of sea urchin transcription factors (Calzone *et al.*, 1991; Coffman *et al.*, 1992). If DNA-binding activity for a particular target site probe is present in nuclear extract, affinity chromatography should always work to purify the protein or proteins responsible for that DNA-binding activity. The procedure purifies specific DNA-binding proteins from other proteins by applying nuclear extract to affinity columns bearing DNA target sites which are then subjected to amino acid sequencing. Partial amino acid sequences are then used to derive nucleic acid probes for the isolation of the gene encoding the DNA-binding protein. The ability of the affinity column to retain transcription factors is due to the relatively high affinity of the protein for its specific DNA target site. Non-specific proteins are removed from the column during subsequent washes and specific DNA-binding proteins are eluted from the column with a salt gradient. Early attempts at site-specific affinity chromatography employed target site containing plasmids adsorbed to cellulose (Rosenfeld and Kelly, 1986) or target sites present in monomeric oligonucleotides attached to agarose (Wu *et al.*, 1987; Blanks and McLaughlin, 1988). A major advance in DNA affinity chromatography was the coupling

of multimerized oligonucleotides, containing DNA target sites, to an agarose support (Kadonaga and Tjian, 1986). This greatly increases the number of DNA target sites which can be covalently attached to the column matrix, providing a high concentration of specific sites within the column to retain DNA-binding proteins. Typical proteins can be purified 100- to 1000-fold with ≥30% yields (Kadonaga, 1991; Calzone *et al.*, 1991; Coffman *et al.*, 1992). Several DNA-binding proteins purified by affinity chromatography are presented in Table 2. This is by no means an exhaustive list; it simply provides several examples of different types of transcription factors, present in a variety of organisms, which have been purified by this method.

Affinity column fractions often contain a substantial number of proteins which may hinder identification of the DNA-binding protein of interest. Thus it is often necessary to further identify which protein in the mixture is responsible for specific-DNA binding. The first step is to identify the column fractions which contain the previously observed DNA-binding activity using gel retardation assays. Once the DNA-binding activity has been localized to a set of column fractions, several different procedures may be used for the identification of the protein responsible for specific complex formation. A procedure often used to identify sea urchin transcription factors is the "Southwestern" blot (Calzone *et al.*, 1991; Coffman *et al.*, 1992). In this procedure, samples of column fractions containing DNA-binding activity are separated by standard SDS-protein gels. The separated proteins are transferred to nitrocellulose, refolded, and incubated with a probe containing the DNA target site of interest. Proteins which bind the probe are likely to be the proteins responsible for complex formation.

Transcription factors which need to form heterodimers for DNA-binding activity are not detectable by the Southwestern method. A second method, the 2-dimensional gel shift (Coffman *et al.*, 1992, Ch. 2), is utilized to detect such complexes. In this assay,

complexes formed in a standard gel shift reaction are separated by electrophoresis (the first dimension), then the gel lane containing the DNA-protein complexes is placed across the top of an SDS-protein gel. Electrophoresis in the second dimension separates the components of the DNA-protein complexes and resolves the individual polypeptides into discrete bands. An analysis of the resulting pattern of polypeptides can elucidate the stoichiometries of the proteins binding to their DNA target site. Between the two methods described here, we have been able to identify, by molecular weight, all of the proteins which bind to *cis*-regulatory sites in the CyIIIa promoter.

As mentioned previously, affinity chromatography requires large amounts of nuclear material for the purification of sufficient quantities of transcription factors. An advantage of the sea urchin system is the ability to rear very large quantities of synchronously growing embryos, to a desired developmental stage, which can then be processed for nuclear proteins (Calzone *et al.*, 1991; Coffman *et al.*, 1992). To date, we have grown and processed over 10^{11} embryos, which represents about 4×10^{13} nuclei. We have developed an automated affinity chromatography system which allows the simultaneous isolation of up to a dozen different DNA-binding proteins (Coffman *et al.*, 1992). Examples of the recoveries of DNA-binding protein activity from several factors binding target sites in the CyIIIa promoter are presented in Table 3 (Coffman *et al.*, 1992). Factors which are present in very low abundance, such as the factor binding to the P4 site of the CyIIIa regulatory region (table 3), may be purified from several different rounds of affinity chromatography until enough protein has accumulated for protein sequence analysis. Factors which bind to specific target sites in the Endo16 and SM50 target sites are now being isolated by automated affinity chromatography (C.-H. Yuh, K. Makabe, J. Coffman and E. Davidson, unpublished). The sea urchin embryo system provides us with sufficient amounts of nuclear extract from which transcription factors may be isolated.

Once a protein has been accumulated in sufficient quantities, partial amino acid sequence is obtained and used to design oligonucleotide probes for the identification of the corresponding gene. In our experiences, N-terminal protein sequencing is not possible as the proteins are most likely modified at the N-terminal residue. We therefore generate a series of peptides using a lysylendopeptidase (*Achromobacter* protease I, often called "Lys-C"). This protease specifically cleaves the carboxy terminal of lysine residues and generally produces peptides of sufficient length for subsequent gene cloning procedures. The peptides are separated by reverse phase HPLC (Char *et al.*, 1993; Coffman and Davidson, 1994; Ch. 2) and individual peaks, containing single peptide species, are collected. Amino acid sequence is obtained from automated Edman degradation sequencing as described (Char *et al.*, 1993) and several different peptide sequences are usually obtained from each purified protein. The procedures for the actual indentification of the gene encoding the transcription factor are described in the next section.

Isolation of genes encoding sea urchin transcription factors

The mRNAs encoding several different sea urchin transcription factors belong to the "rare" class of messenger RNAs (Cutting *et al.*, 1990; Wang *et al.*, 1994). While many of these mRNAs may be stored maternally, the number present per average cell, as development continues, steadily decreases until only one or two copies of the mRNA are present per average cell (Cutting *et al.*, 1990; Wang *et al.*, 1994; Ch. 2). The frequency of transcription factor mRNAs in a population of embryo mRNAs is thus relatively low, and the detection of cDNA clones encoding transcription factors is problematic.

Several considerations are necessary when designing nucleic acid probes derived from amino acid sequences. Since the amino acid code is degenerate, a given amino acid present in a protein coding region, may be encoded by any of several different codons. Therefore, one must account for all possible codon possibilities when designing nucleic acid probes. The degeneracy of such a probe refers to the number of different sequences present in the population of probe molecules. For example, an oligonucleotide encoding a pentapeptide of methionine residues has a degeneracy of one since methionine is encoded by a single codon. An oligonucleotide encoding a pentapeptide of prolines accounting for all possible codons is 1024-fold degenerate (4 codons, five positions = 4^5). Peptides composed of amino acids encoded by only a few different codons should therefore be selected to design nucleic acid probes.

We have used three different screening strategies for isolating the cDNA clones encoding sea urchin transcription factors. The first method is the direct screening of a phage cDNA library with a pool of degenerate oligonucleotide probes. After hybridization, probe which is not specifically hybridized to complementary target sequences is removed with washes of a quaternary alkylammonium salt such as tetramethyl ammonium chloride (Wood *et al.*, 1985; Calzone *et al.*, 1991; Anderson *et al.*, 1994). Unlike normal stringency washes using sodium salts, quaternary ammonium salt washes select for probes hybridizing with a low number of mismatches across the entire length of the probe. Using this method, Anderson *et al.* (1994) screened a genomic library for DNA-binding protein target sites and were able to wash to a stringency of 4-5 base mismatches out of a total probe length of 29 nucleotides. The quaternary alkylammonium salt method was also used to isolate the cDNA clone encoding the P3A2 transcription factor (Calzone *et al.*, 1991).

There are a number of disadvantages to the direct screening of a cDNA library with a pool of degenerate oligonucleotides, the main disadvantage being the low concentration of specific probe in the pool of degenerate oligonucleotides. This results in a probe of low specific activity, which makes the identification of positive clones extremely difficult. If we suppose that a plaque contains 10^7 phage, of which 50% are exposed for probe binding, and a single sequence oligonucleotide probe is labeled with high specific activity ^{32}P (e.g., 4 X 10^8 CPM/50 ng of a 30-mer oligonucleotide), then one might expect about 500 CPM of probe binding per positive plaque. If a degenerate probe is used directly in the library screen, the number of labeled molecules which contain the sequence complementary to the actual clone will decrease as the degeneracy of the probe pool increases. This can reach a point in which insufficient numbers of probe molecules are bound to target sites and are not readily detectable.

As a way to increase the relative concentration of target sequences to probe sequences, we and others have utilized a plasmid-based cDNA library screening method. This provides roughly an order of magnitude increase in the number of cDNA target sites available for hybridizing to the degenerate probe pool. The plasmid cDNA library is subdivided into aliquots of several thousand clones. The plasmids of each aliquot are recovered and digested to release the cDNA inserts, which are then separated by gel electrophoresis and transferred to an appropriate membrane. Using this method, the number of specific target sequences present in a single band is about 10^8 (given a plasmid MW of 3.3 X 10^6 and digesting 1 μg of plasmid from a 2500 clone pool). The membrane containing the digested clones is then hybridized and washed with the quaternary ammonium salts described above. Aliquots containing positive clones are further subdivided and rescreened until a small number of possible clones remain. The desired clone is then recovered by the direct hybridization of the probe to the bacterial colonies. This method has been used to isolate clones for the SpTEF-1 homologue (J.

Xian and E. Davidson, unpublished) and clones for at least one homeodomain protein (P. Martinez and E. Davidson, unpublished).

Non-degenerate probes may be obtained from partial amino acid sequences through the use of the polymerase chain reaction (PCR). The PCR process amplifies DNA sequence which is flanked by two oligonucleotide primers from minute amounts of nucleic acid template. If two degenerate primers are used for PCR, the amplified nucleic acid sequence will serve as an excellent, non-degenerate probe for subsequent library screening. Two different strategies may be employed. Degenerate primers derived from the terminal 6-7 amino acids of a particularly long peptide may be used with PCR to amplify the nucleic acid sequence encoding the internal amino acids (Lee *et al.*, 1988). This method was used to identify the protein binding to the CyIIIa P7II site (Coffman and Davidson, 1994) and the Sp(G/C)F1 protein which binds to the CyIIIa P8 site (Ch. 2). If two short peptide sequences are available, degenerate primers may be designed to amplify the nucleic acid sequence encoding the protein between the two primers. This method will work if the primers are oriented in the appropriate direction and nucleic acid lacking introns, such as cDNA, is used as the template.

The PCR method of identifying non-degenerate transcription factor probes has a number of advantages over the direct use of degenerate oligonucleotide probes. First, since the probes derived from PCR are non-degenerate, higher stringency conditions may be used to isolate an appropriate clone. Second, a variety of different templates, such as cDNAs from many different tissues or developmental stages, may be examined simultaneously. This eliminates the need to screen many different libraries to identify the clone of interest. Oligonucleotides which are too degenerate in sequence for direct library screening may be used successfully in PCR amplification. In addition, oligonucleotides derived from two short peptide sequences which are of insufficient

length to be used for direct library screening may work perfectly well in PCR amplification.

Future prospects

The number of transcription factors which regulate early sea urchin gene expression has been estimated by 2-dimensional gel electrophoresis (Harrington *et al.*, 1992). DNA-binding proteins usually interact with DNA sequences through a region of basic residues and this feature allows the enrichment of transcription factors in nuclear extract with cation-exchange chromatography. When nuclear extracts from 24 hr sea urchin embryos were chromatographically fractionated and analyzed by 2-dimensional electrophoresis, about 265 polypeptides, representing about 100 different transcription factors, were observed (Harrington *et al.*, 1992). This suggests that genes expressed in the early sea urchin embryo are regulated by a relatively small number of DNA-binding proteins (Coffman and Davidson, 1994). With the availability of large quantities of nuclear extract, the identification of *cis*-regulatory target sites from differentially expressed genes, and the rapid development of techniques to sequence smaller amounts of proteins, it is conceivable that we could isolate all of the relevant transcription factors present in the sea urchin embryo. We would then have the molecular components necessary to understand the transcriptional regulatory network of the sea urchin embryo.

Organization of the thesis

Chapter 2 describes the molecular characterization of the Sp(G/C)F1 gene which encodes a novel DNA binding protein with high affinities for G/C rich DNA target sites. Five different forms of the Sp(G/C)F1 protein are present in embryonic nuclear extracts. By examining recombinant protein, we have shown that the different Sp(G/C)F1 polypeptides share a common DNA-binding domain but have different N-terminal sequences. The differential inclusion of proline-rich, N-terminal sequences into each of the five polypeptides may alter the transcriptional activation potential of these proteins. Chapter 2 is being prepared for publication.

Chapter 3 presents a model for Sp(G/C)F1 protein function in the embryo. Quantitative gel mobility shift assays demonstrate that this protein self-associates and must bind DNA cooperatively, most likely as a dimer. As a test of this hypothesis, we used electron microscopy to observe Sp(G/C)F1 protein bound to *cis*-regulatory sites in the CyIIIa promoter. When bound to DNA, this protein is able to self-associate by "looping out" intervening sections of DNA. Therefore, the Sp(G/C)F1 protein is an excellent candidate for an inter-regulatory module communicator. Chapter 3 is being prepared for publication.

Chapter 4 presents an immunological examination of the nuclear concentrations of two previously cloned DNA binding proteins, P3A1 and P3A2. These proteins are present maternally, and both recognize the same target site in several sea urchin gene *cis*-regulatory domains. Only P3A2 protein is present in sufficient quantities later in development (post-cleavage) to have any meaningful effect on gene regulation. Chapter 4 is being prepared for publication.

Chapter 5 presents a brief summary of the work presented in this thesis and suggests several areas for future investigations.

The appendices include three papers relating to sea urchin gene regulation in which I have participated:

Appendix A details the analysis of negative-acting DNA binding protein target sites present in the CyIIIa *cis*-regulatory region. This has been published: Hough-Evans *et al.*, 1990.

Appendix B describes the molecular characterization and cloning of the P3A2 gene which encodes the founding member of a new family of DNA binding proteins. This has been published: Calzone *et al.*, 1991.

Appendix C demonstrates an immunological detection scheme to analyze reporter gene expression from three different territorial-specific *trans*-genes. This has been published: Zeller *et al.*, 1992.

REFERENCES

Akasaka, K., Frudakis, T., Killian, C.E., George, N.C., Yamasu, K., Khaner, O. and Wilt, F.H. (1994). Genomic organization of a gene encoding the spicule matrix protein SM30 in the sea urchin *Strongylocentrotus purpuratus*. *J. Biol. Chem.* 269:20592-20958.

Akasaka, K., Ueda, T., Higashiakagawa, T., Yamada, K. and Shimada, H. (1990). Spatial pattern of arylsulfatase mRNA expression in sea urchin embryos. *Develop. Growth Differ.* 32:9-13.

Anderson, R., Britten, R.J. and Davidson, E.H. (1994). Repeated sequence target sites for maternal DNA-binding proteins in genes activated in early sea urchin development. *Dev. Biol.* 163:11-18.

Angerer, L.M., Chambers, S.A., Yang, Q., Venkatesan, M., Angerer, R.C. and Simpson, R.T. (1988). Expression of a collagen gene in mesenchyme lineages of the *Strongylocentrotus purpuratus* embryo. *Genes Dev.* 2:239-246.

Angerer, L.M., Dolecki, G.J., Gagnon, M., Lum, R., Wang, G., Yang, Q., Humphries, T. and Angerer, R.C. (1989). Progressively restricted expression of a homeobox gene within the aboral ectoderm of developing sea urchin embryos. *Genes Dev.* 3:370-383.

Beg, A.A. and Baldwin, A.S. (1993). The IκB proteins: Multifunctional regulators of Rel/NF-κB transcription factors. *Genes Dev.* 7:2064-2070.

Benson, S., Sucov, H., Stephens, L., Davidson, E. and Wilt, F. (1987). A lineage-specific gene encoding a major spicule matrix protein of the sea urchin spicule. I. Authentification of the cloned gene and its developmental expression. *Dev. Biol.* 120:499-506.

Bisgrove, B.W., Andrews, M.E. and Raff, R.A. (1991). Fibropellins, products of an EGF repeat-containing gene, form a unique extracellular matrix structure that surrounds the sea urchin embryo. *Dev. Biol.* 146:89-99.

Blank, V., Kourilsky, P. and Isreal, A. (1992). NF-κB and related proteins: Rel/Dorsal homologies meet ankyrin-like repeats. *Trends Bioc. Sci.* 17:135-140.

Blanks, R. and McLaughlin, L.W. (1988). An oligodeoxynucleotide affinity column for the isolation of sequence specific DNA binding proteins. *Nuc. Acids Res.* 16:10283-10299.

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

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

Cameron, R.A. and Davidson, E.H. (1992). Cell type specification during sea urchin development. *Trends Gen.* 7:212-218.

Cameron, R.A., Hough-Evans, B.R., Britten, R.J. and Davidson, E.H. (1987). Lineage and fate of each blastomere of the eight-cell sea urchin embryo. *Genes Dev.* 1:75-84.

Char, B.R., Bell, J.R., Dovala, J., Coffman, J.A., Harrington, M.G., Becerra, J.C., Davidson, E.H., Calzone, F.J. and Maxson, R. (1993). SpOct, a gene encoding the major octamer-binding protein in sea urchin embryos: Expression profile, evolutionary relationships, and DNA binding of expressed protein. *Dev. Biol.* 158:350-363.

Clark, A.R. and Doherty, K. (1993). Negative regulation of transcription in eukaryotes. *Biochem. J.* 296:521-541.

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

Coffman, J.A. and Davidson, E.H. (1994). Regulation of gene expression in the sea urchin embryo. *J.mar. biol. Ass. U.K.* 74:17-26.

Coffman, J.A., Moore, J.G., Calzone, F.J., Britten, R.J., Hood, L.E. and Davidson, E.H. (1992). Automated sequential affinity chromatography of sea urchin DNA binding proteins. *Mol. Mar. Biol. and Biotech.* 1:136-146.

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.

Cutting, A.E., Hoog, C.H., Calzone, F.J., Britten, R.J. and Davidson, E.H. (1990). Rare maternal mRNAs code for regulatory proteins that control lineage specific gene expression in the sea urchin embryo. *Proc. Natl. Acad. Sci. USA.* 87:7953-7959.

D'Alessio, M., Ramirez, F., Suzuki, H., Solursh, M. and Gambino, R. (1989). Structure and developmental expression of a sea urchin fibrillar collagen gene. *Proc. Natl. Acad. Sci. USA.* 86:9303-9307.

D'Alessio, M., Ramirez, F., Suzuki, H., Solursh, M. and Gambino, R. (1990). Cloning of a fibrillar collagene gene expressed in the mesenchymal cells of the developing sea urchin embryo. *J. Biol. Chem.* 265:7050-7054.

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

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

Dynlacht, B.D., Attardi, L.D., Admon, A., Freeman, M. and Tjian, R. (1989). Functional analysis of NTF-1, a developmentally regulated *Drosophila* transcription factor that binds neuronal *cis* elements. *Genes Dev.* 3:1677-1688.

Efiok, B.J.S., Chiorini, J.A. and Safer, B. (1994). A key transcription factor for eukaryotic initiation factor-2α is strongly homologous to developmental transcription factors and may link metabolic genes to cellular growth and development. *J. Biol. Chem.* 269:18921-18930.

Eldon, E.D., Angerer, L.M., Angerer, R.C. and Klein, W.H. (1987). Spec3: Embryonic expression of a sea urchin gene whose product is involved in ectoderm ciliogenesis. *Genes Dev.* 1:1280-1292.

England, B.P., Admon, A. and Tjian, R. (1992). Cloning of *Drosophila* transcription factor Adf-1 reveals homology to Myb oncoproteins. *Proc. Natl. Acad. Sci. USA.* 89:683-687.

Evans, T. and Felsenfeld, G. (1989). The erythroid-specific transcription factor Eryf1: A new finger protein. *Cell* 58:877-885.

Farach, M.C., Valdizan, M., Park, H.R., Decker, G.L. and Lennarz, W.J. (1987). Developmental expression of a cell-surface protein involved in calcium uptake and skeleton formation in sea urchin embryos. *Dev. Biol.* 122:320-331.

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

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

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

Fried, M.G. and Crothers, D.M. (1981). Equilibria and kinetics of lac repressor-operator interactions by polyacrylamide gel electrophoresis. *Nuc. Acids Res.* 9:6505-6524.

Fu, X.-Y., Schindler, C., Improta, T., Aebersold, R. and Darnell, J.E. (1992). The proteins of ISGF-3, the interferon α-induced transcriptional activators, define a gene family involved in signal transduction. *Proc. Natl. Acad. Sci. USA.* 89:7840-7843.

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

Gan, L., Zhang, W. and Klein, W.H. (1990b). Repetitive DNA sequences linked to the sea urchin Spec genes contain transcriptional enhancer-like elements. *Dev. Biol.* 139:186-196.

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

Ghosh, S., Gifford, A.M., Riviere, L.R., Tempst, P., Nolan, G.P. and Baltimore, D. (1990). Cloning of the p50 DNA binding subunit of NF-κB: Homology to *rel* and *dorsal*. *Cell* 62:1019-1029.

Grimwald, J.E., Gagnon, M.L., Yang, Q., Angerer, R.C. and Angerer, L.M. (1991). Expression of two mRNAs encoding EGF-related proteins identifies subregions of sea urchin embryonic ectoderm. *Dev. Biol.* 143:44-57.

Hardin, P.E., Angerer, L.M., Hardin, S.H., Angerer, R.C. and Klein, W.H. (1988). Spec2 genes of Strongylocentrotus purpuratus: Structure and differential expression in embryonic aboral ectoderm cells. *J. Mol. Biol.* 202:417-431.

Harkey, M.A., Whiteley, H.R. and Whiteley, A.H. (1992). Differential expression of the MSP130 gene among skeletal lineage cells in the sea urchin embryo - A 3-dimensional in situ hybridization analysis. *Mech. Dev.* 37:173-184.

Harlow, P., Watkins, E., Thornton, R.D. and Nemer, M. (1989). Structure of an ectodermally expressed sea urchin metallothionein gene and characterization of its metal-responsive region. *Mol. Cell. Biol.* 9:5445-5455.

Harrington, M.G., Coffman, J.A., Calzone, F.J., Hood, L.E., Britten, R.J. and Davidson, E.H. (1992). Complexity of sea urchin embryo nuclear proteins that contain basic domains. *Proc. Natl. Acad. Sci. USA.* 89:6252-6256.

Höög, C.H., Calzone, F.J., Cutting, A.E., Britten, R.J. and Davidson, E.H. (1991). Gene regulatory factors of the sea urchin embryo II. Two dissimilar proteins, P3A1 and P3A2, bind to the same target sites that are required for early territorial gene expression. *Development* 112:351-364.

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

Hough-Evans, B.R., Franks, R.R., Cameron, R.A., Britten, R.J. and Davidson, E.H. (1987). Correct cell type-specific expression of a fusion gene injected into sea urchin eggs. *Dev. Biol.* 121:576-579.

Hough-Evans, B.R., Franks, R.R., Zeller, R.W., Britten, R.J. and Davidson, E.H. (1990). Negative spatial regulation of the lineage specific CyIIIa actin gene in the sea urchin embryo. *Development* 110:41-50.

Hu, Y.-F., Lüscher, B., Admon, A., Mermod, N. and Tjian, R. (1990). Transcription factor AP-4 contains multiple dimerization domains that regulate dimer specificity. *Genes Dev.* 4:1741-1752.

Hwang, J.-J., Chambon, P. and Davidson, I. (1993). Characterization of the transcription activation function and the DNA binding domain of transcriptional enhancer factor-1. *EMBO J.* 12:2337-2348.

Jarrett, H.W. (1993). Affinity chromatography with nucleic acid polymers. *J. Chrom.* 618:315-339.

Johnson, P.F. and McKnight, S.L. (1989). Eukaryotic transcriptional regulatory proteins. *Annu. Rev. Biochem.* 58:799-839.

Kadonaga, J.T. (1991). Purification of sequence-specific binding proteins by DNA affinity chromatography. *Meth. Enzym.* 208:10-23.

Kadonaga, J.T. and Tjian, R. (1986). Affinity purification of sequence-specific DNA binding proteins. *Proc. Natl. Acad. Sci. USA.* 83:5889-5893.

Kadonaga, J.T., Carner, K.R., Masiarz, F.R. and Tjian, R. (1987). Isolation of cDNA encoding transcription factor Sp1 and functional analysis of the DNA binding domain. *Cell* 51:1079-1090.

Kaoshima, H., Shigesada, S., Satake, M., Ito, Y., Miyoshi, H., Ohki, M., Pepling, M. and Gergen, P. (1993). The runt domain identifies a new family of heteromeric transcriptional regulators. *Trends Gen.* 9:338-341.

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

Kieran, M., Blank, V., Logeat, F., Vandekerckhove, J., Lottspeich, F., Le Bail, O., Urban, M.B., Kourilsky, P., Baeuerle, P.A. and Isreal, A. (1990). The DNA binding subunit of NF-κB is identical to factor KBF1 and homologous to the *rel* oncogene product. *Cell* 62:1007-1018.

Kingston, R.E. and Green, M.R. (1994). Modeling eukaryotic transcriptional activation. *Curr. Biol.* 4:325-332.

Lee, C.C., Wu, X., Gibbs, R.A., Cook, R.G., Muzny, D.M. and Caskey, C.T. (1988). Generation of cDNA probes directed by amino acid sequence: Cloning of urate oxidase. *Science* 239:1288-1291.

Li, L. and Olson, E.N. (1992). Regulation of muscle cell growth and differentiation by the MyoD family of helix-loop-helix proteins. *Adv. Cancer Res.* 58:95-119.

Lynn, D.A., Angerer, L.M., Bruskin, A.M., Klein, W.H. and Angerer, R.C. (1983). Localization of a family of mRNAs in a single cell type and its precursors in sea urchin embryos. *Proc. Natl. Acad. Sci. USA*. 80:2656-2660.

McMahon, A.P., Flytzanis, C.N., Hough-Evans, B.R., Katula, K.S., Britten, R.J. and Davidson, E.H. (1985). Introduction of cloned DNA into sea urchin egg cytoplasm: Replication and persistance during embryogenesis. *Dev. Biol.* 108:420-430.

Mitchell, P.J. and Tjian, R. (1989). Transcriptional regulation in mammalian cells by sequence-specific DNA binding proteins. *Science* 245:371-378.

Nemer, M., Thornton, R.D., Stuebing, E.W. and Harlow, P. (1991). Structure, spatial, and temporal expression of two sea urchin metallothionein genes, $SpMTB_1$ and SpMTA. *J. Biol. Chem.* 266:6586-6593.

Niemeyer, C.C. and Flytzanis, C.N. (1993). Upstream elements involved in the embryonic regulation of the sea urchin CyIIIb actin gene: Temporal and spatial specific interactions at a single *cis*-acting element. *Dev. Biol.* 156:293-302.

Nocente-McGrath, C., Brenner, C.A. and Ernst, S.G. (1989). Endo16, a lineage-specific protein of the sea urchin embryo, is first expressed just prior to gastrulation. *Dev. Biol.* 136:264-272.

Norman, C., Runswick, M., Pollock, R. and Treisman, R. (1988). Isolation and properties of cDNA clones encoding SRF, a transcription factor that binds to the *c-fos* serum response element. *Cell* 55:989-1003.

Ogawa, E., Inuzuka, M., Maruyama, M., Satake, M., Naito-Fujimoto, M., Ito, Y. and Shigesada, K. (1993a). Molecular cloning and characterization of PEBP2β, the heterodimeric partner of a novel *Drosophila runt*-related DNA binding protein PEBP2α. *Virology* 194:314-331.

Ogawa, E., Maruyama, M., Kagoshima, H., Inuzuka, M., Lu, J., Satake, M., Shigesada, K. and Ito, Y. (1993b). PEBP2/PEA2 represents a family of transcription factors

homologous to the products of the *Drosophila* runt gene and the human *AML1* gene. *Proc. Natl. Acad. Sci. USA*. 90:6859-6863.

Pabo, C.O. and Sauer, R.T. (1992). Transcription factors: Structural families and principles of DNA recognition. *Annu. Rev. Biochem.* 61:1053-1095.

Parr, B.A., Parks, A.L., and Raff, R.A. (1989). Promoter structure and protein sequence of Msp130, a sea urchin primary mesenchyme cell glycoprotein. *J. Biol. Chem.* 265:1408-1413.

Perkins, K.K., Admon, A., Patel, N. and Tjian, R. (1990). The *Drosophila* fos-related AP-1 protein is a developmentally regulated transcription factor. *Genes Dev.* 4:822-834.

Ptashne, M. (1988). How eukaryotic transcriptional activators work. *Nature* 335:683-689.

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

Reynolds, S.D., Angerer, L.M., Palis, J., Nasir, A. and Angerer, R.C. (1992). Early mRNAs, spatially restricted along the animal-vegetal axis of sea urchin embryos, include one encoding a protein related to tolloid and BMP-1. *Development* 114:769-786.

Rosenfeld, P.J. and Kelly, T.J. (1986). Purification of nuclear factor-I by DNA recognition site affinity chromatography. *J. Biol. Chem.* 261:1398-1408.

Santoro, C., Mermod, N., Andrews, P.C. and Tjian, R. (1988). A family of human CCAAT-box-binding proteins active in transcription and DNA replication: cloning and expression of multiple cDNAs. *Nature* 334:218-224.

Schwarz-Sommer, Z., Huijser, P., Nacken, W., Saedler, H. and Sommer, H. (1990). Genetic control of flower development by homeotic genes in *Antirrhinum majus*. *Science* 250:931-936.

Singh, H. (1993). Specific recognition site probes for isolating genes encoding DNA-binding proteins. *Meth. Enzym.* 218:551-567.

Singh, H., LeBowitz, H., Baldwin, A.S. and Sharp, P.A. (1988). Molecular cloning of an enhancer binding protein: Isolation by screening of an expression library with a recognition site DNA. *Cell* 52:415-423.

Soeller, W.C., Oh, C.E. and Kornberg, T.B. (1993). Isolation of cDNAs encoding the *Drosophila* GAGA transcription factor. *Mol. Cell. Biol.* 13:7961-7970.

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

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

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

Vinson, C.R., LaMarco, K.L., Johnson, P.F., Landschultz, W.H., and McKnight, S.L. (1988). In situ detection of sequence-specific DNA binding activity specified by a recombinant bacteriophage. *Genes Dev.* 2:801-806.

Virbasius, C.-m. A., Virbasius, J.V. and Scarpulla, R.C. (1993). NRF-1, an activator involved in nuclear-mitochondrial interactions, utilizes a new DNA-binding domain conserved in a family of developmental regulators. *Genes Dev.* 7:2431-2445.

Wakao, H., Gouilleux, F. and Groner, B. (1994). Mammary gland factor (MGF) is a novel member of the cytokine transcription factor gene family and confers the prolactin response. *EMBO J* 13:2182-2191.

Wang, D.G.-W., Kirchhamer, C.V., Britten, R.J. and Davidson, E.H. (1994). SpZ12-1, a negative regulator required for spatial control of the territory-specific CyIIIa gene in the sea urchin embryo. Submitted.

Wang, L.-H., Tsai, S.Y., Cook, R.G., Beattie, W.G., Tsai, M.-J. and O'Malley, B.W. (1989). COUP transcription factor is a member of the steroid receptor superfamily. *Nature* 340:163-166.

Wessel, G.M., Goldberg, L., Lennarz, W.J. and Klein, W.H. (1989a). Gastrulation in the sea urchin embryo is accompanied by the accumulation of an endoderm-specific mRNA. *Dev. Biol.* 136:526-536.

Wessel, G.M., Zhang, W., Tomlinson, C.R., Lennarz, W.J. and Klein, W.H. (1989b). Transcription of the Spec1-like gene of *Lytechinus* is selectively inhibited in response to disruption of the extracellular matrix. *Development* 106:355-365.

Wiederrecht, G., Seto, D. and Parker, C.S. (1988). Isolation of the gene encoding the *S. cerevisiae* heat shock transcription factor. *Cell* 54:841-853.

Wilkinson, D.G. and Nemer, M. (1987). Metallothionein genes MTa and MTb expressed under distinct quantitative and tissue-specific regulation in sea urchin embryos. *Mol. Cell. Biol.* 7:48-58.

Williams, T., Admon, A., Lüscher, B. and Tjian, R. (1988). Cloning and expression of AP-2, a cell-type-specific transcription factor that activates inducible enhancer elements. *Genes Dev.* 2:1557-1569.

Wood, W.I., Gitschier, J., Laskey, L.A. and Lawn, R.M. (1985). Base composition-independent hybridization in tetramethyl ammonium chloride - a method for oligonucleotide screening of highly complex gene libraries. *Proc. Natl. Acad. Sci. USA.* 82:1585-1588.

Wu, C., Wilson, S., Walker, B., Dawid, I., Paisley, T., Zimarino, V. and Ueda, H. (1987). Purification and properties of *Drosophila* heat shock activator protein. *Science* 238:1247-1253.

Xiang, M., Bedard, P.-A., Wessel, G.M., Filion, M., Brandhorst, B.P. and Klein, W.H. (1988). Tandem duplication and divergence of a sea urchin protein belonging to the troponin C superfamily. *J. Biol. Chem.* 263:17173-17180.

Xiang, M., Ge, T., Tomlinson, C.R. and Klein, W.H. (1991). Structure and promoter activity of the LpS1 genes of *Lytechinus pictus*: Duplicated exons account for LpS1 proteins with eight different calcium-binding domains. *J. Biol. Chem.* 266:10524-10533.

Xiao, J.H., Davidson, I., Matthes, H., Garnier, J.-M. and Chambon, P. (1991). Cloning, expression, and transcriptional properties of the human enhancer factor TEF-1. *Cell* 65:551-568.

Yang, Q., Angerer, L.M. and Angerer, R.C. (1989). Structure and tissue-specific developmental expression of a sea urchin arylsulfatase gene. *Dev. Biol.* 135:53-65.

Yuh, C.-H., Ransick, A., Martinez, P., Britten, R.J. and Davidson, E.H. (1994). Complexity and organization of DNA-protein interactions in the 5'-regulatory region of an endoderm-specific marker gene in the sea urchin embryo. *Mech. Dev.* 47:165-186.

Zeller, R.W., Cameron, R.A., Franks, R.R., Britten, R.J. and Davidson, E.H. (1992). Territorial expression of three different *trans*-genes in early sea urchin embryos detected by a whole-mount flourescence procedure. *Dev. Biol.* 151:382-390.

Table 1. Spatially regulated genes in the sea urchin embryo (After Coffman and Davidson, 1992).

Gene	Protein Product	Territory*	Species*	References
Spec1	Ca2+ binding protein	AE	S.p.	A
Spec2 (a-d)	Ca2+ binding protein		S.p.	B
LvS1	Ca2+ binding protein		L.v.	C
LpS2 (a,b)	Ca2+ binding protein		L.p.	D, E
SpMTa	Metallothionein		S.p.	F, G, H
CyIIIa	Cytoskeletal Actin		S.p.	I
CyIIIb	Cytoskeletal Actin		S.p.	I
SpARS	Arylsulfatase		S.p.	J
ARS	Arylsulfatase		H.p.	K
Hbox1	Transcription Factor		S.p., T.g.	L
SpEGF1	ECM Protein	AE+OE	S.p.	M, N
SpEGFII	ECM Protein		S.p.	M
SpHe	Metalloendoprotease		S.p.	O
SpAN	Metalloendoprotease		S.p.	O
Spec3	Ciliary Protein		S.p.	P
SpMTb1	Metallothionein	Global	S.p.	H
CyI	Cytoskeletal Actin	Early	S.p.	I
CyIIb	Cytoskeletal Actin	OE+V Late	S.p.	I
Endo16	Cell Surface Protein	V	S.p.	Q
LvN1.2	Cytoplasmic Protein		L.v.	R
CyIIa	Cytoskeletal Actin	V+SM	S.p.	I
Collagen	ECM Protein		S.p.	S
COLL1α	ECM Protein		P.l.	T
COLL2α	ECM Protein		P.l.	U
SM50	Spicule Matrix protein	SM	S.p.	V, W, X
SM30	Spicule Matrix protein		S.p.	Y
MSP130	Cell-surface glycoprotein		S.p.	Z, AA

Table 1, continued.

* Territory abbreviations: AE - aboral ectoderm, OE - oral ectoderm, SM - skeletogenic mesenchyme, V - vegetal plate. Species abbreviations: H.p. - *Hemicentrotus pulcherrimus*, L.p. - *Lytechinus pictus*, L.v. - *Lytechinus variegatus*, P.l. - *Paracentrotus lividus*, S.p. - *Strongylocentrotus purpuratus*, T.g. - *Tripneustes gratilla.*

References:

A) Lynn *et al.*, 1983. B) Hardin *et al.*, 1988. C) Wessel *et al.*, 1989b. D) Xiang *et al.*, 1988. E) Xiang *et al.*, 1991. F) Harlow *et al.*, 1989. G) Wilkinson *et al.*, 1987. H) Nemer *et al.*, 1991. I) Cox *et al.*, 1986. J) Yang *et al.*, 1989. K) Akasaka *et al.*, 1990. L) Angerer *et al.*, 1989. M) Grimwade *et al.*, 1991. N) Bisgrove *et al.*, 1991. O) Reynolds *et al.*, 1992. P) Eldon *et al.*, 1987. Q) Nocente-McGrath *et al.*, 1989. R) Wessel *et al.*, 1989a. S) Angerer *et al.*, 1988. T) D'Alessio *et al.*, 1989. U) D'Alessio *et al.*, 1990. V) Benson *et al.*, 1987. W) Katoh-Fukui *et al.*, 1991. X) Sucov *et al.*, 1987. Y) George *et al.*, 1991. Z) Parr *et al.*, 1989. AA) Harkey *et al.*, 1992.

Table 2. Some transcription Factors purified by DNA affinity chromatography.*

Factor	Factor Class†	Organism	References
Coup	ZF-HR	Chicken	Wang *et al.*, 1989
Adf-1	HTH	*Drosophila*	England *et al.*, 1992
dFRA	bZIP	*Drosophila*	Perkins *et al.*, 1990
dJRA	bZIP	*Drosophila*	Perkins *et al.*, 1990
GAGA	ZF	*Drosophila*	Soeller *et al.*, 1993
NTF-1	Novel	*Drosophila*	Dynlacht *et al.*, 1989
AP-2	HSH	Human	Williams *et al.*, 1988
AP-4	HLH	Human	Hu *et al.*, 1990
CTF/NF-I	Basic Domain	Human	Santoro *et al.*, 1988
Eryf1	ZF	Human	Evans & Felsenfeld, 1989
NF-κB (p50)	Rel-like	Human	Ghosh *et al.*, 1990 Kieran *et al.*, 1990
NRF-1	P3A2-like	Human	Virbasius *et al.*, 1993
α-Pal = NRF-1	P3A2-like	Human	Efiok *et al.*, 1994
PEBP2α	Runt-Domain	Human	Ogawa *et al.*, 1993b
PEBP2β	PEBP2α partner	Human	Ogawa *et al.*, 1993a
Sp1	ZF	Human	Kadonaga *et al.*, 1987
SRF	MADS box	Human	Norman *et al.*, 1988
MGF	Stat protein	Sheep	Wakao *et al.*, 1994
HSF1	Novel	Yeast	Wiederrecht *et al.*, 1988
SpP3A2	P3A2-like	Sea Urchin Embryo	Calzone *et al.*, 1991
Sp(G/C)F1	Novel	Sea Urchin Embryo	Zeller *et al.*, 1994 Coffman *et al.*, 1992
SpRunt	Runt Domain	Sea Urchin Embryo	J. Coffman & E. Davidson, unpublished; Coffman *et al.*, 1992
SpTEF-1	TEA Domain	Sea Urchin Embryo	J. Xian & E. Davidson, unpublished; Coffman *et al.*, 1992
Sp-Oct	Pou-HD	Sea Urchin Embryo	Char *et al.*, 1993; Coffman *et al.*, 1992
SpP7II	Novel	Sea Urchin Embryo	Coffman & Davidson, 1994, unpublished

Table 2 continued.

* Many transcription factors have been characterized by DNA affinity chromatography. In all examples listed here, except for the COUP and HSF1 factors, partial amino acid sequence from the affinity purified protein was used to design oligonucleotide probes for the identification of the corresponding gene. Affinity purified COUP or HSF1 protein was used as an antigen to make specific antibodies. The antibodies were then used to screen an expression library and recover the appropriate cDNA clone.

† Class abbreviations: bZIP - basic leucine-zipper domain. HSH - helix-span-helix. HTH - helix-turn-helix. MADS-box - a group of DNA binding proteins named after the founding members MCM1, AG, DEF A and SRF (Schwarz-Sommer *et al.*, 1990). P3A2-like - a member of a new family of transcription factors with homologies to P3A2. Pou-HD - POU/Homeo Domain. Rel-like - a member of the rel family of proteins (reviewed by Blank *et al.*, 1992; Beg and Baldwin, 1993). Runt Domain - a Runt family member (Kagoshima *et al.*, 1993). Stat protein - related to a new class of DNA-binding ligand receptor proteins (Fu *et al.*, 1992). TEA Domain - related to the TEF-1 class of transcription factors (Hwang *et al.*, 1993). ZF - zinc finger protein. ZF-HR - zinc finger containing hormone receptor. For those classes not explicitly referenced, a review of the structural properties of these DNA-binding proteins may be found in Pabo and Sauer (1992).

Table 3. Recovery of DNA-binding proteins interacting with *cis*-elements in the CyIIIa promoter (After Coffman *et al.*, 1992).

Factor	Quantity Input (Moles)	Quantity Recovered (Moles)	Yield (%)	Starting Specific Activity (moles/g)	Final Specific Activity (moles/g)	Fold[1] enrichment
P8II*	6.8×10^{-9}	3.5×10^{-9}	51	9.4×10^{-8}	5.2×10^{-6}	55
P3A2	2.1×10^{-9}	3.5×10^{-10}	17	2.9×10^{-8}	4.6×10^{-6}	160
P7I**	9.3×10^{-10}	2.5×10^{-10}	27	1.3×10^{-8}	6.9×10^{-7}	53
P3B†	5.5×10^{-9}	1.1×10^{-9}	20	7.6×10^{-8}	5.9×10^{-6}	78
P5‡	4.5×10^{-10}	7.4×10^{-11}	16	6.3×10^{-9}	4.1×10^{-7}	66
P4	1.4×10^{-9}	1.6×10^{-10}	11	1.9×10^{-8}	1.9×10^{-6}	100

Definitions: **Yield** is the quotient of molar recovery to molar input. **Specific Activity** is the quotient of the number of moles of specific factor to the number of grams of total protein. **Starting** specific activity refers to nuclear extract, **Final** specific activity refers to that in pooled fractions. **Fold enrichment** is the quotient of final specific activity to starting specific activity.

1 The fold enrichment values reported here are from a single affinity column purification step. For P3A2, Calzone *et al.* (1991) reported a fold enrichment value of 33, after the first affinity column purification step. This value increased to 1600 after a second affinity column purification.

* This factor is now known as Sp(G/C)F1 (Ch. 1 & 2). The amount of protein isolated here was recovered from three tandem affinity columns. The amounts reported for the other proteins represent recoveries from single affinity columns.

** This factor is now known as SpRunt (J. Coffman and E. Davidson, unpublished).

† The most prominent protein occupying the P3B site is SpOct (Char *et al.*, 1993).

Table 3 continued.

‡ This protein has recently been sequenced and shows homologies to the mammalian factor TEF-1 (Xiao *et al.*, 1991; Hwang *et al.*, 1993) and is now known as SpTEF-1 (J. Xian and E. Davidson, unpublished).

Figure 1. Strategy for the rapid identification of DNA-binding protein target sites. (A) Map of the CyIIIa regulatory domain. The 2.3 kb CyIIIa regulatory domain has been mapped by subdividing the promoter into small fragments which are examined for the presence of *cis*-target sites by gel shift assays as described in the text. Fragments which were bound specifically by DNA-binding proteins are labeled "+", those that did not show evidence of DNA-protein interactions are labeled "-." Comparing the positions of the various fragments allows sites of DNA-protein interaction to be assigned; these are shown as filled boxes. There are around 20 DNA-binding protein target sites serviced by about 10 different transcription factors in the CyIIIa promoter (Calzone *et al.*, 1988; Thézé *et al.*, 1990; Coffman *et al.*, 1992). This diagram is redrawn from Thézé *et al.* (1990).

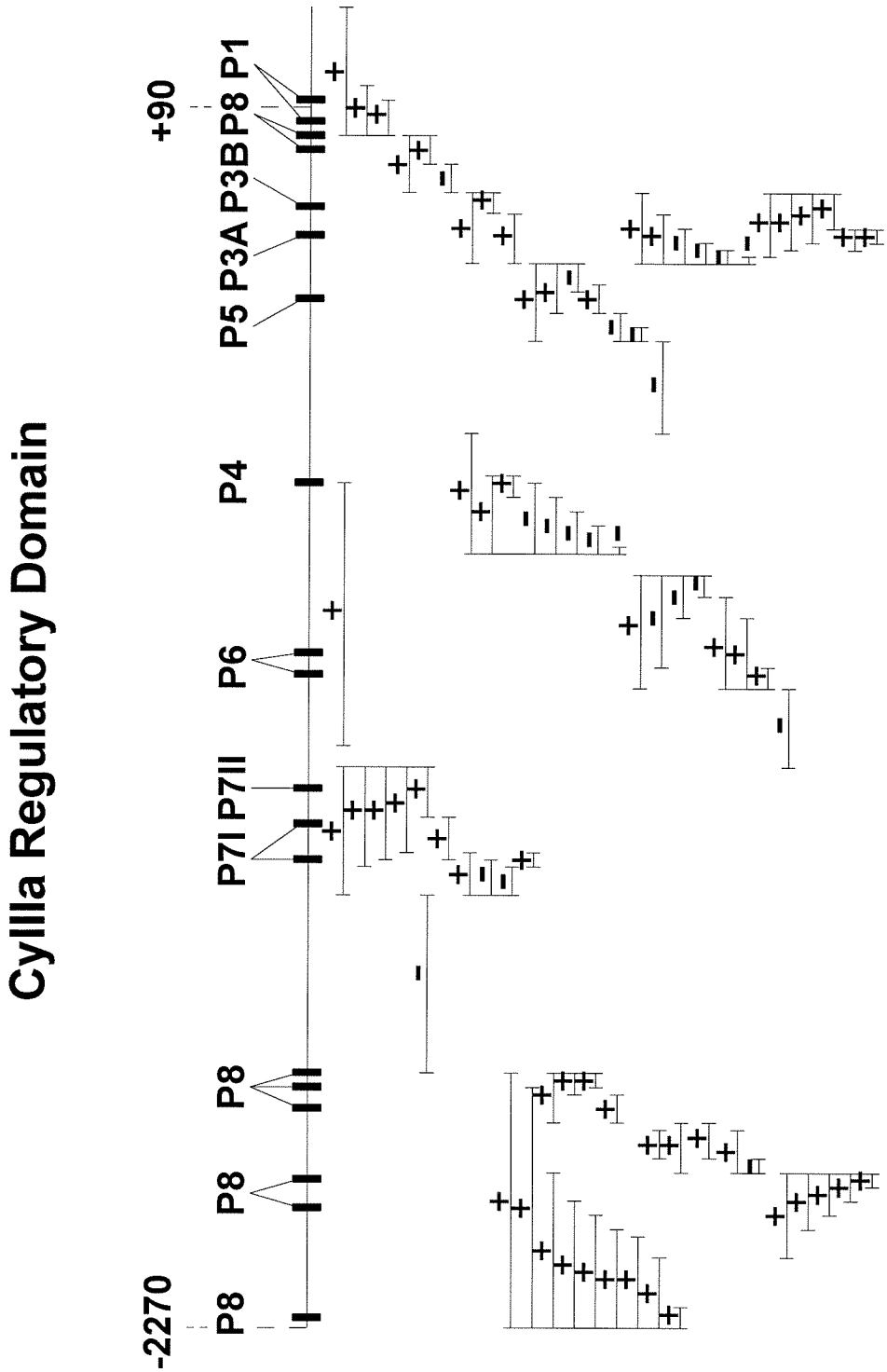
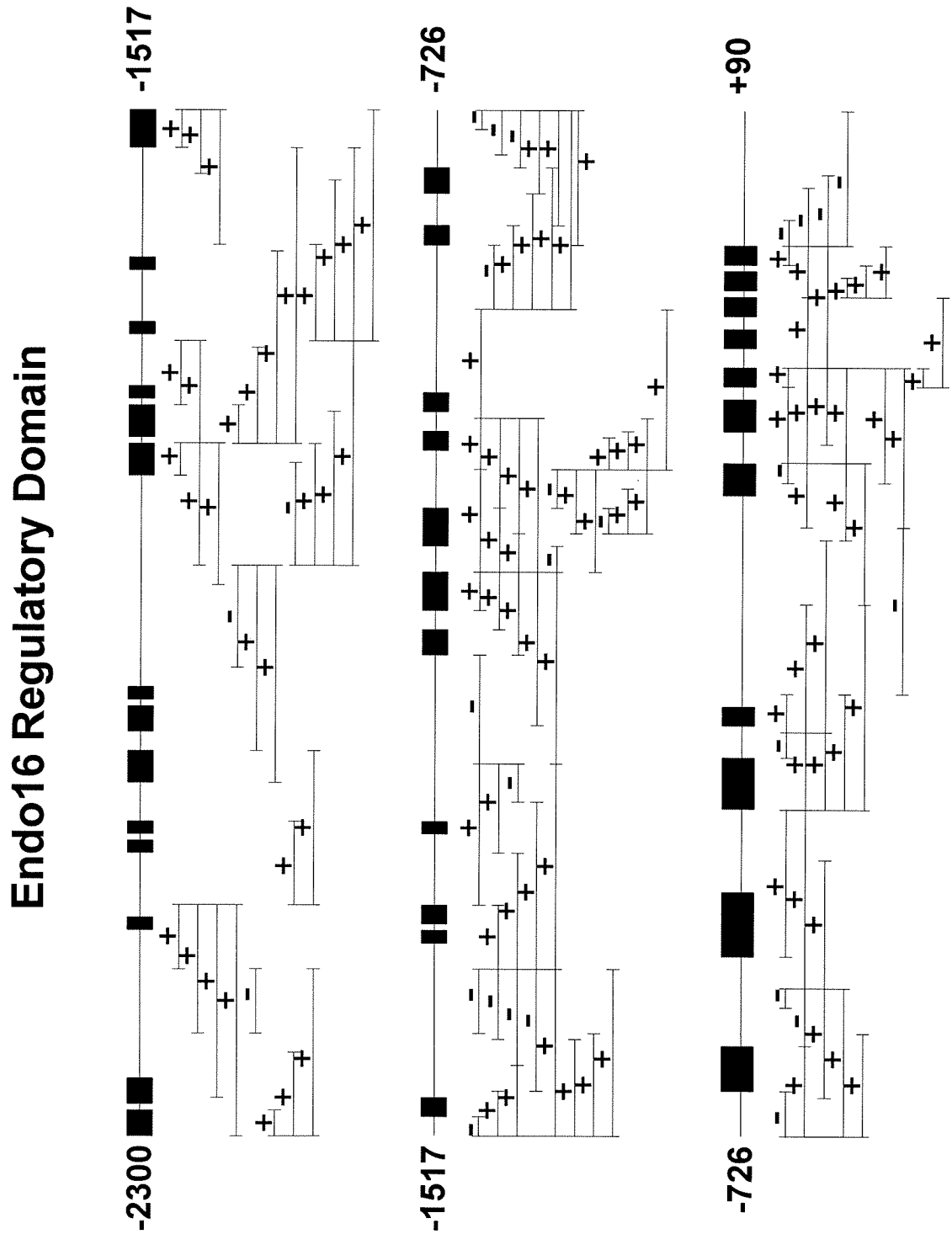

CylIIa Regulatory Domain
+90
-2270
P1
P8
P3B
P3A
P5
P4
P6
P7II
P7I
P8
P8
P8

Figure 1. Strategy for the rapid identification of DNA-binding protein target sites. (B) Map of the Endo16 regulatory domain. The 2.3 kb Endo16 promoter was examined for DNA-binding protein targets as described for CyIIIa. The complex temporal and spatial patterns of expression of this gene is reflected in the number of transcription factor target sites. The Endo16 promoter contains 38 target sites which are serviced by about 14 different factors (Yuh *et al.*, 1994). This diagram is redrawn from Yuh *et al.* (1994).


Endo16 Regulatory Domain
-2300
-1517
-1517
-726
-726
+90

Chapter 2

Sp(G/C)F1, a sea urchin embryo transcription factor, exists as five nested variants encoded by a single mRNA

Robert W. Zeller, James A. Coffman, Michael G. Harrington, Roy J. Britten and Eric H. Davidson

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

ABSTRACT

Several *Strongylocentrotus purpuratus* gene *cis*-regulatory regions contain assymetric C_4 sequences which are core elements of target sites for a specific DNA-protein interaction. Blastula stage nuclear extract contains 5 proteins which specifically bind to these target sites, resulting in a characteristic pattern of complexes in gel mobility shift assays. We used automated affinity chromatography to purify a protein which binds to these sites and have isolated the corresponding cDNA. This protein, Sp(G/C)F1, is a novel sea urchin DNA-binding protein with no overall homology to proteins reported in the databases currently available. The DNA binding domain of this protein was identified by a deletion analysis. As demonstrated by protein translated *in vitro* and bacterial protein expression, a single Sp(G/C)F1 mRNA serves as a template for the synthesis of five DNA-binding polypeptides. We show that these five polypeptides are most likely produced by differential usage of a nested set of AUG start codons in the Sp(G/C)F1 cDNA and thus contain variable amounts of a proline-rich N-terminal domain. Since proline-rich regions often serve as transcriptional activation domains, the five Sp(G/C)F1 proteins apparently possess different "activation potentials."

INTRODUCTION

Assymetric G/C-rich target sites have been identified in the regulatory regions of several sea urchin genes. Although sea urchin nuclear proteins which recognize these target sites have been previously described, the corresponding genes for these factors have not been isolated. A positive-acting G_6 target site present in the *Lytechinus pictus* LpS1 *cis*-regulatory domain is bound by two different nuclear proteins (Xiang *et al.*, 1991). These proteins, termed "G-string" binding factors, are differentially localized in the embryo and have different DNA binding characteristics. They may be similar to the mammalian IF1 factor which binds G_7 target sites in the $\alpha 1$ and $\alpha 2$ collagen promoters (Karsenty and Crombrugghe, 1991). A third DNA-binding protein recognizing G/C-rich target sites, suGF1, has been recently isolated from nuclear extract of the sea urchin *Parachinus angulosus* (Hapgood and Patterson, 1994; Patterson and Hapgood, 1994). This protein has a high affinity for poly-G targets sites found in the histone genes, and will also recognize the G_6 target site present in the LpS1 promoter. If a nucleosome core is positioned *in vitro* over the G-rich target site, the suGF1 protein is prevented from binding DNA. The suGF1 protein may gain access to its recognition site shortly after replication and may help to maintain nucleosome-free DNA (Patterson and Hapgood, 1994).

In the *CyIIIa* and *Endo16 cis*-regulatory regions, sites previously referred to as "P8" or "P2" sites (hereafter referred to as Sp(G/C)F1 sites) typically contain a C_4 core element (5' CCCC 3') and are bound specifically by five proteins present in 24 hr blastula stage nuclear extract (Calzone *et al.*, 1988; Thézé e*t al.*, 1990; Coffman *et al.*, 1992; Yuh *et al.*, 1994). The relative positions of these sites, as well as sites for other DNA-binding proteins present in these promoters, are depicted in Fig. 1A. An interesting feature of these sites is that they often occur throughout the entire gene regulatory domain in a variety of patterns. In the *CyIIIa* cytoskeletal actin gene promoter, about 50% of the

Sp(G/C)F1 target sites are found in a cluster at the 5' end of the regulatory domain. Several target sites are also located in the central and 3' domains of the regulatory region. Previous *in vivo* site competition experiments demonstrated that Sp(G/C)F1 sites most likely serve as general positive regulators of *CyIIIa* (Franks *et al.*, 1990). There are 23 Sp(G/C)F1 binding sites in the *Endo16* promoter, scattered throughout the entire *cis*-regulatory domain (Yuh *et al.*, 1994). The function of the *Endo16* Sp(G/C)F1 target sites has not yet been determined. In these and other regulatory domains, Sp(G/C)F1 sites are often found in close proximity to target sites for other DNA binding proteins (Fig. 1A). This suggests that the factor(s) binding to the Sp(G/C)F1 sites are capable of interacting with each other and/or with other factors to influence gene expression.

Here we report the isolation of the Sp(G/C)F1 protein and the corresponding cDNA. This factor was initially identified as the 55 kD protein which gives rise to the slowest migrating complex in a 2-dimensional gel shift assay using an Sp(G/C)F1 target site probe (Coffman *et al.*, 1992). The 55 kD protein was purified from sea urchin nuclear extract by affinity chromatography, digested with a lysylendopeptidase and subjected to microsequence analysis. Nucleic acid probes designed from the resulting amino acid sequence were used to identify Sp(G/C)F1 clones present in an early embryo cDNA library. We show that a single Sp(G/C)F1 mRNA produces five polypeptides which share a common DNA-binding domain and C-terminus but differ at the N-terminus. A proline-rich region, often implicated as functioning as a transcriptional activation domain, is located in the N-terminal domain. The differential inclusion of this proline-rich region in the various forms of the Sp(G/C)F1 protein may have functional significance for Sp(G/C)F1 regulated gene expression in the embryo.

MATERIALS AND METHODS

Protein purification and microsequencing

The 55 kD form of the Sp(G/C)F1 protein was purified by tandem affinity chromatography and identified by gel mobility shift assays as reported by Coffman *et al.* (1992) from 24 hr *Strongylocentrotus purpuratus* embryo nuclear extract. The Sp(G/C)F1 site oligonucleotide sequence used to purify the protein was

5' ACCTCACCCTCCCCCCCCCCCCCTCC 3'

and corresponds to the oligonucleotide sequence pair 33/34 of Thézé *et al.* (1990). The procedure for generating peptides and obtaining amino acid sequence was described in detail by Char *et al.* (1993); here we present a brief protocol. The Sp(G/C)F1 protein eluted with the 0.6-0.8 *M* KCl fractions was precipitated with trichloroacetic acid and alkylated with 4-vinylpyridine as described by Tempst *et al.* (1990). After separation from other proteins by SDS-PAGE, the Sp(G/C)F1 protein was electroeluted from a gel slice and concentrated into ~100 µl final volume. To generate peptides, a lysylendopeptidase (*Achromobacter* protease I, Wako) was added to a final concentration of 0.1 ng/µl and incubated overnight at 37°C. Peptides were separated by reverse-phase HPLC and individually collected peaks were sequenced by automated Edman degradation. Several different peptide sequences were obtained and are presented in Table 1.

Isolation of the Sp(G/C)F1 cDNA

We used the Polymerase Chain Reaction (PCR) to isolate an Sp(G/C)F1 probe. Two degenerate oligonucleotide primers, corresponding to the terminal six or seven amino acids of the P8MPNE peptide (Table 1), were used to amplifiy the sequence

encoding the internal 18 amino acids of the P8MPNE peptide from genomic DNA template. The sequence encoding the internal 18 amino acids was used to design a non-degenerate oligonucleotide probe which was subsequently used to screen an embryo cDNA library. For PCR amplification, the "sense" oligo sequence was (*Bam*HI site underlined):

5' CGGGATCCCATGCCNAAYGARTTYCTNCAYCA 3'

and the "antisense" oligonucleotide sequence was (*Kpn*I site underlined):

5' GGGGTACCCAARTCRTTRTTRTANAGYTGCAT 3'.

Restriction sites were added to the ends of the primers to facilitate subcloning into an appropriate sequencing vector. The primers, 600 ng of *S. purpuratus* genomic DNA and TAQ DNA polymerase were added together in a 25 µl PCR reaction. This reaction contained 80 µ*M* each primer, 400 µ*M* each nucleotide, 5U TAQ; 30 m*M* Tricine, pH 8.4, 2 m*M* $MgCl_2$, 5 m*M* β-mercaptoethanol, 0.01% gelatin and 0.1% Thesit. After an initial denaturation step (3 min at 94°C) the reaction mixture was subjected to 30 amplication cycles (parameters: 94°C 1 min, 48°C 1 min, 72°C 1 min). A ~120 bp amplified fragment was gel purified, subcloned, sequenced and found to encode the P8MPNE peptide (this corresponds to nt 498-584 in the cDNA sequence). From this sequence a 48 nt non-degenerate oligonucleotide was designed:

5' CAGAATCCACAAAGTGAGCGACAGCTTTTCTACAACGACGTAGCCATG 3'

(this corresponds to nt 519-566 in the cDNA sequence). This oligonucleotide was used as a probe to screen 8 X 10^5 clones from a 4 hr embryo lambda ZAP cDNA library as described by Calzone *et al.* (1991). Three cDNA clones were isolated, the longest of which (3.3 kb) was completely sequenced. The other two clones were found to be shorter fragments of the longest clone. Sequencing templates were generated by either limited DNAse I digestion (Eberle, 1993) or by a transposon facilitated procedure (Strathman *et al.*, 1991). Sequence data was obtained from both standard ^{35}S dideoxy sequencing reactions (Sanger *et al.*, 1977) using Sequenase Version 2.0 or from dye-labeled primer

cycle sequencing reactions analyzed on an ABI 373A automated DNA sequencer according to the manufacturer's instructions. Sequences were assembled using the IBI-Pustell sequence analysis program. Database searches were performed using the BLAST and FASTA search programs (Altschul *et al.*, 1990; Pearson and Lipman, 1988).

Protein expression

The various Sp(G/C)F1 expression constructs were made in the pRSET series of vectors (Invitrogen, San Diego, CA) which insert a six histidine tag for nickel-NTA affinity chromatography (Hochuli *et al.*, 1987) at the N-terminus of the protein. An epitope tag, the T7 tag, is also inserted between the six histidine residues and the vector polylinker region. This epitope is recognized by a monoclonal antibody which in our experiments identifies the N-terminus of the expressed protein (Lutz-Freyermuth *et al.*, 1990; Tsai *et al.*, 1992). We used PCR to amplify the entire Sp(G/C)F1 coding region for subsequent insertion into the expression vector. This was necessary because no restriction enzyme sites were present upstream of the initiation codon. A "sense" oligonucleotide primer was designed to incorporate a *Bam*HI site (underlined) just upstream of the Sp(G/C)F1 initiation codon: (double underlined)

5' CGCGGATCCGATGTCCACTCTGCCCCAGCCC 3'.

The "antisense" oligonucleotide primer corresponded to a T3 RNA polymerase site flanking the cDNA vector polylinker. These two oligonucleotides, the cDNA clone template and Vent DNA polymerase (NEB), were mixed in a final volume of 25 µl and subjected to 25 PCR cycles (exact reaction conditions: 1X Vent buffer (NEB), 100 µ*M* each nucleotide, 1 µ*M* each oligonucleotide, 8 m*M* $MgSO_4$; 10 ng supercoiled plasmid template). The cycling conditions consisted of an initial 3 min denaturation at 96¡C followed by 25 cycles of 96°C for 30 sec, 60°C annealing for 30 sec and 2 min extension at 72°C. The resulting PCR product was digested with *Bam*HI and *Bgl*II to release a

~1.7 kb fragment containing the entire Sp(G/C)F1 coding region and about 175 nt of 3' untranslated region (the *Bgl*II site is located at nt 1548 in the cDNA sequence). This fragment was subcloned into *Bam*HI/*Bgl*II digested pRSET vector to create the full-length Sp(G/C)F1 expression construct. Subsequent Sp(G/C)F1 expression clones were derived from the full-length clone, using both restriction enzyme digestion and PCR as necessary.

Bacterial protein expression was induced by the addition of 1 m*M* IPTG (final concentration) to a log phase culture of *E coli* strain pLys-S transformed with the Sp(G/C)F1 pRSET plasmid as described by Studier *et al.* (1986). The full-length Sp(G/C)F1 protein is expressed in inclusion bodies which were purified by the method of Lin and Cheng (1991). To refold the protein, Sp(G/C)F1 inclusion bodies (collected from 0.25 liter of bacterial culture) were dissolved in 2.5 ml of 8 *M* urea, 20 m*M* HEPES, pH 7.9, 0.1 m*M* EDTA, 1 m*M* DTT, 1 m*M* $MgCl_2$ and incubated on ice for 1 hr. Next, 2.5 ml of renaturation buffer (10 m*M* HEPES, pH 7.9, 0.1% NP-40, 1 m*M* DTT, 100 m*M* KCl, 10% glycerol) were added, and the preparation was mixed and incubated on ice for an additional hour. Next, 5 ml of renaturation buffer was added and incubated for two additional hours on ice. After bringing the volume to 30 ml with renaturation buffer, the mixture was slowly stirred at 4°C overnight. The protein was concentrated in Centricon centrifugation concentrators and stored in aliquots at -70°C. Based on the mass of the refolded Sp(G/C)F1 protein (data not shown) and quantitative DNA-binding measurements (see below), we estimated a recovery of 0.1% DNA-binding activity in the refolded protein. This restoration of activity is similar in level to what has been observed for refolded, recombinant SpP3A2 protein (Calzone *et al.*, 1991).

Capped or uncapped mRNA for *in vitro* translation templates was made from T7 mMESSAGE mMACHINE or MEGAscript kits (Ambion, Austin, TX) using linearized plasmid DNA templates. Protein translated *in vitro* was made using the IVT kit

according to the manufacturer's instructions (Ambion, Austin, TX) in 25 µl reaction volumes.

Gel mobility shift assays

Enhanced mobility shift assays were performed essentially as described by Calzone *et al.* (1988). Sp(G/C)F1 target site probes used in gel shift reactions included the 33/34 oligonucleotide pair from the *CyIIIa* regulatory region (Thézé *et al.*, 1990) and a 61 bp cloned fragment of a second Sp(G/C)F1 site in the *CyIIIa* promoter (nt positions -2027 to -1982 in Thézé *et al.*, 1990). Reaction volumes were typically 20 µl and contained 0.2-5 µg poly dI-dC as nonspecific competitor. For recombinant protein gel shifts, up to 4 µl of refolded protein was used with or without nonspecific competitor DNA, depending on the assay. Quantitative gel shifts were analyzed as described by Calzone *et al.* (1988, 1991). In these experiments, P_0 is the concentration (mole liter^{-1}) of protein molecules able to bind Sp(G/C)F1 target sites and K_r measures the preference the protein displays for Sp(G/C)F1 target sites, compared with its affinity for non-specific DNA. The amounts of free and protein-complexed probe were used to calculate the values of P_0 and K_r by a least-squares fit of the data to the equations presented in Calzone *et al.* (1988, 1991). These values are reported in Table 2.

Affinity purified sea urchin Sp(G/C)F1 protein from the 0.6 *M* KCl eluted column fractions was used at 0.025-0.05 µl per 20 µl reaction (about 0.25 n*M* final concentration). Gel shifts using protein translated *in vitro* contained from 0.25-2 µl of protein per 20 µl reaction (about 0.25 n*M* final protein concentration). Procedures for two-dimensional gel shifts were described by Coffman *et al.* (1992). For supershifting experiments, antibody was added directly to a 10 µl gel shift reaction using the 33/34 oligonucleotide probe and incubated on ice for 10-15 min before gel loading. Up to 1 µl

of either the monoclonal T7 tag antibody (Novagen, Madison, WI) or the polyclonal Sp(G/C)F1 antibody was used in each reaction.

Antibody procedures

A carboxy terminal portion of the Sp(G/C)F1 protein (amino acids 253-486) was expressed in *E. coli* and isolated by nickel-NTA chromatography as previously described (Zeller *et al.*, 1994a). This served as an antigen for the generation of rabbit antibodies by Cocalico Biologicals (Reamstown, PA). Procedures for SDS-PAGE and immunoblotting were described by Zeller *et al.* (1994a).

Nucleic acid blots and RNAase protection assays

Nucleic acids were digested, blotted to nylon membranes (MSI Inc., Westboro, MA) and hybridized at 65°C with 10^6 CPM probe/ml in 5X Denhardt's, 6X SSPE, 0.5% SDS, 50 mg/ml salmon sperm DNA (1X Denhardt's is 0.2 mg/ml each of polyvinylpyrrolidone, bovine serum albumen and Ficoll 400; 1X SSPE is 150 m*M* NaCl, 1 m*M* Na_2EDTA, 150 m*M* NaH_2PO_4, pH 7.4). The blot was washed at high stringency at 65¡C in a final wash of 0.1X SSPE and 1.0% SDS. The random primed, ^{32}P-labeled probe corresponded to either the entire Sp(G/C)F1 coding region or a portion of the cDNA clone (nt 1017-1740). Quantitative RNAase protection assays were done according to Lee *et al.* (1986) using duplicate 50 μg samples of total embryo RNA for each time point. The antisense riboprobe for these measurements was labeled with ^{32}P-UTP and corresponds to positions 1510-1743 in the cDNA clone sequence. The amount of total RNA per sea urchin embryo is 2.8 ng (Davidson, 1986). The formulas used to determine transcript prevalence per embryo were those of Lee *et al.* (1986).

RESULTS

Isolation of the Sp(G/C)F1 protein and cloning of the corresponding cDNA

The Sp(G/C)F1 target site occurs at least 38 times in the *CyIIIa* and *Endo16* promoters (Calzone *et al.*, 1988; Thézé *et al.*, 1990; Yuh *et al.*, 1994; Fig. 1A). These sites were aligned and used to derive a consensus sequence for the region surrounding the assymetric core CCCC, as shown in Fig. 1B. Requirements for particular nucleotides are relatively stringent at only 10 out of 19 positions outside of the CCCC core, but even so, when taken together, the allowable sequence has a low probability of random occurrence.

The 55 kD Sp(G/C)F1 protein which binds this site is one of the most prevalent DNA-binding proteins in 24 hr nuclear extract (Calzone *et al.*, 1988; Coffman *et al.*, 1992). When bound to an Sp(G/C)F1 target site, the 55 kD protein forms the slowest migrating complex observed in gel shift reactions (Coffman *et al.*, 1992). Affinity chromatography was used to purify the Sp(G/C)F1 protein, most of which eluted in the 0.6-0.9 *M* KCl fractions, as reported earlier by Coffman *et al.* (1992). Over 50% of the Sp(G/C)F1 protein present in 24 hr nuclear extract (representing about 2 X 10^9 embryos) was recovered after automated affinity chromatography, resulting in a 55-fold enrichment. After concentration, alkylation (Tempst *et al.*, 1990) and gel purification, the protein was digested with a lysylendopeptidase (*Achromobacter* protease I). This protease specifically cleaves the carboxy terminal of lysine residues and generally produces peptides of greater length than those generated by trypsin digestion. Long peptides are of particular value when trying to use amino acid sequence as a basis for designing specific probes for subsequent gene isolation. The resulting peptides were separated by reverse phase HPLC, and sequenced by automated Edman degradation. Several peptide sequences were obtained that were of sufficient length and amino acid

composition for use in designing specific probes for the isolation of the Sp(G/C)F1 cDNA, and these are presented in Table 1.

In order to avoid the marginal hybridization signals we often obtain when screening phage libraries with degenerate oligonucleotide probes, we designed primers for PCR to prepare a non-degenerate Sp(G/C)F1 probe. Degenerate primers corresponding to the terminal 6-7 amino acids on both ends of the complete P8MPNE peptide (Table 1) were used to amplify the nucleotide sequence encoding the internal 18 amino acids of this peptide from genomic DNA. The amplified fragment (~120 bp) was sequenced and found to encode the P8MPNE peptide (nt 498-584 of Fig. 2). A non-degenerate oligonucleotide was then designed (nt 519-566 of Fig. 2) and used as a probe to screen 8 X 10^5 clones from a 4 hr sea urchin embryo cDNA library. Three clones were isolated, the longest of which was 3.3 kb in length. The complete double-stranded sequence of this clone is shown in Fig. 2. The two smaller clones were partially sequenced and found to be fragments of the longest clone.

Sequence features of Sp(G/C)F1 cDNA

The longest open reading frame in the Sp(G/C)F1 cDNA encodes a protein of 486 amino acids with a calculated mass of 55 kD, the same mass observed for the sea urchin protein (Coffman *et al.*, 1992; Fig. 2). The predicted amino acid sequence shows no significant homologies with any known proteins from several different sequence databases (PIR, GenPept, Swiss Prot). The amino terminal region contains nearly half of all the methionine residues present in the Sp(G/C)F1 protein (10/21 residues). The first methionine in the predicted ORF does not conform to an initiation consensus sequence (Kozak, 1991; Cavener and Ray, 1991), but this codon is preceded by four in-frame stop codons. The 486 amino acid ORF thus represents the longest possible protein which could be translated from this cDNA. As discussed below, four additional methionines

which encode Sp(G/C)F1 polypeptides are also present in the amino terminal region at positions 34, 102, 125/128 and 153. These methionine codons also lack initiation consensus sequences, a feature that becomes important in the analysis of Sp(G/C)F1 translation described below.

The 105 amino terminal residues contain 69% of the prolines (22/32) found in the protein, and this results in a proline concentration of 21% in this region. Proline residues are often implicated in certain classes of transcriptional activation domains (Mitchell and Tjian, 1989), and this region could constitute an Sp(G/C)F1 activation domain. Many of the N-terminal methionine and proline residues are found in a pentapeptide repeat (N/S V S M P) that is present in seven tandem copies. We have not determined a function for this repeat, nor are similar repeats present in other proteins deposited in the current databases. The central region of the protein is generally devoid of prolines (2/32). It contains a strongly basic domain (residues 315-332) which could be involved in DNA binding, as discussed below. A potential heptad repeat of hydrophobic residues is present in several locations in the Sp(G/C)F1 protein. Two adjacent heptad repeats are located N-terminal to the strongly basic domain. The first repeat spans 22 amino acids (residues 234-255) and is immediately followed by a repeat spanning 63 amino acids (residues 256-318). A three heptad repeat domain is located at the N-terminus (residues 13-34) and is included only in the 55 kD form of the protein Sp(G/C)F1 protein. Hydrophobic heptad repeats are commonly found in proteins which adopt a "coiled-coil" structure such as myosin (McLachlan and Karn, 1983; LeBlanc and Leinwand, 1991) and in the leucine zipper class of DNA-binding proteins (Landschulz *et al.*, 1988).

The Sp(G/C)F1 gene encodes five nested proteins interacting with Sp(G/C)F1 DNA target sites

To ascertain whether the Sp(G/C)F1 protein is identical to the proteins in sea urchin nuclear extract which bind the G/C-rich target sites, we generated polyclonal antibodies specific to the carboxy terminus of the recombinant protein. A fragment of the cDNA clone encoding amino acids 253-486 was expressed in bacteria using the pRSET series of expression vectors. The bacteria containing the recombinant protein were lysed and the lysate passed over nickel-NTA columns for purification (Hochuli *et al.*, 1987). In this manner we were able to purify the carboxy terminal Sp(G/C)F1 polypeptide to an estimated 95% purity, as judged from silver stained protein gels (data not shown).

We used the anti-C-terminal Sp(G/C)F1 antibody to examine sea urchin nuclear proteins affinity purified with the Sp(G/C)F1 binding site. The affinity column efficiently extracts nuclear proteins interacting with this target site. An aliquot of the 0.6 *M* KCl-eluted fraction was separated on a standard SDS protein gel, transferred to nitrocellulose and incubated with the Sp(G/C)F1 antibody (see Materials and Methods). As shown in Fig. 3A, the antibodies detect five polypeptides with approximate masses of 55 kD, 51 kD, 45 kD, 42 kD and 40 kD, respectively, which correspond to the proteins shown previously to specifically bind to the Sp(G/C)F1 site (Coffman *et al.*, 1992). Thus the anti-C-terminal Sp(G/C)F1 antibody recognizes the sea urchin proteins which bind specifically to the Sp(G/C)F1 site *in vitro*. We therefore believe these are the same proteins which occupy Sp(G/C)F1 target sites *in vivo*.

The five polypeptides recognized by the anti-C-terminal antibody described above could be either Sp(G/C)F1 degradation products resulting from the preparation of nuclear extract or different variants of the Sp(G/C)F1 protein. To distinguish between these possibilities, we made Sp(G/C)F1 protein *in vitro* and used this protein in a two-dimensional gel shift assay described by Coffman *et al.* (1992). In this assay, complexes formed in a gel mobility shift reaction are separated in a tube gel (the first dimension). After separation, the tube gel is removed and placed horizontally across the top of an SDS-protein gel. Electrophoresis in this gel (the second dimension) separates the

polypeptide components from DNA-protein complexes. From this type of mobility shift, it is possible to obtain the mass of a protein binding to a specific target site, and also to identify DNA-protein complexes formed by multiple polypeptide species. Complexes resulting from a protein monomer or homomultimer binding to a DNA target site will be detected as a single band in the two-dimensional gel shift. Complexes formed by protein heteromultimers are detected as a series of *vertical* bands.

An mRNA made directly from the Sp(G/C)F1 cDNA clone was used as the template to synthesize protein *in vitro*. When this protein was used in a two-dimensional gel shift, five major bands were observed in a *diagonal* pattern as shown in Fig. 3B. The arrangement of these bands indicates that each DNA-protein complex is formed by a single protein species binding to a DNA probe. The estimated sizes of these polypeptides are 55 kD, 50 kD, 43 kD, 40 kD and 37 kD, respectively, and are thus in close agreement with the sizes observed for affinity purified Sp(G/C)F1 proteins reported above and by Coffman *et al.* (1992). By this type of assay, we are not able to determine if an Sp(G/C)F1 protein monomer or homomultimer is forming the DNA-protein complex. As we describe elsewhere, we have shown by quantitative gel shift assays that Sp(G/C)F1 binds DNA cooperatively, most likely as a dimer (Zeller *et al.*, 1994b). The two important conclusions from the two-dimensional gel shift reported here are that five polypeptides are produced from a single Sp(G/C)F1 mRNA; and that the DNA-protein complexes formed by these polypeptides accurately reproduce the pattern of complexes previously observed for the Sp(G/C)F1 proteins in nuclear extract (Coffman *et al.*, 1992). In recombinant Sp(G/C)F1 protein preparations the variants are present in the ratios 37 kD, 32.8; (40+43 kD), 22.1; 50 kD, 22.4 and 55 kD, 22.7.

The multiple Sp(G/C)F1 polypeptides contain different N-terminal sequences but common C-terminal domains

As previously mentioned, the N-terminus contains a number of methionines which could serve as translation start sites. In fact, polypeptides of the observed masses could be encoded by proteins which begin with methionines at amino acid positions 1, 34, 102, 125/128 and 153, respectively (see Fig. 2). As demonstrated above, all five polypeptides are capable of complexing an Sp(G/C)F1 target site. This suggests the Sp(G/C)F1 DNA binding domain is located in the middle of the protein and the differences between the various polypeptides must therefore be localized to either the N- or C-terminal domains. To determine the relationship between the Sp(G/C)F1 polypeptides capable of forming complexes with DNA, we used antibodies specific to either the N- or C-terminal domains of the recombinant protein in gel mobility supershift assays. One bacterially expressed Sp(G/C)F1 protein contains a T7 tag epitope upstream of the natural initiation codon, which is specifically recognized by a commercially available monoclonal antibody (see Materials and Methods). To identify the C-terminal of the polypeptides, we used the same polyclonal antibody employed in the immunoblot described, which recognizes epitopes lying within amino acids 253-486.

The recombinant protein we used for this experiment is expressed in an inactive form in inclusion bodies. DNA-binding activity was restored to the Sp(G/C)F1 protein by slowly refolding the protein from purified and denatured inclusion bodies as detailed in Material and Methods. The recombinant Sp(G/C)F1 protein was incubated in a 10 µl gel shift reaction, using the 33/34 oligonucleotide pair as a probe. Increasing amounts of either antibody were added to the gel shift reactions and allowed to bind to DNA-protein complexes. Both antibodies used here retard the mobility of the complexes as shown in Fig. 4 ("supershift"), but the patterns of the supershifts are different. The T7 tag antibody, which specifically recognizes the N-terminal of the protein, only supershifts the slowest migrating band (Fig. 4, lanes 3-6). This is the 55 kD protein detected in the two-dimensional gel shift experiment (Fig. 3B). The shorter polypeptides, i.e., the faster migrating complexes, are not supershifted by this antibody, indicating they lack the T7

tag. Unlike the amino terminal antibody, the Sp(G/C)F1 specific C-terminal antibody supershifts all of the complexes (Fig. 4, lanes 7-10) indicating that all polypeptides contain the C-terminal of the Sp(G/C)F1 protein. These results imply that the multiple Sp(G/C)F1 polypeptides share a common DNA-binding domain and C-terminus but have different N-terminal sequences.

When the carboxy terminal antibody reacts with affinity purified sea urchin protein, not all the polypeptides are equally supershifted (Fig. 3B, lanes 12-15). The argument that all five polypeptides are probably Sp(G/C)F1 gene products rests on the observation that, as shown in Fig. 3A, the Sp(G/C)F1 antibody recognizes five polypeptides of the same molecular weights as are purified by binding to Sp(G/C)F1 target sites. These polypeptides also correspond in size to proteins present in Sp(G/C)F1 affinity column fractions detected by silver staining SDS-protein gels (Coffman *et al.*, 1992; data not shown). The failure to supershift all of the nuclear extract complexes in the experiment of Fig. 3B suggests that some unrelated protein might be present that co-migrates with the 43 kD Sp(G/C)F1 polypeptide. However, it is also possible that the polyclonal antibody may have difficulty recognizing some native forms of Sp(G/C)F1 protein that are present in the supershift reactions. Alternatively, the smaller polypeptides may be post-translationally modified in such a way as to mask the Sp(G/C)F1 epitope. We know that several other sea urchin DNA-binding proteins are differentially post-translationally modified (Wang *et al.*, 1994; M. Harrington, R. Zeller and E. Davidson, unpublished). Such a result would not be unexpected, since the antibody was generated against a recombinant protein which lacks most eukaryotic post-translational modifications. Evidence that the nuclear extract proteins are modified is that the observed molecular weights of the smaller Sp(G/C)F1 proteins in nuclear extract are slightly larger than their corresponding recombinant counterparts (e.g., 45 *vs* 43 kD, 42 *vs* 40 kD and 40 *vs* 37 kD). In any case, the important point from the supershift and immunoblot experiments is that there are at least five Sp(G/C)F1 protein species

generated from a single mRNA, which specifically bind to the Sp(G/C)F1 target site. These proteins differentially utilize a nested set of initiation codons and therefore contain different N-terminal domains.

We measured the affinity of each Sp(G/C)F1 polypeptide for its specific DNA target site in a probe-excess titration gel shift. In this experiment, we measure two parameters: P_0, which is the concentration of protein molecules able to bind DNA; and K_r, which is the ratio of the equilibrium constants for specific (K_s) and non-specific (K_n) DNA binding (Emerson *et al.*, 1985; Calzone *et al.*, 1988). The concentration of specific DNA target sites is defined as D_s and the amount of non-specific sites is defined as D_n. The concentration of protein bound in complexes with specific DNA is defined as PD_s. The amount of specific complex present at equilibrium is described as follows:

$$PD_s = \frac{P_0 K_r D_s}{D_n + K_r D_s} \qquad (1)$$

Using a constant amount of protein, we measured the amounts of free and Sp(G/C)F1-complexed probe for several different concentrations of Sp(G/C)F1 probe. To obtain values for P_0 and K_r, we performed a least-squares fit of the data to equation (1). Results are listed in Table 2. The K_r values for each Sp(G/C)F1 peptide are very similar to each other (8.55×10^4 to 2.14×10^5). The complexes formed by the 40 and 43 kD polypeptides are not easily resolved in this experiment; for that reason we combined these two complexes for analysis. If we sum all four Sp(G/C)F1 complexes and recalculate K_r, we obtain a K_r of 1.22×10^5. This agrees well with the K_r ($\sim 1 \times 10^5$) reported by Calzone *et al.* (1988) for the summation of the five nuclear extract Sp(G/C)F1 complexes formed with the *CyIIIa* P2 target site probe.

Determination of the DNA binding domain

As indicated from the gel mobility shift results, the Sp(G/C)F1 DNA-binding domain is most likely positioned in the middle or carboxy third of the protein, i.e., the regions common to all of the various forms. To determine the region responsible for DNA-binding, we made a series of Sp(G/C)F1 expression constructs for use in gel mobility shift assays (Fig. 5A). A combination of PCR and standard subcloning techniques were used to create N-terminal, C-terminal and internal deletions from the original full-length Sp(G/C)F1 expression clone. Protein from each mutant was made from an *in vitro* translation reaction and used in a standard gel shift assay using the 33/34 probe as depicted in Fig. 5B. In each case, specific DNA-binding was tested by competing the complex with a 50-fold molar excess of unlabeled probe. While not intended as an exhaustive set of protein mutants, these constructs did provide us with a rough estimate of the DNA-binding domain. The results are summarized as follows: Protein mutants with amino terminal deletions beyond amino acid 143 (Fig. 5B, lanes 3-9) or carboxy terminal deletions before amino acid 353 did not bind DNA (Fig. 5B, lanes 24-31). This excludes most of the proline containing N-terminal region. Three mutants with internal deletions of 31-47 amino acids each also failed to bind DNA (Fig. 5B, lanes 10-15). The smallest protein mutant with DNA-binding activity spanned the central half of the protein from amino acids 143-353 (Fig. 5B, lanes 36-37). The previously described basic domain is included in this region and most likely participates in DNA binding (see Fig. 2).

Sp(G/C)F1 is a single copy gene

We have already demonstrated that five proteins present in nuclear extract bind the Sp(G/C)F1 target site. We have also shown that these proteins can be produced from a single Sp(G/C)F1 mRNA template. It is therefore important to determine if the Sp(G/C)F1 gene is a member of a related gene family in which different proteins might

be produced by different family members. We examined the Sp(G/C)F1 gene copy number by hybridizing *S. purpuratus* genomic DNA with a Sp(G/C)F1 specific probe. Restriction digested genomic DNA from three different male sea urchins was probed with a full-length Sp(G/C)F1 coding region probe, as detailed in Materials and Methods. Two restriction enzymes used to digest the genomic DNA cut the Sp(G/C)F1 cDNA at a single site (*Bam*HI and *Hind*III) while the third restriction enzyme (*Eco*RI) does not cut the cDNA. The resulting pattern of hybridized bands, shown in Fig. 6, demonstrates that Sp(G/C)F1 is a single copy gene. Individuals one and two probably have single *Bam*HI alleles (lanes 1 and 4) while individual three has two *Bam*HI alleles (lane 7). Individual one has a single *Eco*RI allele (lane 3) while individuals two and three have two *Eco*RI alleles (lanes 6 and 9). The pattern for the *Hind*III alleles is more complex. Individual two has a single *Hind*III allele (lane 5). Individuals one and three share a common *Hind*III allele (lanes 2 and 8) which is different from individual two's allele. Individual one probably has a second *Hind*III alllele not shared with individual three (lane 2).

Measurement of Sp(G/C)F1 transcript prevalence

The prevalence of Sp(G/C)F1 mRNA was examined by quantitative RNAase protection assays (Lee *et al.*, 1986). In these assays, samples of embryo RNA are hybridized with an excess of labeled antisense RNA probe. After digestion of single-stranded RNAs with RNAases, the remaining RNA hybrids are precipitated and the resulting amount of "protected" label is determined. This provides an estimate of the number of transcripts present in the RNA sample. To determine Sp(G/C)F1 prevalence, duplicate samples of ovary or embryo total RNA were hybridized with a ^{32}P-labeled probe from the 3' untranslated region of the Sp(G/C)F1 cDNA (nt 1510-1743). The number of mRNA molecules per mass of total embryo RNA was determined as described (Lee *et al.*, 1986). The number of Sp(G/C)F1 mRNA molecules per embryo is shown in

the graphs in Fig. 7. The mRNA is detected in sea urchin ovaries, and is present at about 5000 copies in the unfertilized egg. Assuming an equal cellular distribution, the amount of mRNA per cell steadily decreases until 72 hr post-fertilization, when there are only about one to two mRNAs per cell. On a "per embryo" basis, the amount of mRNA peaks at about 9 hr of development, when there are about 8000 molecules per embryo. The levels of Sp(G/C)F1 binding activity previously measured in 7 and 24 hr nuclear extract (Calzone *et al.*, 1988) can easily be accounted for by new protein synthesis, given the levels of Sp(G/C)F1 mRNA per embryo, and the rate of sea urchin embryo protein synthesis (Davidson, 1986). Thus Calzone *et al.* (1988) reported that there are approximately 2 X 10^6 molecules of what we know now is Sp(G/C)F1 protein. This amount of protein can be synthesized from 5000 mRNAs in a little over 3 hr (two proteins $mRNA^{-1}$ min^{-1} and 5000 mRNA = 6 X 10^5 proteins hr^{-1} or about 3.3 hr for 2 X 10^6 proteins).

DISCUSSION

The Sp(G/C)F1 target site appears in the *cis*-regulatory regions of several different sea urchin genes. We used affinity chromatography to isolate a novel protein which specifically binds to this site and have cloned the corresponding cDNA. Of the eight sea urchin nuclear proteins so far isolated by affinity chromatography using *CyIIIa* target sites, three are novel DNA-binding proteins unrelated to any other known proteins (e.g., Calzone *et al.*, 1991; Coffman and Davidson, 1994). In addition to Sp(G/C)F1 these are SpP3A2 (Calzone *et al.*, 1991) and a protein which specifically binds to the P7II site (Hough-Evans *et al.*, 1990; Calzone *et al.*, 1988; Coffman and Davidson, 1994). Both are negative regulators of spatial gene expression, while Sp(G/C)F1 appears to act positively in this gene.

Differential translation produces a family of Sp(G/C)F1 proteins

Gel shift reactions in which embryo nuclear extract is reacted with a synthetic double-stranded oligonucleotide bearing a single Sp(G/C)F1 target site typically show four to five DNA-protein complexes. It has been unclear whether these bands were due to the binding of different proteins or protein degradation products. We demonstrate here that these four to five complexes are due to discrete protein products of the single copy Sp(G/C)F1 gene. Both protein translated *in vitro* from synthetic mRNA and bacterially expressed Sp(G/C)F1 include five polypeptide species originating from the single mRNA template. The five Sp(G/C)F1 proteins that are present in nuclear extract have sizes similar to the recombinant polypeptide species and exhibit patterns of DNA-protein complexes identical to those observed for recombinant Sp(G/C)F1. We therefore believe that Sp(G/C)F1 polypeptides present in nuclear extracts are also produced from a single mRNA. By using N-terminal and C-terminal specific antibodies, we demonstrated that the different Sp(G/C)F1 forms share common C-terminal sequences but differ in the N-terminal residues. The amino terminal domain of the Sp(G/C)F1 protein contains five methionines which may act as initiation codons during translation. Methionines at positions 1, 34, 102, 125/128 and 153 will produce Sp(G/C)F1 proteins close to the masses observed in a recombinant protein preparation. Since strong initiation consensus sequences do not exist at any of these N-terminal methionine residues, it seems likely that the various forms of Sp(G/C)F1 are produced by a "ribosome scanning" mechanism (Descombes and Schibler, 1991). In this situation, ribosomes which fail to initiate translation at the first start codon may continue down the mRNA and begin translation at a downstream start codon. The absence of a strong initiation consensus sequence surrounding the initial start codon may contribute to the operation of this mechanism in Sp(G/C)F1 mRNA.

This interesting phenomenon is not unique to Sp(G/C)F1. Alternate forms of several other DNA binding proteins are now known to be expressed in a similar manner such as C/EBP (Ossipow *et al.*, 1993) and the antagonistic pair LIP and LAP (Descombes and Schibler, 1991). These three proteins are members of the leucine zipper class of DNA binding proteins first described by Landschulz *et al.* (1988). C/EBP is expressed in a restricted set of cell types including adipocytes and hepatocytes. A potent transcriptional activation domain is present at the N-terminus of the dominant (longest) form of the protein. The shorter form of the protein, produced from the same mRNA, lacks the N-terminal domain and thus fails to activate transcription in target genes. Like C/EBP, LIP and LAP have similar structural features. The LAP protein is the dominant (longest) form and has a potent activation domain at the N-terminal. The LIP protein, produced from the same mRNA, is shorter and lacks the N-terminal activation domain. The shorter forms of these proteins can act as transcriptional repressors by competing with the longer proteins for DNA target sites. The lack of a transactivation domain on the short protein forms will thus inhibit gene transcription. Similar, alternative regulatory functions may exist among the different Sp(G/C)F1 proteins in the sea urchin embryo.

Sp(G/C)F1 proteins probably have differential transcriptional activation potential.

There are several classes of transcriptional activation domains, each categorized by particular amino acid composition and sequence organization (see Mitchell and Tjian, 1989 for review). One class is characterized by a high proline content, sometimes including homopolymeric proline stretches within the activation domain (Gerber *et al.*, 1994). Examples of proline-rich activation domains are found in several transcription factors including CTF/NF-I (Mermod *et al.*, 1989) and C/EBP (Friedman and McKnight, 1990; Pei and Shih, 1991). The C-terminal 100 amino acid CTF/NF-I activation domain has a 25% proline content, and partial deletions of this region suggest that this activation

domain is partially redundant. The N-terminal region of C/EBP contains two activator domains, one of which is proline rich. This 45 amino acid domain has a stretch of eight homopolymeric prolines as well as an additional 14 proline residues for an overall proline content of 49%. Point mutations introduced into the proline octamer changing up to four prolines into serines did not appreciably effect transactivation activity (Pei and Shih, 1991). Thus the activation potential of proline-rich domains may depend on overall proline content rather than on local regions within the domain.

Like C/EBP and CTF/NF-I, the sea urchin Sp(G/C)F1 protein contains an N-terminal proline-rich region. The average proline content is 21% over the N-terminal 105 amino acids. In addition, there is a second proline-rich region in the C-terminal domain of the protein (8% over 113 amino acids). If the N-terminal proline-rich domain serves as a transcriptional activation region, then each of the five major Sp(G/C)F1 species would likely have different "potentials" for transcriptional activation. If the longest protein is considered to have 100% of the activation domain (22/22 residues), it might be expected to have the largest "activation potential." The smaller Sp(G/C)F1 proteins will include less of the proline-rich domain and thus may have less ability to activate transcription. The amounts of the proline-rich domain present in the other Sp(G/C)F1 forms are, relative to the 55 kD protein, 77% (17/22), 9% (2/22), 0% (0/22) and 0% (0/22) for the 50 kD, 43 kD, 40 kD, and 37 kD proteins respectively. Thus only the 55 and 50 kD proteins retain significant N-terminal proline-rich domains. Unlike the two larger Sp(G/C)F1 proteins, the three smaller forms (43-37 kD) include less than 10% of the proline-rich region. The three smallest Sp(G/C)F1 proteins could serve as competitive repressors, by occupying sites which would otherwise be serviced by the 55 or 50 kD forms.

The function of Sp(G/C)F1 proteins in sea urchin gene regulatory domains

Several sea urchin gene *cis*-regulatory domains, expressed in non-overlapping embryonic territories, have been found to contain Sp(G/C)F1 target sites (Calzone *et al.*, 1988; Thézé *et al.*, 1990). The affinity of the Sp(G/C)F1 protein for these sites has been measured, and is here described by the dimensionless term K_r (Calzone et al., 1988; K_r is the ratio of the equilibrium constant for the interaction of the protein with its specific target site, divided by the equilibrium constant for the interaction of the protein with non-specific DNA). A large K_r value implies a high affinity for a specific DNA target site. The Sp(G/C)F1 target sites may be placed into two classes: a "strong" class consisting of long stretches of overlapping or extended core sequences made up of C residues, typically ≥ 10 (C_{10}), which display K_r values of ~10^6; and a "weak" class (C_4) that displays K_r values of ~10^5 (Calzone et al., 1988; Thézé *et al.*, 1990). Both of these types of site have K_r values that require highly specific interactions with their target sites. Sp(G/C)F1 sites are found in clusters in the 5' region of the *CyIIIa cis*-regulatory domain, which is expressed in the aboral ectoderm (Calzone *et al.*, 1988; Thézé *et al.*, 1990), or they may be scattered throughout the entire regulatory domain, as is the case in the *Endo 16* gene, which is expressed in the vegetal plate and gut, as summarized in Fig. 1 (Yuh *et al.*, 1994). Similar target sites exist in three other known gene regulatory regions, *viz* those of *SM50* (Sucov *et al.*, 1987), *msp130* (Parr *et al.*, 1990) and *SM30* genes (Akasaka *et al.*, 1994). All three of these genes are expressed in the skeletogenic mesenchyme territory (Benson *et al.*, 1987; Harkey *et al.*, 1992; George *et al.*, 1991). Sp(G/C)F1 protein binding activity is present in eggs (Calzone *et al.*, 1994) and this factor is one of the most prevalent in 24 hr blastula nuclear extracts (Calzone *et al.*, 1988; Coffman *et al.*, 1992). Genes apparently serviced by Sp(G/C)F1 are expressed in many regions of the embryo. These observations suggest that this factor is present throughout the entire embryo, but this remains to be shown directly.

The main functional data for Sp(G/C)F1 sites comes from studies of *CyIIIa-CAT* reporter transgenes. In a promoter deletion experiment, Flytzanis *et al.* (1987) showed

that elimination of the 5' cluster of Sp(G/C)F1 sites sharply reduces *CyIIIa-CAT* expression. A similar result was observed by Franks *et al.* (1989) in a series of *in vivo* target site competition studies in which excess amounts of DNA-binding protein target sites were co-injected with *CyIIIa-CAT* reporter constructs into sea urchin zygotes. When fragments containing clusters of Sp(G/C)F1 target sites were co-injected with *CyIIIa-CAT*, the level of CAT activity was stoichiometrically reduced. If the Sp(G/C)F1 variants are in fact the proteins that bind to these sites *in vivo*, then at least in *CyIIIa*, a major role of at least some of the Sp(G/C)F1 variants is positive regulation of gene expression.

REFERENCES

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

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

Benson, S., Sucov, H., Stephens, L., Davidson, E. and Wilt, F. (1987) A lineage-specific gene encoding a major matrix protein of the sea urchin embryo spicule I. Authentication of the cloned gene and its developmental expression. *Dev. Biol.* **120**:499-506.

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

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

Cavener, D.R. and Ray, S.C. (1991). Eukaryotic start and stop translation sites. *Nuc. Acid Res.* **19**:3185-3192.

Char, B.R., Bell, J.R., Dovala, J., Coffman, J.A., Harrington, M.G., Becerra, J.C., Davidson, E.H., Calzone, F.J. and Maxson, R. (1993). SpOct, a gene encoding the major octamer-binding protein in sea urchin embryos: Expression profile, evolutionary relationships, and DNA binding of expressed protein. *Dev. Biol.* **158**:350-363.

Coffman, J.A. and Davidson, E.H. (1994). Regulation of gene expression in the sea urchin embryo. *J. mar. biol. Ass. U.K.* **74**:17-26.

Coffman, J.A., Moore, J.G., Calzone, F.J., Britten, R.J., Hood, L.E. and Davidson, E.H. (1992). Automated sequential affinity chromatography of sea urchin embryo DNA binding proteins. *Mol. mar. bio. and Biotech.* **1**:136-146.

Cutting, A.E., Höög, C.H., Calzone, F.J., Britten, R.J. and Davidson, E.H. (1990). Rare maternal mRNAs code for regulatory proteins that control lineage-specific gene expression in the sea urchin embryo. *Proc. Natl. Acad. Sci. USA.* **87**:7953-7957.

Davidson, E.H. (1986). *Gene Activity in Early Development.* Third Edition. Academic Press: Orlando, Florida.

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

Descombes, P. and Schibler, U. (1991). A liver-enriched transcriptional activator protein, LAP, and a transcriptional inhibitory protein, LIP, are translated from the same mRNA. *Cell* **67**:569-579.

Eberle, J.R. (1993). Generation of bi-directional deletions using DNAse I. *Biotechniques* **14**:408-411.

Emerson, B.M., Lewis, C.D. and Felsenfeld, G. (1985). Interaction of specific nuclear factors with the nuclease-hypersensitive region of the chicken adult β-globin gene: Nature of the binding domain. *Cell* **41**:21-30.

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

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

Friedman, A.D. and McKnight, S.L. (1990). Identification of two polypeptide segments of CCAAT/enhancer-binding protein required for transcriptional activation of the serum albumin gene. *Genes Dev.* **4**:1416-1426.

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

Gerber, H.-P., Seipel, K., Georgiev, O., Höfferer, M., Hug., M., Rusconi, S. and Schaffner, W. (1994). Transcriptional activation modulated by homopolymeric glutamine and proline stretches. *Science* **263**:808-811.

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

Harkey, M.A., Whiteley, H.R. and Whiteley, A.H. (1992). Differential expression of the msp130 gene among skeletal lineage cells in the sea urchin embryo- A 3-dimensional insitu hybridization analysis. *Mech. Dev.* **37**:173-184.

Hochuli, E., Dobeli, H. and Schacher, A. (1987). New metal chelate adsorbant selective for proteins and peptides containing neighboring histidine residues. *J. Chrom.* **411**:177-184.

Hough-Evans, B. R., Franks, R. R., Zeller, R. W., Britten, R. J., and Davidson, E. H. Negative spatial regulation of the lineage specific *CyIIIa* actin gene in the sea urchin embryo. *Development* **110**:41-50, 1990.

Karsenty, G. and de Crombrugghe, B. (1991). Conservation of binding sites for regulatory factors in the coordinately expressed a1 (I) and a2 (I) collagen promoters. *Biochem. Biophys. Res. Commun.* **177**, 538-544.

Kozak, M. (1991). Structural features in eukaryotic messenger RNAs that modulate the initiation of translation. *J. Biol. Chem.* **266**:9867-9870.

Landschulz, W. H., Johnson, P. F. and McKnight, S. L. The leucine zipper: A hypothetical structure common to a new class of DNA binding proteins. *Science* **240**:1759-1764.

LeBlanc, J.M. and Leinwand, L.A. (1991). The diversity of myosin-based contractile systems in eukaryotic cells. *Am. Zool.* **31**:514-521.

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. *J. Molec. Biol.* **188**:173-183.

Lin, K-h. and Cheng, S-y. (1991). An efficient method to purify active eukaryotic proteins from the inclusion bodies in *Escherichia coli*. *Biotechniques* **11**:748-753.

Lutz-Freyermuth, C., Query, C.C. and Keene, J.D. (1990). Quantitative determination that one of two potential RNA-binding domains of the A protein component of the U1 small nuclear ribonucleoprotein complex binds with high affinity to stem-loop II of U1 RNA. *Proc. Natl. Acad. Sci. USA.* **87**:6393-6397.

McLachlan, A.D. and Karn, J. (1983). Periodic features in the amino acid sequence of nematode myosin rod. *J. Mol. Biol.* **164**:605-626.

Mermod, N., O'Neill, E.A., Kelly, T.J. and Tjian, R. (1989). The proline-rich transcriptional activator of CTF/NF-I is distinct from the replication and DNA binding domain. *Cell* **58**:741-753

Mitchell, P.J. and Tjian, R. (1989). Transcriptional regulation in mammalian cells by sequence-specific DNA binding proteins. *Science* **245**:371-378.

Ossipow, V., Descombes, P. and Schibler, U. (1993). CCAAT/enhancer-binding protein mRNA is translated into multiple proteins with different activation potentials. *Proc. Natl. Acad. Sci. USA*. **90**:8219-8223.

Parr, B.A., Parks, A.L. and Raff, R.A. (1990). Promoter structure and protein sequence of msp130, a lipid-anchored sea urchin glycoprotein. *J. Biol. Chem.* **265**:1408-1413.

Patterson, D. and Hapgood, J. (1994). suGF1 binds in the major groove of its oligo(dG) oligo(dC) recognition sequence and is excluded by a positioned nucleosome core. *Mol. Cell. Biol.* **14**:1410-1418.

Pearson, W.R. and Lipman, D.J. (1988). Improved tools for biological sequence comparison. *Proc. Nat. Acad. Sci. USA*. **85**:2444-2448.

Pei, D. and Shih, C. (1991). An "attenuator domain" is sandwiched by two distinct transactivator domains in the transcription factor C/EBP. *Mol. Cell. Biol.* **11**:1480-1487.

Strathman, M., Hamilton, B.A., Mayeda, C.A., Simon, M.I., Meyerowitz, E.M. and Palazzolo, M.J. (1991). Transposon-facilitated DNA sequencing. *Proc. Nat. Acad. Sci. USA*. **88**:1247-1250.

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

Tempst, P., Link, A.J., Riviere, L.R., Fleming, M. and Elicone, C. (1990). Internal sequence analysis of proteins separated on polyacrylamide gels at the submicrogram level: Improved methods, applications and gene cloning strategies. *Electrophoresis* **11**:537-553.

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

Tsai, D.E., Kenan, D.J. and Keene, J.D. (1992). *In vitro* selection of an RNA epitope immunologically cross-reactive with a peptide. *Proc. Natl. Acad. Sci. USA*. **89**:8864-8868.

Wang, D.G.-W, Kirchhamer, C. V., Britten, R. J. and Davidson, E. H. (1994). SpZ12-1, a negative regulator required for spatial control of the territory-specific *CyIIIa* gene in the sea urchin embryo. Submitted.

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

Yuh, C.-H., Ransick, A., Martinez, P., Britten, R.J. and Davidson, E.H. (1994). Complexity and organization of DNA-protein interactions in the 5'-regulatory region of an endoderm-specific marker gene in the sea urchin embryo. *Mech. Dev.* **47**:165-186.

Zeller, R.W., Britten, R.J. and Davidson, E.H. (1994a). Developmental utilization of P3A1 and P3A2: Two proteins which recognize the same DNA target site in several sea urchin gene regulatory regions. Submitted.

Zeller, R.W., Griffith, J.D., Moore, J.G., Kirkhhamer, C.V., Britten, R.J. and Davidson, E.H. (1994b). A multimerizing transcription factor of sea urchin embryos capable of looping DNA. Submitted.

Table 1: Sp(G/C)F1 polypeptides obtained from amino acid sequencing.

Peptide sequence	Position in predicted amino acid sequence
P8MPNE: MPNEFLHQ?PQSEXQLFYNDVAMQLYNNDF	131-160
P8ELQN: ELENMATXIASVRQQLLHK	233-251
SU078: ITAPASELNSILPXVTGIASXNMVSSVNXAV	440-470
SU079: TEAHQQATQVGINSLSINK	421-439
SU080: EFHGYLLEQQK	170-178
SU081: GYMTPGAMEMVSQK	398-411
SU082: SEVRPLMHLLSK	385-397
SU083: SVNSAVTQQSVPTVNLNTQL	465-484
SU106: WXWDTTSYIGN	179-190

Table 2. P_0 and K_r measurements of recombinant Sp(G/C)F1 proteins.

Sp(G/C)F1 protein	P_0 [a]	K_r [b]
55 kD	2.13×10^{-11}	9.68×10^4
50 kD	2.28×10^{-11}	8.56×10^4
43/40 kD [c]	1.45×10^{-11}	1.11×10^5
37 kD	1.48×10^{-11}	2.14×10^5
Total	7.34×10^{-11}	

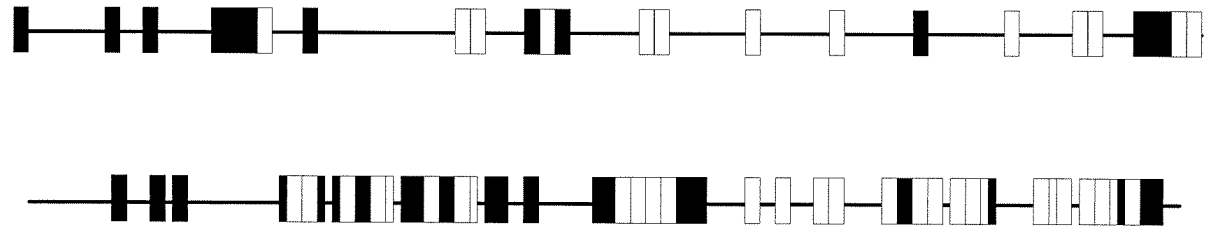
[a] P_0 measurements report the concentrations (M l^{-1})of recombinant protein molecules which are able to bind DNA. This represents about 0.1% of the total amount of Sp(G/C)F1 polypeptides present in the recombinant protein preparation.

[b] K_r measurements report the ratio of the equilibrium constants for specific:non-specific DNA binding.

[c] The complexes containing the 40 and 43 kD polypepetides are not easily resolved. The amounts of both complexes were combined for the P_0 and K_r measurements.

Fig.1. Sp(G/C)F1 binding sites in *cis*-regulatory regions. (A) Arrangement of Sp(G/C)F1 DNA target sites in *CyIIIa* and *Endo16* promoters. Sp(G/C)F1 target sites are displayed as filled boxes, DNA target sites for other factors are shown as empty boxes. The sites were initially characterized by gel shift analysis or sequence homology (Calzone *et al.*, 1988; Thézé *et al.*, 1990; Yuh *et al.*, 1994). Sp(G/C)F1 sites in both promoters are often found closely associated with other DNA binding factor target sites. Not including Sp(G/C)F1, there are eight or nine distinct transcription factors binding to the CyIIIa regulatory domain, some of which have mutliple target sites (Thézé *et al.*, 1990, unpublished data). For Endo 16, there are 12 distinct factors, in addition to Sp(G/C)F1 (Yuh *et al.*, 1994). (B) Consensus Sp(G/C)F1 target site. The consensus sequence for Sp(G/C)F1 target sites was compiled from the 38 *CyIIIa* and *Endo16* target sites depicted in (a). Nucleotides labeled "H" indicate positions in which A, C or T frequently occur. There is a strong bias against G at these positions. The numbers below this sequence refer to the number of Sp(G/C)F1 target sites, out of 38, which match the consensus sequence at the indicated nucleotide positions.

A



B **Sp(G/C)F1 target site consensus:**

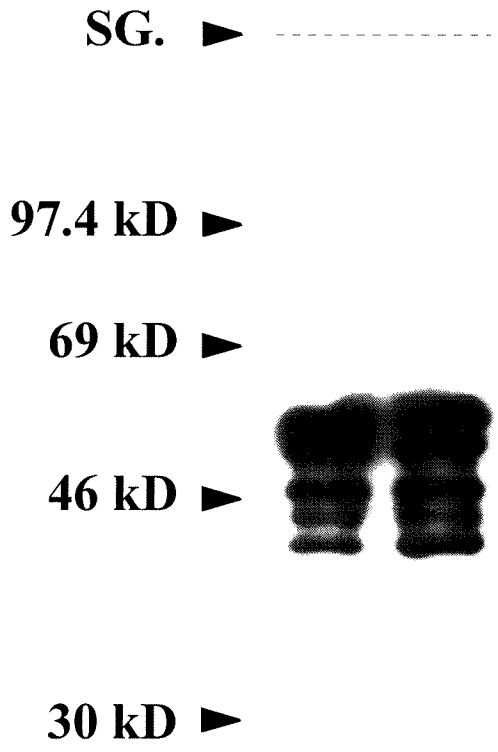
A/T	C/T	C/T	H	N	H	A/T	C	C	C	C	T	H	H	N	H	N	A/T	A/C	N	A/T	T	C/T
31	29	28	34		33	28	38	38	38	38	15	36	35		33		29	27		25	15	26

Fig. 2. cDNA sequence of Sp(G/C)F1 and predicted amino acid sequence of the protein. The 3285 nt sequence excludes the two terminal *Eco*RI sites used in the cDNA library construction. The longest open reading frame encodes a 486 amino acid protein with a calculated molecular weight of 55 kD. This is the same mass measured for the native sea urchin Sp(G/C)F1 protein (see text). Four stop codons upstream and in frame of the predicted open reading frame are shown in bold. A pentapeptide repeat of the form (N/S)VSMP occurs seven times and is underlined. Proline residues in the putative N-terminal transactivation domain are indicated with asterisks (*). Methionines which are believed to serve as alternative initiation codons are located at positions 34, 102, 125/128 and 153 and are marked with daggers (†). A domain of basic residues extends from residues 315-332 and is double underlined. This region is encompassed by the minimal DNA binding domain (residues 223-353) as described in the text. Three regions of the protein contain heptad repeats of hydrophobic residues (L, I, V, F, M, Y or W). These are located from amino acids 13-34, 234-255 and 256-319. Three potential polyadenylation signals of the form AATAAA are double underlined and occur in the 3' untranslated region.

```
GGATTTTGGGGGCATAATTTTGCTATTGATCAAGGATAGCGGGCCCGAATTTACTCATTTTTTAGTGACTTGACGAGGATCCAACAGAGGTGAGTGAGGA  100

GTGAGGTATGTCCACTCTGCCCCAGCCCCTTTCCCACTGCCTGCTGAACCAGGTACACCCCGCTCTCAACCTGCCCCAGACAGGGGTCATCACAGACATC  200
        M† S  T  L  P* Q  P* L  S  H  C  L  L  N  Q  V  H  P* A  L  N  L  P* Q  T  G  V  I  T  D  I    31
                                                |_______heptad_______|_______heptad_______|_______heptad
AAGCCCATGATCAGTAATAAACCTCCTACACAGGAGGTCAAACCAAACATCCTAGCAACTGGCTTGCCCTATCCTCCACTCAACGTGCCTAGGCTACCCG  300
 K  P* M† I  S  N  K  P* P* T  Q  E  V  K  P* N  I  L  A  T  G  L  P* Y  P* P* L  N  V  P* R  L  P*    64
_______|
TCATGCCCAATGTGTCTCTGCCTAGTGTCTCTATGCCGAGTGTGTCTATGCCCAATGTCTCCATGCCCAACGCATCCATGCCCAGTGTTTCGATGCCCAA  400
V  M  P* N  V  S  L  P* S  V  S  M  P* S  V  S  M  P* N  V  S  M  P* N  A  S  M  P* S  V  S  M  P* N    98

TGTGTCCATGCCAAGTATTCCTCATCACAACTTACAGGGTAACTTAGGCCAATTACTCAACAACAGTAATTCCCAAAAAATGTCCCAAATGAAAAAGTGC  500
__V  S  M†_P* S  I  P* H  H  N  L  Q  G  N  L  G  Q  L  L  N  N  S  N  S  Q  K  M†_S  Q  M†_K  K  C   131

CCCAACGAGTTTTTACATCAGAATCCACAAAGTGAGCGACAGCTTTTCTACAACGACGTAGCCATGCAACTGTATAACAGTGACTTCAACAAGTTTGCTT  600
 P  N  E  F  L  H  Q  N  P  Q  S  E  R  Q  L  F  Y  N  D  V  A  M†_Q  L  Y  N  S  D  F  N  K  F  A   164

CCAAGAAGGGATTTCATGGCTACCTGTTAGAGCAACAGAAGTGGAGGTGGGATACCCACAGCTACATAGGTAACCTGGAGACTAGAGTACATAACTTGCT  700
S  K  K  G  F  H  G  Y  L  L  E  Q  Q  K  W  R  W  D  T  H  S  Y  I  G  N  L  E  T  R  V  H  N  L  L  198

CATTAATCCAAACAGTGGGGTTGCACAGAATGTTGCTCGCTACCGCAGTGTCCCCATCAAATGTAAAAGTGAGGATGTGAAGCGATGTAAAGCCACGTCC  800
  I  N  P  N  S  G  V  A  Q  N  V  A  R  Y  R  S  V  P  I  K  C  K  S  E  D  V  K  R  C  K  A  T  S   231
                                                                        |<-DNA BINDING DOMAIN <<<-
AAAGAGCTTGAGAACATGGCAACCCGTATTGCCAGTGTACGGCAGCAGCTGCTACACAAAAAGGGCACCTTGCTGACATCCAGCGATAACAGCGTTATAG  900
 K  E  L  E  N  M  A  T  R  I  A  S  V  R  Q  Q  L  L  H  K  K  G  T  L  L  T  S  S  D  N  S  V  I   264
       |_______heptad_______|_______heptad_______|_______heptad_______|_______heptad_______|  |_______heptad_______|_____
TGTGGCAGAATGAGCTAGCCTACATAGAACAGCTGTTTGACAGGACTGATCAGATGTACAATGAGGTGTTATCTACCCTGGCAAGTGTCAACCAGACCTT 1000
V  W  Q  N  E  L  A  Y  I  E  Q  L  F  D  R  T  D  Q  M  Y  N  E  V  L  S  T  L  A  S  V  N  Q  T  F  298
__heptad_______|_______heptad_______|_______heptad_______|_______heptad_______|_______heptad_______|_______heptad_______|
CTCCCACCTTCAGACAAGCTTCACAGCAGAAGCTGCAGAGTTGGCAGATCGTAGGCGCTTGTGGAGGAGGAGAAAGGAGAACAACCGCAAGAGACGCAAG 1100
  S  H  L  Q  T  S  F  T  A  E  A  A  E  L  A  D  R  R  R  L  W  R  R  R  K  E  N  N  R  K  R  R  K   331
_______heptad_______|_______heptad_______|_______heptad_______|     Basic Domain
CGCATGGAGAAACAACTTGAAAAGATTGAGCAGCGATCTTGTGAGCTTCTCTTTCATATCACATCCCGGGGAGCATATGACCGGGTGCGTTCCCACCCAG 1200
_R  M  E  K  Q  L  E  K  I  E  Q  R  S  C  E  L  L  F  H  I  T  S  R  G  A  Y  D  R  V  R  S  H  P   364
                                    ->>> DNA BINDING DOMAIN->|
AGATGCCTCGTATTGGACCCAGCGAGGTGAACACAGACATGTTAAATGGGATTAAATCTAAATCCGAAGTGAGGCCTCTTATGCACCTACTCAGTAAGGG 1300
E  M  P  R  I  G  P  S  E  V  N  T  D  M  L  N  G  I  K  S  K  S  E  V  R  P  L  M  H  L  L  S  K  G  398

TTACATGACCCCTGGTGCAATGGAGATGGTCTCTCAAAAGATCCAAAAACTAGAGTGTGGTATTAAGACTGAAGCGCACCAACAGGCAACCCAGGTTGGT 1400
  Y  M  T  P  G  A  M  E  M  V  S  Q  K  I  Q  K  L  E  C  G  I  K  T  E  A  H  Q  Q  A  T  Q  V  G   431

ATCAACTCCCTGTCGATCAACAAAATTACAGCACCTGCTTCAGAGCTAAACTCCATACTGCCTCCTGTCACTGGAATTGCCTCATCAAATATGGTGTCAT 1500
 I  N  S  L  S  I  N  K  I  T  A  P  A  S  E  L  N  S  I  L  P  P  V  T  G  I  A  S  S  N  M  V  S   464

CTGTAAACTCAGCTGTGACACAACAATCAGTGCCCACAGTAAATCTTAACACTCAATTAGCGAAGTAAAGACATTTTAACCAAGTCACAGCGACTTTGCC 1600
S  V  N  S  A  V  T  Q  Q  S  V  P  T  V  N  L  N  T  Q  L  A  K  *                                  486

ACATTGCCGAGTGTTTGACATTGAGTAGGCTGTACTCTACTCCACACTGTTTTAACCAACATTGTATTATGTATGAGCATACTCTTACATGGCAAAATGT 1700

ACATTTATTATATTATGCAGGAATGTTGCATCAGTTTAGATCTAACCAAAACAGTTTTAGATTACTTGTTCTTTTTTTTACCAGGTGTACATGTAAATTT 1800

TAACTGGTGAAATTTCACTTCGCTTGAACTCTTTAAAAAATCACAATCTTCTTCATCCTACTAGTAATATACATGTAGGATGGCCCACACTTGGGAAAGA 1900

CATCATTTACCAAATTCCCTAAAGCTGTAATCAAACTCATTGCGGTGGAATAAAAATTGATTACTTGCCTTGCCACATATATTACATGAATCGAATTGTA 2000

AGAAATTGTCCTGTTACTTTGTTCTTTATTGATCATGTCTAGGACTGTCCAAATTCAATGGCTTAGTTTATCCTGGATCGGTGCTTAATCTTAGACTTTC 2100

CTCTTGGCTGATTTGTCACTGATTCATTGTGTTTGTAAGTTGTAATAGAGCTGTAGGGTAAATCTCTTCTCTTTTTTTTTGGGGGGGGTGGAGGGGAATG 2200

GGTTGTTTGGGGGGTGTTTATGTTGACTTTTCCCCCCATAATATTCCTTCATGGACCAAATGTATAGAACAATTACTACCCGTTATTACATGTACATTTG 2300

TCAGTCTCTTAATGTTATTTTATAATGCTTTATTTTTAGGCAGCTAGAGTAGAATAGTTTCTGCATTCTGGACATTTGTTTTACTTTGAAATATATTATG 2400

TTCATTCTTTTTATCATAGCCACTTACCCGCTACCTCGGAAGTAGCCAGTCTGAGACCTGGATTTTTAAACATGGTTTAAAGAAATGGGTGAGATATTTT 2500

AAGTGTACCAATCATCCGAAAAAAAAATGATCTGTGTAGTGTAAAGCTAGAATGTATACCGGTAAATAAACTCAACCAAAAATATTTACCAGGCATTATAT 2600

AGGTAAACCGATTAAGGAGTATTAAAAAAAAATGAAGGGTAATATTAAATCCCTCGTGCAGGGCTTGCACAGAACCTAAATCCGACTTAGTTCCGTGCGAG 2700

CCCTATTCATGCAAGTACGTACGTGTGCGTTTACTACTGGTAGTCTACTCAACAAGACGATATACTGTTGTATTTAACTTGCAATATCTCTTTAATAATA 2800

GAAGTAATATTTGAAGCTCATATTTTTTTAATACATTATAGCCACACATTGAACCCCTTACATCAAGAGCAAAATTTGTTTTAGTGATTGGTACACTGTA 2900

AACATTTTTGTTGTTGCTGAGTTCTTAATAAGTGCTACTTGATCTTTGCCATTTCCTTTATGGTGAAATCATTCAAACTATGTGATTACCAGTAACCTTC 3000

CAATCAAGTTGAACCATGAAATATGGTTTGTATTGCTCGTCTATCATTCAAATTATTCAGAATGTGATGGTCATATTTCAGCTCCAAAACGAAATAAATC 3100

GGGTCACTCAAGCATTGAAGTTGAGCCATTGCTTTGGCTTTAGGGTGTAAATGATTGCTTAATTTTTTACTAATGATTCAATGTTTTTGACAAAGCATTA 3200

AAATCAACAGGTTCAAATTTGCTTCTGGTGTGTAAATCTAATTATTTATGACAACACTAAATTTTGTACAAGTCTGAATCAAGCC                  3285
```

Fig. 3. Evidence that Sp(G/C)F1 is the protein present in embryo nuclear extract that binds P8 DNA target sites. (A) Antibody immunoblot of embryo nuclear extract. A fragment of the Sp(G/C)F1 cDNA encoding the C-terminal 233 amino acids (residues 253-486) was used as an antigen in the production of rabbit polyclonal antibodies. A sample of affinity purified Sp(G/C)F1 from sea urchin embryo nuclear extract was separated in a standard SDS protein gel, blotted to nitrocellulose and incubated with the specific C-terminal antibody. The antibody binding was detected by a chemiluminescent procedure, as detailed in Materials and Methods. Five bands are recognized by the antibody, which have masses of 55 kD, 51 kD, 45 kD, 42 kD and 40 kD, respectively. The boundary between the stacking gel and the separating gel is marked "SG." (B) Two-dimensional gel mobility shift assay of Sp(G/C)F1. An mRNA was produced directly from the Sp(G/C)F1 cDNA clone and used as a template to translate Sp(G/C)F1 protein *in vitro*. This protein was used in a two-dimensional gel mobility shift assay to determine the stoichiometry of protein binding on a single-site oligonucleotide probe (oligonucleotide pair 33/34 from Thézé *et al.*, 1990). The horizontal axis shows the direction of migration of the complexes which were separated by tube gel electrophoresis. A diagram of the tube gel showing the five Sp(G/C)F1-DNA complexes is shown. The top of the tube gel is oriented towards the left. The vertical dimension shows the direction of a standard SDS-protein gel separation. Several shifted bands are evident in a diagonal pattern. Each individual complex is due to a monomeric or homomultimeric binding of a single sized polypeptide as detailed in the text. The estimated masses of these polypeptides are 55 kD, 50 kD, 43 kD, 40 kD and 37 kD, respectively.

A **Immunoblot of Affinity Purified Sp(G/C)F1 Protein**



B **Two-Dimensional Gel Shift**

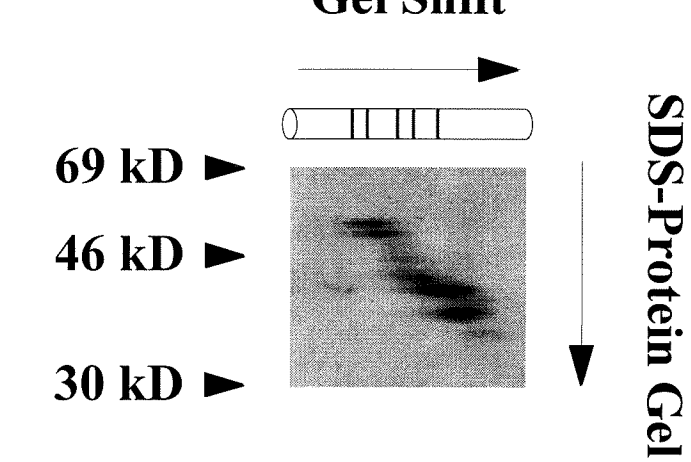


Fig. 4. Supershift gel mobility assay analysis of Sp(G/C)F1 polypeptides. Bacterially expressed Sp(G/C)F1 protein was used in a series of supershift gel shift assays to examine the polypeptide structure of the various Sp(G/C)F1 protein forms. A specific monoclonal antibody recognizing a vector-introduced tag (the T7 tag) was used to identify the N-terminal of the protein while a polyclonal Sp(G/C)F1 antibody was used to identify the C-terminal of the protein. Increasing amounts of either antibody were added to Sp(G/C)F1 gel shift reactions before loading on a polyacrylamide gel. Both antibodies have a "supershifting" effect of increased complex retardation. Lane 1 shows the 33/34 probe alone. Lane 2 shows a control Sp(G/C)F1 gel shift. Lanes 3-6 show the effects of adding increasing amounts of the T7 TAG monoclonal antibody. As the amount of antibody is increased, only the slowest migrating band is retarded indicating that this protein is only N-terminal T7 tag containing polypeptide. Lanes 7-10 show the effects of adding increasing amounts of the C-terminal specific antibody. All complexes are retarded indicating that each polypeptide has the C-terminal antigen. If the Sp(G/C)F1 antibody is added in increasing amounts to affinity purified Sp(G/C)F1 protein the larger complexes are supershifted but not all of the smaller ones are (lanes 12-15). The presence of other Sp(G/C)F1 target site binding proteins thus can not be ruled out, as discussed in the text.

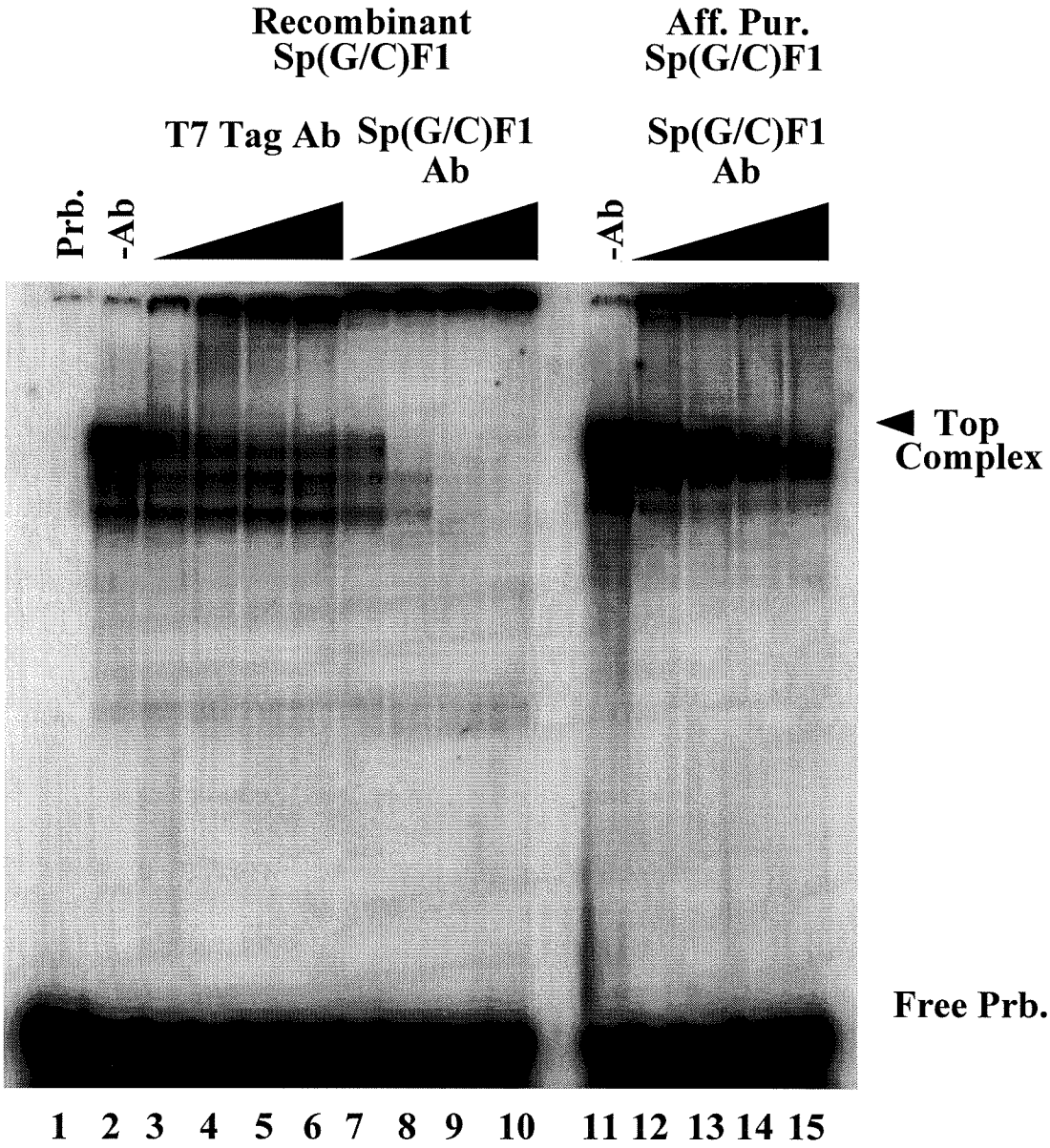

Recombinant
Sp(G/C)F1
Aff. Pur.
Sp(G/C)F1
T7 Tag Ab
Sp(G/C)F1
Ab
Sp(G/C)F1
Ab
Prb.
-Ab
-Ab
Top
Complex
Free Prb.
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15

Fig. 5. Determination of the Sp(G/C)F1 DNA binding domain. (A) Maps of protein expression constructs. All constructs were derived from the full-length Sp(G/C)F1 expressing clone as described in Materials and Methods. The numbering of the amino acid positions begins with the Sp(G/C)F1 initiation codon. The N-terminal histidine and T7 tags add 33 amino acids to the overall protein size. DNA binding activity is indicated for each construct in the right column.

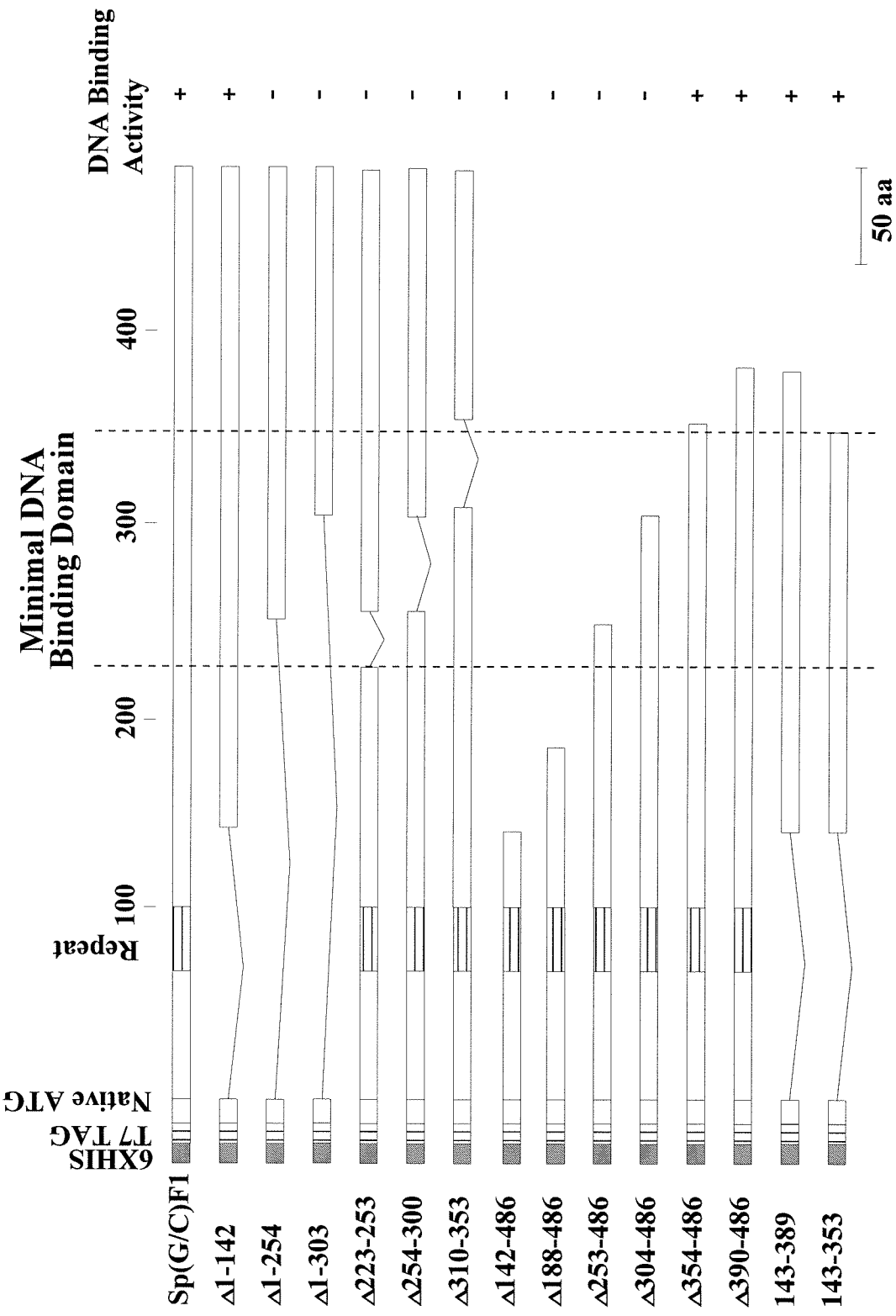

6XHIS
T7 TAG
Native ATG
Repeat
100
200
Minimal DNA
Binding Domain
300
400
DNA Binding
Activity
50 aa
Sp(G/C)F1
Δ1-142
Δ1-254
Δ1-303
Δ223-253
Δ254-300
Δ310-353
Δ142-486
Δ188-486
Δ253-486
Δ304-486
Δ354-486
Δ390-486
143-389
143-353
+
+
-
-
-
-
-
-
-
-
+
+
+
+

Fig. 5. Determination of the Sp(G/C)F1 DNA binding domain. (B) Gel mobility shift assay of Sp(G/C)F1 protein constructs. Protein translated *in vitro* was incubated with Sp(G/C)F1 target site (oligonucleotide pair 33/34, see Materials and Methods) and analyzed for DNA binding activity. Specific complex formation was tested by the addition of a 50-fold excess of unlabeled probe. Control reactions using *in vitro* translation reactions lacking an RNA template are shown in lanes 15-17 and 32-33 with (+) or without (-) specific competitor. Control reactions with full-length Sp(G/C)F1 protein with (+) or without (-) specific competitor are shown in lanes 1-2, 18-19 and 34-35. Only the first 142 N-terminal amino acids are dispensable for DNA binding (lanes 3-4). C-terminal deletions up to the last 132 amino acids are dispensable for DNA binding (lanes 20-23). Further deletions eliminate DNA binding activity (lanes 24-31). Minimal polypeptides capable of binding DNA are shown in lanes 36-39. The smallest polypeptide is 210 amino acids in length (lanes 36-37) and corresponds to positions 143-353 in the full Sp(G/C)F1 protein sequence displayed in Fig. 1. The minimal region for DNA binding probably extends from amino acids 223-353.

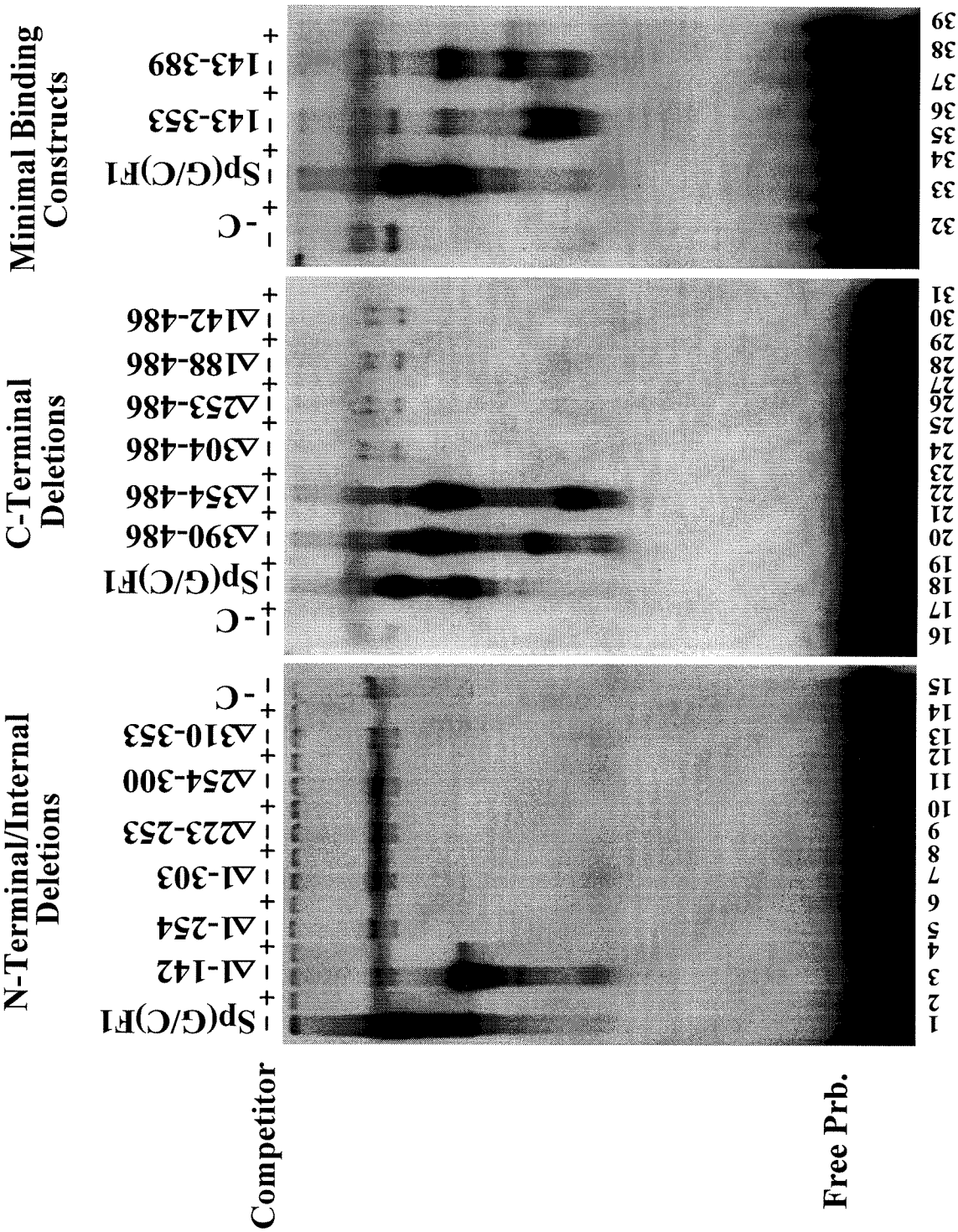

N-Terminal/Internal Deletions
Sp(G/C)F1
Δ1-142
Δ1-254
Δ1-303
Δ223-253
Δ254-300
Δ310-353
-C
C-Terminal Deletions
Sp(G/C)F1
Δ390-486
Δ354-486
Δ304-486
Δ253-486
Δ188-486
Δ142-486
Minimal Binding Constructs
-C
Sp(G/C)F1
143-353
143-389
Competitor
Free Prb.
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15
16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31
32 33 34 35 36 37 38 39

Fig. 6. Analysis of the Sp(G/C)F1 gene in the sea urchin embryo. (A) Genome blot. Three different individual genomic DNAs (numbered 1, 2 or 3; 10 μg each sample) were digested with either *Bam*HI (B), *Hin*dIII (H) or *Eco*RI (E), separated by electrophoresis, blotted and probed with the entire Sp(G/C)F1 coding sequence. The resulting pattern of bands indicates that the Sp(G/C)F1 gene must be present in a single copy in the *S. purpuratus* genome, as described in the text.

Sp(G/C)F1 Genome Blot

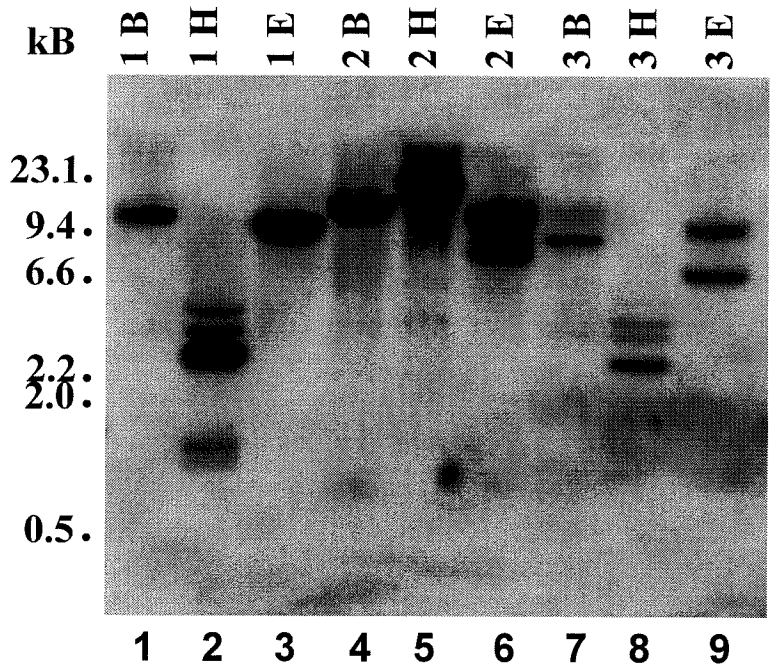
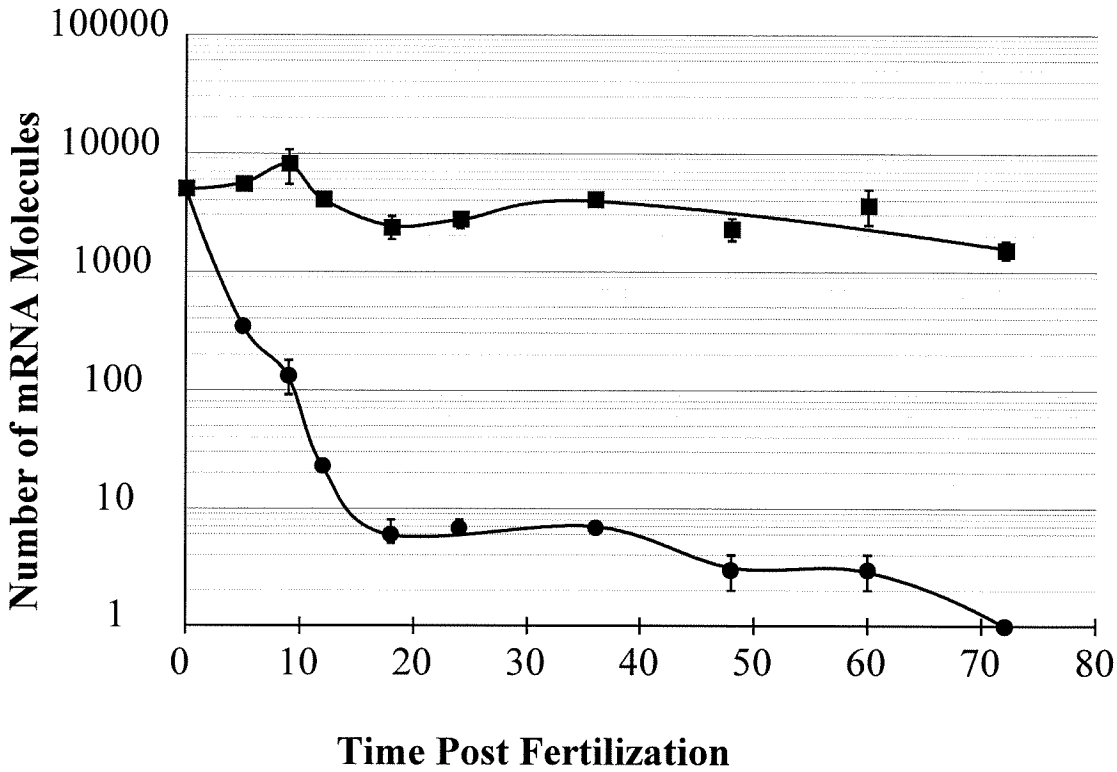


Fig. 7. Sp(G/C)F1 transcript prevalence. Duplicate samples of 50 μg of total sea urchin embryo RNA were used in an RNA protection assay with a 233 nt antisense probe (nt 1510-1743 in the cDNA sequence presented in Fig. 1). The number of molecules of mRNA were calculated according to Lee *et al.*, (1986), assuming a value of 2.8 ng total RNA per embryo (Davidson, 1986). A plot of the number of molecules of Sp(G/C)F1 mRNA per embryo (squares) or per average cell (circles) is shown. Error bars indicate high and low estimates, respectively. The levels of mRNA per embryo peak around 9 hr post-fertilization, and then decrease to around 1500 copies per embryo at 72 hr. On an average per-cell basis, Sp(G/C)F1 mRNA is present at around 5000 copies per egg and, following a sharp decline, is present at about ten molecules per average cell by 24 hr. At 72 hr there is only about one copy of Sp(G/C)F1 mRNA per average cell.

Sp(G/C)F1 mRNA in the Embryo



Chapter 3

A multimerizing transcription factor of sea urchin embryos capable of looping DNA

Robert W. Zeller*, Jack D. Griffith†, James G. Moore*, Carmen V. Kirchhamer*, Roy J. Britten* and Eric H. Davidson*

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

†Lineberger Comprehensive Cancer Center, University of North Carolina, Chapel Hill, NC 27599-7295

ABSTRACT

Sp(G/C)F1 is a recently cloned sea urchin transcription factor that recognizes target sites in several different sea urchin genes. We find that in gel shift experiments this factor is able to multimerize. A quantitative simulation of the gel shift results suggests that Sp(G/C)F1 molecules that are bound to DNA target sites may also bind to one another, thus associating several DNA probe molecules. Sp(G/C)F1 might therefore be able to loop DNA molecules bearing its target sites at distant locations. We demonstrate this prediction by electron microscopy, and using the well-characterized *cis*-regulatory domain of the *CyIIIa* gene, we show that the loop conformations predicted from the known Sp(G/C)F1 target site locations are actually formed *in vitro*. We speculate that the multimerization of this factor *in vivo* may function to bring distant regions of extended regulatory domains into immediate proximity so that they can interact with one another.

INTRODUCTION

Sp(G/C)F1 is a sea urchin embryo transcription factor that binds tightly to certain target site sequences surrounding a core CCCCN sequence, where N is usually T. Most of our knowledge of the function of this protein *in vivo* derives from studies on the *CyIIIa* cytoskeletal actin gene of *Strongylocentrotus purpuratus*, which is expressed exclusively in the aboral ectoderm of the embryo and larva. As shown in Fig. 1, the distal end of this regulatory domain (1-4) includes a cluster of Sp(G/C)F1 sites (cluster A). If these sites are deleted (5) expression of the *CyIIIa·CAT* transgene is sharply reduced. The same result is obtained if the factor is sequestered away from the *CyIIIa·CAT* fusion gene by co-introduction of excess target sites (i.e., *in vivo* competition [6,7]), though this does not affect the spatial domain of the remaining expression. Similar *in vivo* competition results were obtained with the Sp(G/C)F1 sites of region E of Fig. 1 (6,7). In the CyIIIa gene these sites thus mediate a positive transcriptional activating function. A protein binding specifically to these sites, called Sp(G/C)F1, was purified from embryo nuclear extracts, and cloned by Zeller *et al.* (11), though it remains possible that other proteins recognizing these sites could also be present in the nucleus. Though in sequence Sp(G/C)F1 resembles no other known protein, a specific DNA binding domain has been identified experimentally. The N-terminal region of Sp(G/C)F1 contains a high concentration of proline residues, a feature associated with a transcriptional activation function in some other known transcription factors (8-11). A particularly interesting characteristic of Sp(G/C)F1 is that it is present in five different forms of 37, 40, 43, 50 and 55 kD mass. These forms differ in that they include increasing extents of the N-terminal proline-rich domain (11). All five forms are translated from the same mRNA, by alternative utilization of one of five in-frame ATG codons (11).

For some time we have been puzzled by the prevalence and distribution of Sp(G/C)F1 target sites in both the *CyIIIa cis*-regulatory domain and that of another gene

that we have been interested in, the *Endo16* gene. *Endo16* is expressed in the vegetal plate and archenteron of the embryo (12,13). The ~2500 ntp *cis*-regulatory domain of *Endo16* contains >20 distinct sites where Sp(G/C)F1 binds specifically. As in *CyIIIa* (Fig. 1) these sites usually occur in clusters, and often lie close to sites where other transcription factors bind. Sp(G/C)F1 sites also appear in the regulatory domain of *SM30*, a skeletogenic gene (14). Thus this factor appears to service genes expressed at least in three different embryonic territories, *viz* aboral ectoderm, vegetal plate and skeletogenic territories. We measured the relative equilibrium constants for the interactions of Sp(G/C)F1 with three of its *CyIIIa* target sites, and found (11,15) that the preference of this protein for these particular sites (with reference to poly(dA/dT)) ranges from $5x10^4$ to ~10^6. Thus Sp(G/C)F1 is a DNA binding protein which recognizes its target sites with high specificity. In this communication we describe a new functional characteristic of Sp(G/C)F1, *viz* its ability to multimerize with itself. Perhaps this feature explains both the distribution and prevalence of Sp(G/C)F1 sites: interactions between bound Sp(G/C)F1 molecules might have the function of bringing into immediate proximity other transcription factors bound at distant locations within the regulatory domain so that these factors are able to form functional complexes with one another. This would imply that the DNA sequence intervening between interacting Sp(G/C)F1 molecules can be looped out. Though we have no evidence for such conformations *in vivo*, we show in the following that the expected DNA loops indeed form in a purified system in which the only macromolecular components are recombinant Sp(G/C)F1 and the *CyIIIa* regulatory DNA sequence represented in Fig. 1.

MATERIALS AND METHODS

Protein production and gel mobility shift assays

Recombinant Sp(G/C)F1 protein was expressed in *E. coli* (11). Gel mobility shift assays were performed essentially as described (15). The target site probe used in the gel mobility shifts was a 61 bp fragment of the *CyIIIa* regulatory DNA located at -2027 to -1983 in the sequence (16). It was labeled by the end fill reaction (15). The concentration of active Sp(G/C)F1 was determined by probe excess titration (11,15) to be 5.8 n*M* divided among the five mass variants. Aliquots containing various amounts of the protein (4 µl vol) were dispensed into tubes on ice. To each of these tubes was added 16 µl of a premixed solution of the probe (15-50 p*M* final) and 1 X binding buffer (83 m*M* KCl, 5 m*M* $MgCl_2$, 20 µ*M* EDTA, 700 µ*M* DTT, 14 m*M* Hepes-KOH, pH 7.9, 2% glycerol). The complexes were allowed to form on ice for 10-15 min then were loaded directly to an 8% polyacrylamide gel (30:0.8 acrylamide:bis-acrylamide) previously pre-run for an hour. After electrophoresis at 250 V, the gel was dried, exposed to a Kodak imaging screen, and subsequently read in a Molecular Dynamics Phosphorimager. Data were converted to cpm by reference to a known probe sample that was counted in both scintillation counter and in the Phosphorimager.

Electron microscopy of DNA-protein complexes

Samples were prepared for EM by incubating 1 ng of the 2.3 *CyIIIa* regulation region, i.e., positions -2160 to +181 (16) with Sp(G/C)F1 protein for 10 min on ice in a buffer containing 83 m*M* KCl, 5 m*M* $MgCl_2$, 20 µ*M* EDTA, 700 µ*M* DTT, 14 m*M* Hepes-KOH, pH 7.9, 2% glycerol (final volume of 25 µl, final protein concentration 1.75 n*M*). Following incubation, the sample was treated with 0.6% glutaraldehyde for 15 min on

ice, then chromatographed over 2 ml Biogel A5M (Biorad, Inc.) columns equilibrated in 10 m*M* Tris, 0.1 m*M* EDTA (pH 7.5). The peak fraction containing the DNA-protein complexes was treated as described (17). In brief, thin carbon foils supported by 400 mesh copper grids were exposed to a glow discharge, and the complexes were adsorbed for 30 sec in a buffer containing 2 m*M* spermidine and 150 m*M* NaCl. The grids were then rinsed with sequential washes of water and graded ethanol solutions to 100% ethanol. The samples were air-dried and rotary shadowcast with tungsten at 10^{-7} torr.

Images were recorded on sheet or 35 mm film using a Philips CM12 electron microscope. Images for analysis or publication were captured from the film using a Cohu CCD camera and a Data Translation Quick Capture board in a MacIntosh computer. The NIH IMAGE software was used to adjust the contrast in the images for publication. Images were transferred to 35 mm film using a GCC film recorder. Lengths of DNA segments on the micrographs were measured on the photographic prints using a Summagraphics digitizing tablet coupled to a CompuAdd computer programmed with software developed in the laboratory of JDG.

RESULTS

Multimerization of Sp(G/C)F1 on a single-site probe

An example of a gel shift experiment that reveals the multimerization of Sp(G/C)F1 is shown in Fig. 2A. Increasing amounts of renatured recombinant Sp(G/C)F1 were reacted with a 61 nt oligonucleotide probe (see Materials and Methods) that contained a *single target binding site*. The bands formed at the lower protein concentrations (bands 1-5, lanes 1-7) represent complexes formed respectively, by the 37, 40 + 43, 50 and 55 kD variants of the protein. This was shown earlier (11) by the use of 2D gel electrophoresis and immunological identification of the complexes with a

polyclonal antibody that recognizes the C-terminal region of the protein common to all five variants. As protein concentration is increased, more slowly migrating complexes appear. A blurred region of the gel is seen above the 55 kD complex band in lanes 8-11 (region 5-A), representing a series of complexes that are too close to one another in size to be resolved into discrete bands. Above this two more sharply resolved sets of complexes (bands A and B in Fig. 2A) appear at the highest protein concentrations. Since there is only a single target site on the probe, and since no other proteins than the purified recombinant Sp(G/C)F1 are present in the experiment, all the complexes migrating more slowly than that formed by the 55 kD variant must be due to Sp(G/C)F1oSp(G/C)F1 interactions that involve Sp(G/C)F1 molecules that are also bound to the labeled DNA probe.

In order to interpret the results of experiments such as shown in Fig. 2A, we constructed models of the presumed interactions based on simple equilibrium considerations (see Appendix). These models allowed us to test a number of different stoichiometries. The best fit to the data were obtained with what we termed the *dimer model* and are shown in Fig. 2B. The assumptions on which this model is based, and the number of Sp(G/C)F1 molecules and DNA probe molecules assumed in this model for each of the complexes observed, are detailed in the Appendix. Very briefly, the compositions of the individual bands in the dimer model are as follows: *Bands 1-5*, are formed of homodimers of the Sp(G/C)F1 size variants complexed to a single DNA probe molecule, i.e., the 37 kD variant dimerizes only with another 37 kD variant, the 55 kD variant dimerizes only with itself, etc. The steepness with which these complexes accumulate with increasing protein concentration excludes monomeric complexes (cooperativity in DNA binding is not an issue since there is only a single target site per probe). *Band 5-A* in the dimer model consists of unresolved complexes formed between monomers of any of the Sp(G/C)F1 variants and any of the DNAo$(Sp(G/C)F1)_2$ complexes represented in Bands 1-5. Region 5-A complexes result from

Sp(G/C)F1·Sp(G/C)F1 interactions that occur at higher protein concentrations than does formation of the initial complexes with DNA. *Band A* in this model consists of dimers of the (Sp(G/C)F1)$_3$oDNA complexes represented in Region 5-A, i.e., Band A complexes contain six Sp(G/C)F1 molecules and two DNA molecules per complex. *Band B* consists of dimers of the complexes formed in Band A, i.e., Band B complexes consist of twelve Sp(G/C)F1 molecules and four DNA probe molecules per complex. Thus both Band A and Band B consist of complexes in which DNA probe molecules already complexed to Sp(G/C)F1 dimers are brought into a higher order multimeric complex by further Sp(G/C)F1oSp(G/C)F1 interactions.

Sp(G/C)F1 loops DNA that contains multiple target sites

If Sp(G/C)F1 molecules already bound to their DNA target sites are capable of binding specifically to one another, as in the interpretation of Fig. 2A just discussed, then such interactions should occur between Sp(G/C)F1 molecules bound at target sites at distant locations on the same DNA molecule. This would result in the formation of a DNA loop held together by Sp(G/C)F1 complexes that could be visualized by electron microscopy (18-20). We utilized the 2200 ntp *CyIIIa* regulatory domain DNA fragment illustrated in Fig. 1 for a direct test of this prediction, and a 600 nt control fragment that included only the "E" Sp(G/C)F1 cluster of target sites. The latter should display proteinoDNA complexes, but no loops, and this is indeed what was observed (data not shown). However, when reacted with Sp(G/C)F1 protein, the 2200 ntp *CyIIIa* regulatory DNA fragment formed loops at a frequency of over 60% (see Materials and Methods for conditions).

To analyze these loop structures the distances between the ends of the DNA molecules and the bound Sp(G/C)F1 protein at the base of the loop(s) were measured. As Fig. 1 indicates, the Sp(G/C)F1 target sites and site clusters are located in such a fashion

that the loops which could be formed by interactions at any two of these sites will produce ends of different lengths that could be identified unambiguously. Table 1 records the different forms of loop that are thus recognized, categorized according to the site nomenclature of Fig. 1. Table 1 includes data from 59 loop structures that were clearly visualized, and that we examined in detail. Examples of many of these categories of loop structure that we saw are shown in Fig. 3. As indicated in Table 1, most of the possible binary interactions that could occur were in fact included in the sample of 59 structures analyzed. The frequency of given loops as opposed to others reflects the relative magnitude of the equilibrium constants for the target sites at which they form, at least in those cases for which we have data. For example, some sites in region A have an apparent equilibrium constant for Sp(G/C)F1 interaction that is about ten times higher than for the other measured *CyIIIa* sites, probably because of the opportunity for cooperative interactions between adjacent bound Sp(G/C)F1 molecules within this cluster. Correspondingly, a major fraction of all the loops observed involved site cluster A (49 of the 59 structures analyzed; Table 1). Single loops formed by Sp(G/C)F1 interactions at two sites as illustrated in Fig. 3, panels A-E, accounted for 71% of the structures we studied (42 of the 59 structures), and multiple loop structures accounted for the remainder. For example, panels F and G show double loops formed by triple interactions, and panel H shows a five-loop structure formed by interactions at six possible sites.

DISCUSSION

The correlation between the positions of Sp(G/C)F1 target sites in the *CyIIIa* regulatory domain and the positions of loop-forming sites, together with the high frequency of looped structures that we encountered in these experiments, confirm that Sp(G/C)F1 bound to DNA is capable of forming stable multimeric complexes. In the gel

shift experiment of Fig. 2A this property evidently results in the sequestration of several DNA probe molecules per complex (bands A and B), and in the electron microscope study it results in the formation of DNA loops. This conclusion stands independently of the exact stoichiometry of the complexes formed in the gel shift experiment, which will require direct protein cross-linking to establish.

Sp(G/C)F1 evidently behaves very much like the mammalian Sp1 transcription factor. Sp1 is a transcriptional activator that binds GC-rich target sites, which also appear in clusters (21,22). Sp1 also forms higher multimeric complexes when bound to DNA probes, as visualized in gel shift experiments, just as does Sp(G/C)F1 (23). Furthermore, two different groups (19,20) have shown by electron microscopy that *in vitro*, Sp1 loops DNA bearing multiple sites in the same way as does Sp(G/C)F1. However, Sp(G/C)F1 is not a sea urchin version of Sp1, since Sp(G/C)F1 has an entirely dissimilar sequence from Sp1, and in fact it displays no sequence homology with any other known transcription factor (or other protein) (11). Thus two completely unrelated transcription factors share the specific property of looping DNA by formation of stable multimers between factor molecules bound to DNA at distant sites. It follows that this property might reflect an important *in vivo* function.

Following is a speculation on the nature of this function, that is suggested by the distribution of Sp(G/C)F1 sites in the *CyIIIa* regulatory domain. As indicated in Fig. 1, the A-B Sp(G/C)F1 site cluster lies at the distal end of the regulatory domain; the C site cluster occurs toward the middle of the sequence where the Sp(G/C)F1 sites are contiguous to sites for two other transcription factors (labeled 7 and 8 in Fig. 1); the proximally located D-E sites bracket target sites for four additional transcription factors (labeled 1-4 in Fig. 1). Earlier work (2,6,7), and extensive new gene transfer studies to be reported elsewhere, show that these three regions of the *CyIIIa* regulatory domain have distinct functions. The distal region, which consists mainly of the Sp(G/C)F1 target sites of the A-B cluster, controls the level of expression mediated by the proximal and middle

modules of the regulatory domain, and has no spatial control function of its own. The proximal and middle modules each include target sites for both negatively and positively acting transcription factors and they each exercise control over the spatial patterns of *CyIIIa* expression. The proximal module alone (target sites 1-5 plus Sp(G/C)F1 sites D-E in Fig. 1) suffices to set expression in the aboral ectoderm early in development. The middle module (target sites 6-8 plus Sp(G/C)F1 site C in Fig. 1) controls spatial and temporal patterns of expression in postgastrular embryos, but it requires portions of the proximal module for function. All three of these components of the regulatory domain are required for normal embryonic expression of *CyIIIa* transgenes, and they obviously must intercommunicate in some way; early in development distaloproximal intercommunication is required; and later on distalomiddleoproximal intercommunication is required. Formation of stable multimeric complexes between Sp(G/C)F1 molecules bound at the distal, middle, and proximal target sites could provide the specific mechanism for the physical intercommunication of transcription factors bound within individual regulatory modules with one another, and with the basal transcription apparatus that will be located immediately adjacent to the proximal regulatory module. Thus early in development we imagine a distaloproximal DNA loop (*cf.* Fig. 3D) that would bring the activation function of distal Sp(G/C)F1 modules into direct proximity of the basal apparatus and the proximal transcription factors. Later in development multimerization of Sp(G/C)F1 bound at the A-B sites with Sp(G/C)F1 bound at the C and D-E sites would be predicted (*cf.* Figs. 3G and H). Similar arguments could apply to the *Endo16* gene regulatory system mentioned in the Introduction, where Sp(G/C)F1 sites are again associated with *cis*-regulatory modules distributed over more than two kb of DNA. Thus in general terms our speculation is that Sp(G/C)F1 carries out the essential function of a *regulatory module intercommunicator*, as well as serving as a transcriptional activator in its own right. The multimerization and DNA looping behavior of this protein that we observe *in vitro* would reflect this ubiquitously required *in vivo* function, and we

propose that this function explains the widespread prevalence and local distribution of Sp(G/C)F1 sites in the *cis*-regulatory domains of sea urchin genes.

REFERENCES

1. Hough-Evans, B. R., Franks, R. R., Cameron, R. A., Britten, R. J., and Davidson, E. H. (1987). Correct cell type-specific expression of a fusion gene injected into sea urchin eggs. *Dev. Biol.* **121**:576-579.
2. Wang, D. G.-W., Kirchhamer, C. V., Britten, R. J. and Davidson, E. H. (1994). SpZ12-1, a negative regulator required for spatial control of the territory-specific *CyIIIa* gene in the sea urchin embryo. Submitted.
3. Cameron, R. A., Britten, R. J., and Davidson, E. H. (1989). Expression of two actin genes during larval development in the sea urchin *Strongylocentrotus purpuratus*. *Mol. Reprod. Dev.* **1**:149-155.
4. Livant, D., Cutting, A., Britten, R. J., and Davidson, E. H. (1988). An *in vivo* titration of regulatory factors required for expression of a fusion gene in transgenic sea urchin embryos. *Proc. Natl. Acad. Sci. USA.* **85**:7607-7611.
5. Flytzanis, C. N., Britten, R. J., and Davidson, E. H. (1987). Ontogenic activation of a fusion gene introduced into sea urchin eggs. *Proc. Natl. Acad. Sci. USA.* **84**:151-155.
6. Hough-Evans, B. R., Franks, R. R., Zeller, R. W., Britten, R. J., and Davidson, E. H. (1990). Negative spatial regulation of the lineage specific CyIIIa actin gene in the sea urchin embryo. *Development* **110**:41-50.
7. Franks, R. R., Anderson, R., Moore, J. G., Hough-Evans, B. R., Britten, R. J., and Davidson, E. H. (1990). Competitive titration in living sea urchin embryos of regulatory factors required for expression of the CyIIIa actin gene. *Development* **110**:31-40.
8. Pei, D. and Shih, C. (1991). An "attenuator domain" is sandwiched by two distinct transactivator domains in the transcription factor C/EBP. *Mol. Cell. Biol* **11**:1480-1487.

9. Mermod, N., O'Neill, E. A., Kelly, T. J. and Tjian, R. (1989). The proline-rich transcription activator of CTF/NF-I is distinct from the replication and DNA binding domain. *Cell* **58**:741-753.

10. Friedman, A. D. and McKnight, S. L. (1990). Identification of two polypeptide segments of CCAAT/enhancer-binding protein required for transcriptional activation of the serum albumin gene. *Genes Dev.* **4**:1416-1426.

11. Zeller, R.W., Coffman, J.A., Harrington, M.G., Britten, R.J. and Davidson, E.H. (1994). Sp(G/C)F1, a sea urchin embryo transcription factor, exists as five nested variants encoded by a single mRNA. Submitted.

12. Nocente-McGrath, C., Brenner, C. A. and Ernst, S. G. (1989). *Endo16*, a lineage-specific protein of the sea urchin embryo, is first expressed just prior to gastrulation. *Dev. Biol.* **136**:264-272.

13. Ransick, A., Ernst, S., Britten, R. J. and Davidson, E. H. (1993). Whole mount in situ hybridization shows *Endo-16* to be a marker for the vegetal plate territory in sea urchin embryos. *Mech. Dev.* **42**:117-124.

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

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

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

17. Griffith, J. and Christiansen, G. (1978). Electron microscopic visualization of chromatin and other DNA-protein complexes. *Ann. Rev. Biophys. Bioeng.* **7**:19-35.

18. Griffith, J., Hochschild, A., and Ptashne, M. (1986). DNA loops induced by cooperative binding of lambda repressor. *Nature* **322**:750-752.

19. Su, W., Jackson, S., Tjian, R. and Echols, H. (1991). DNA looping between sites for transcriptional activation: Self-association of DNA-bound Sp1. *Genes Dev.* **5**:820-826.

20. Mastrangelo, I. A., Courey, A. J., Wall, J. S., Jackson, S. P. and Hough, P. V. C. (1991). DNA looping and Sp1 multimer links: A mechanism for transcriptional synergism and enhancement. *Proc. Natl. Acad. Sci. USA* **88**:5670-5674.

21. Dynan, W. S. and Tjian, R. (1983). The promoter-specific transcription factor Sp1 binds to upstream sequences in the SV40 early promoter. *Cell* **35**:79-87.

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

23. Pascal, E. and Tjian, R. (1991). Different activation domains of Sp1 govern formation of multimers and mediate transcriptional synergism. *Genes Dev.* **5**:1646-1656.

24. Coffman, J. A., Moore, J. G., Calzone, F. J., Hood, L. E., Britten, R. J. and Davidson, E. H. (1992). Automated sequential affinity chromatography of sea urchin embryo DNA binding proteins. *Mol. Marine Biol. Biotech* **1**:136-146.

25. Char, B. R., Bell, J. R., Dovala, J., Coffman, J. A., Harrington, M. G., Becerra, J. C., Davidson, E. H., Calzone, F. J. and Maxson, R. (1993). SpOct, a gene encoding the major octamer binding protein in sea urchin embryos: Expression profile, evolutionary relationships, and DNA binding of expressed protein. *Dev. Biol.* **158**:350-363.

Table 1. Categories of loops in sample of 59 structures.

No. loops per molecule	Interaction	No. occurrences
1 (71.2%)	AA	2
	AB	4
	AC	9
	AD	9
	AE	3
	BC	5
	BD	0
	BE	1
	CD	4
	CE	2
	DE	3
2 (22.0%)	AAC	2
	ABC	1
	ACD	2
	ACE	3
	BCD	3
	BCE	1
	CDE	1
3 (5.1%)	AABD	1
	ABCD	1
	ABCF	1
5 (1.7%)	AACDEE	1
TOTAL		59

Fig. 1. Locations of Sp(G/C)F1 binding sites in the 2.2 kb *CyIIIa* regulatory domain. Binding sites for Sp(G/C)F1 protein (from unpublished data and refs. 15, 16, 24; are shown as filled ovals. Open boxes indicate the positions of other DNA-binding protein target sites: 1 = P1 site, 2 = SpOct1 site (25), 3 = SpP3A site, 4 = SpTEF1 site, 5 = SpCTF1 site, 6 = the pair of SpZ12-1 sites (2), 7 = P7II site, 8 = Sprunt1 site and 9 = newly discovered site designated the P9 site. The arrow indicates the start of transcription. The regulatory region displayed here extends from -2160 to +133. The percentage values indicate the distance from the various Sp(G/C)F1 sites to the ends of the DNA molecule.

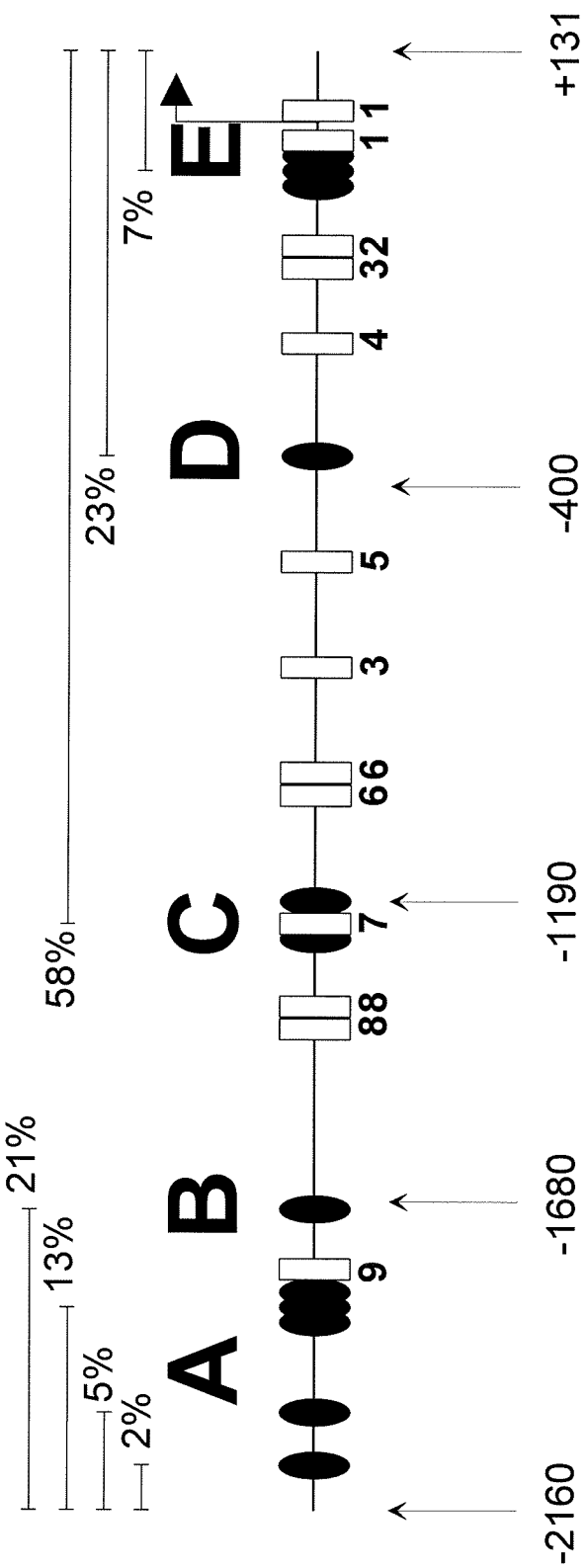

A
B
C
D
E
21%
13%
5%
2%
58%
23%
7%
9
88
7
66
3
5
4
32
11
-2160
-1680
-1190
-400
+131

Fig. 2. Quantitative protein excess gel mobility assay. (A) Phosphorimager report of the labeled complexes formed. A synthetic DNA probe bearing a single Sp(G/C)F1 target site was incubated with increasing amounts of recombinant Sp(G/C)F1 protein. The complexes formed were resolved by electrophoresis through an 8% polyacrylamide gel. The amount of added protein increases from left (lane 1, 1.8×10^{-11} *M*) to right (lane 11, 5.8×10^{-10} *M*). Bands 1-5 indicate the complexes formed by the different Sp(G/C)F1 variants binding to the probe (see text). Region 5-A, Bands A, and B are complexes formed by Sp(G/C)F1·Sp(G/C)F1 interactions, as described in text.

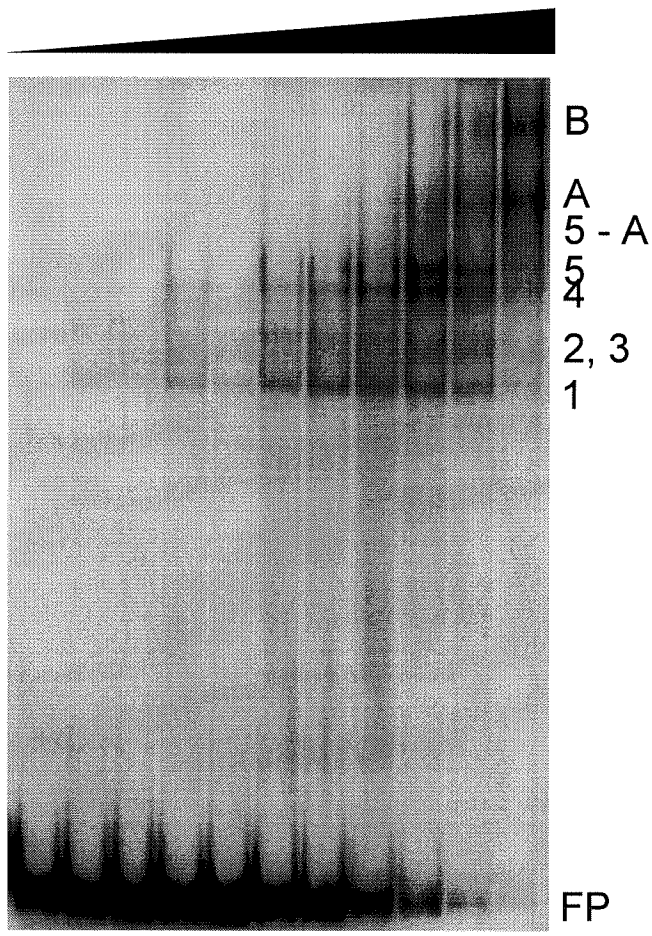

B
A
5 - A
5
4
2, 3
1
FP

Fig. 2. Quantitative protein excess gel mobility assay. (B) Curves generated by the dimer model for Sp(G/C)F1 complex formation discussed in text, and in the Appendix. The solid lines show the functions generated using the least squares solution for the parameters of the model for these data. Each panel represents one of the bands figured in (A), as indicated. The data from the phosphorimager analysis are indicated as open symbols, after normalization, so that the probe present in each complex is expressed as a fraction of total probe in the experiment. D represents free probe.

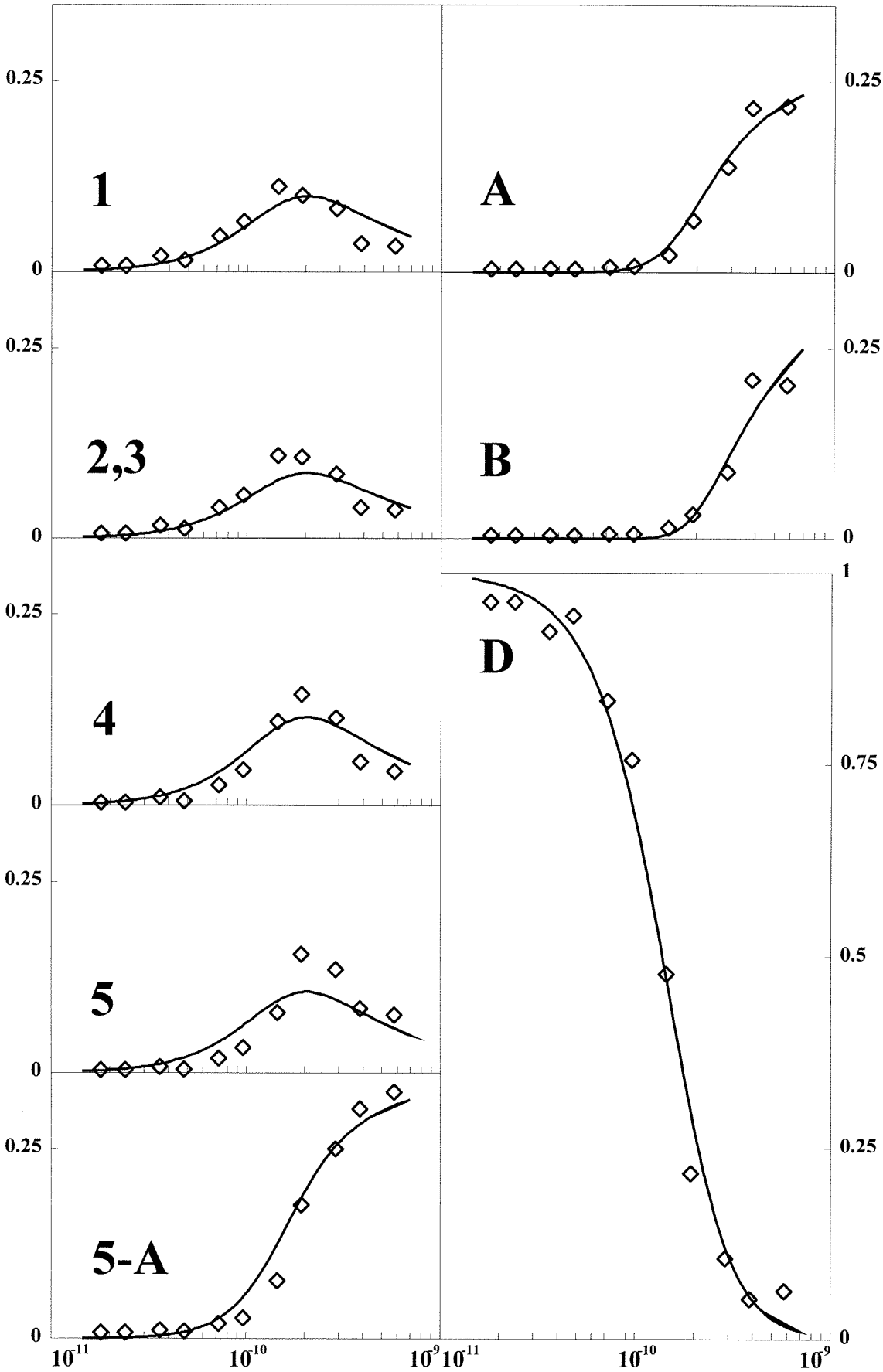
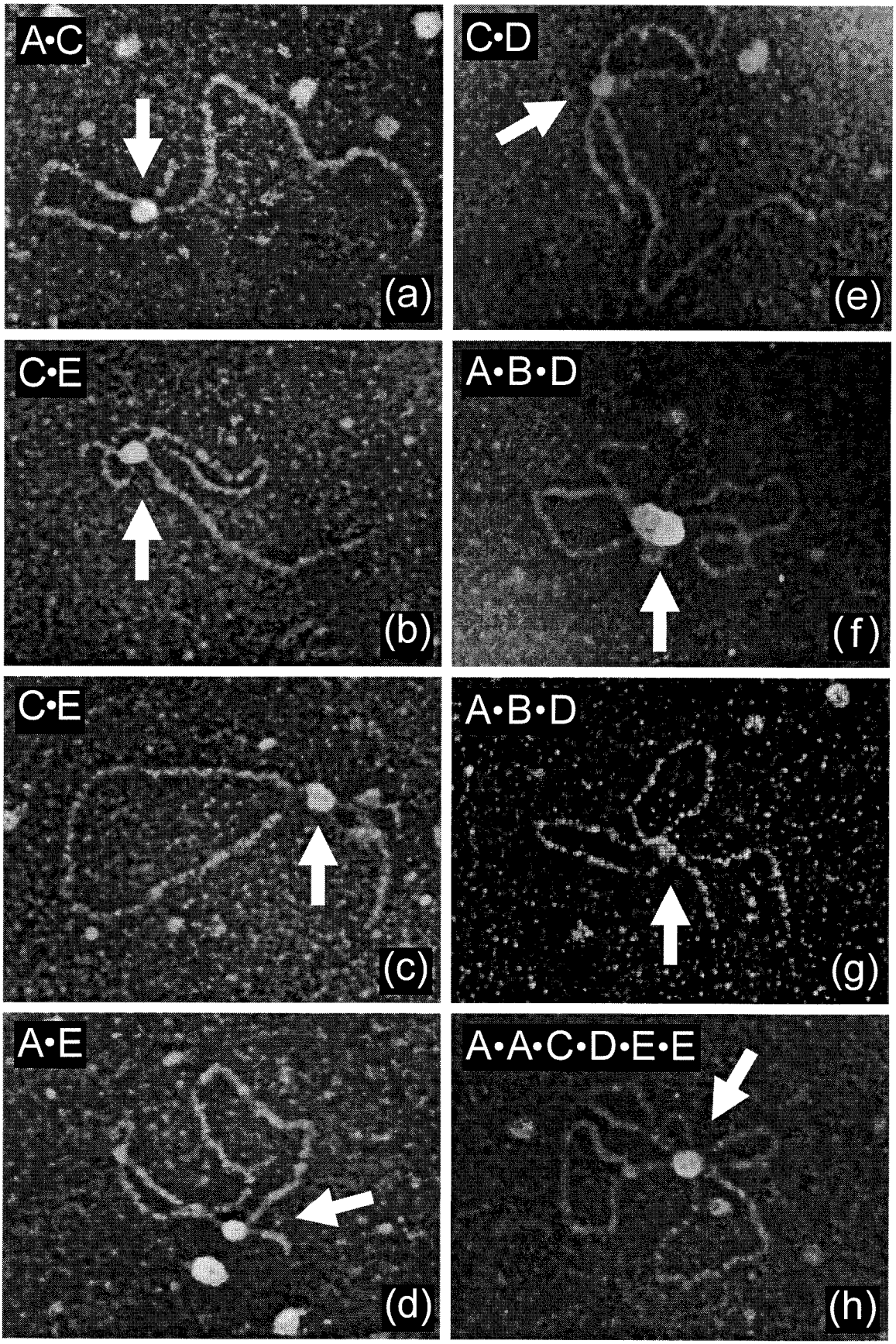

1
A
2,3
B
4
D
5
5-A
0.25
0
1
0.75
0.5
10^{-11}
10^{-10}
10^{-9}

Fig. 3. Loops formed on *CyIIIa* DNA by interactions between bound Sp(G/C)F1 molecules. Locations of Sp(G/C)F1 target sites (A-E) are shown in Fig. 1. The sites at which are bound the Sp(G/C)F1 molecules whose interaction forms the loop are indicated in each panel.


A•C
(a)
C•D
(e)
C•E
(b)
A•B•D
(f)
C•E
(c)
A•B•D
(g)
A•E
(d)
A•A•C•D•E•E
(h)

APPENDIX

The accumulation of probe in the specific complexes shown in Fig. 2A, as a function of protein concentration, was tested against several models that might explain the formation and composition of these complexes. We compared the functions generated from these models to the data obtained from scanning the gels by Phosphorimager, by means of a non-linear least squares procedure. The concentrations of recombinant Sp(G/C)F1 protein and of oligonucleotide probe were known *a priori*, as were the fractions of total protein (ϕ) for each of the Sp(G/C)F1 molecular weight variants: where *i* or *j* denote the 37, 40, 43, 50, and 55 kD variants, and *P* is protein, $P_i = \phi_i P$; $\Sigma\phi_i = 1$. We also knew (24, 11) that bands 1-5 contain respectively *only* the 37, 40, 43, 50 and 55 kD variants; i.e., if these complexes contain multimers of Sp(G/C)F1 protein they are homomultimers and not heteromultimers. Three models, and the degree to which they conformed to the data, expressed as χ^2 values (σ = 10%, n = 13 X 8; 11 X 8 for the two data sets), are shown in the following Table. Here, where *n* is the number of Sp(G/C)F1 (*P*) molecules, and *m* is the number of DNA (D) probe molecules per complex, we have used the convention *n.m* to describe the composition of the complexes:

Model	Band	n.m; Composition	χ^2 (two data sets)
	1-5	1.1 ; P_iD	
"monomer"	Region 5-A	2.1 ; $P_j \cdot P_iD$	18.1, 15.0
	A	4.2 ; $(P_j \cdot P_iD)_2$	
	B	8.4 ; $((P_j \cdot P_iD)2)_2$	

	1-5	2.1 ; $(P_i)_2D$	
"dimer"	Region 5-A	3.1 ; $P_j{\cdot}(P_i)_2D$	5.3, 3.8
	A	6.2 ; $(P_j{\cdot}(Pi)_2D)_2$	
	B	12.4 ; $((P_j{\cdot}(P_i)_2D)_2)_2$	
	1-5	4.1 ; $(P_i)_4D$	
"tetramer"	Region 5-A	5.1 ; $P_j{\cdot}(P_i)_4D$	9.6, 10.2
	A	10.2 ; $(P_j{\cdot}(P_i)_4D)_2$	
	B	12.4 ; $((P_j{\cdot}(P_i)_4D)_2)_2$	

In constructing these models we assumed that all the mass variants of Sp(G/C)F1 may not interact with DNA with exactly the same equilibrium constant (k_N); $k_{N_i} = \rho_i k_N$ (ρ_i values were known from earlier experiments (11)). The equilibrium constant for Sp(G/C)F1·Sp(G/C)F1 interaction is k_E; for the interaction $[P_j{\cdot}(Pi)2D]{\cdot}[P_j(Pi)2D]$ it is kD; for the interaction $(P_j{\cdot}(P_i)_2D)_2{\cdot}(P_j{\cdot}(P_i)_2D)_2$ it is k_Q. Where $S_i = \rho_i \phi_i^2$, and $S = \sum_i S_i$, the equations used to describe the complexes formed in the dimer model are: (Bands 1-5), $PD_{2.1} = S_i k_N{\cdot}P_2{\cdot}D$; (Region 5-A), $PD_{3.1} = k_E{\cdot}S{\cdot}k_N P^3{\cdot}D$; (Band A), $PD_{6.2} = k_D{\cdot}k_E^2{\cdot}S^2{\cdot}k_N^2{\cdot}P^6{\cdot}D^2$; (Band B), $PD_{12.4} = k_Q{\cdot}k_D^2{\cdot}k_E^4{\cdot}S^4{\cdot}k_N^4{\cdot}P^{12}{\cdot}D^4$. It is worth noting also that in this model Bands 1-5 approximately display an expected log-linear relation between mass and mobility in the gel, but this is no longer so for the higher order complexes.

Chapter 4

Developmental utilization of P3A1 and P3A2: Two proteins which recognize the same DNA target site in several sea urchin gene regulatory regions

Robert W. Zeller, Roy J. Britten and Eric H. Davidson

Division of Biology 156-29

California Institute of Technology

Pasadena, CA 91125

ABSTRACT

P3A1 and P3A2 are DNA-binding proteins that interact specifically with the same target sites in the regulatory domains of the *Strongylocentrotus purpuratus CyIIIa* gene and also of several other known genes. In this work we used antibodies raised against recombinant P3A1 and P3A2 to quantitate these transcription factors in eggs and in the nuclear compartments of embryos of various stages. Both proteins are present in unfertilized eggs, and both enter the embryonic nuclei early in development, but only P3A2 remains present in nuclei at functional concentration beyond the early gastrula stage. Combined with earlier measurements of P3A site binding at cleavage stages, P3A1 would be replaced by P3A2 at target sites in genes regulated by these factors.

INTRODUCTION

Interactions at two specific target sites that include the sequence T/CNT/CGCGCA/T ("P3A" sites) are essential to prevent ectopic expression of *S. purpuratus CyIIIa* fusion genes. The *CyIIIa* gene, which encodes a cytoskeletal actin, is normally expressed only in the aboral ectoderm of the embryo and larva (Cox *et al.*, 1986; Cameron *et al.*, 1989). If interactions at the P3A sites are prevented, either by *in vivo* site competition (Franks *et al.*, 1990; Hough-Evans *et al.*, 1990), or by site deletion (C. Kirchhamer and E, Davidson, in preparation), expression of injected *CyIIIa·CAT* fusion genes is dramatically extended to the oral ectoderm. P3A sites are also found in the regulatory domain of the *SM50* spicule matrix protein (Calzone *et al.*, 1991) and in another gene expressed in aboral ectoderm, *Spec1a* (Thiebaud *et al.*, 1990), but the regulatory roles of the P3A sites in respect to expression of these genes has not yet been defined. In previous work we identified and cloned two different embryo nuclear proteins that bind tightly and specifically to the P3A sites of *CyIIIa* and *SM50*. P3A1 is a two Zn-finger protein, which was originally found by screening an expression cDNA library with an oligomerized P3A target site (Höög *et al.*, 1991). Though it displays no sequence similarity to any other known protein outside of the Zn finger binding region, its Zn fingers are most similar to those of the *Drosophila hunchback* transcription factor (Höög *et al.*, 1991). P3A2 is a protein of entirely different and novel structure, which has no sequence resemblance to P3A1, except for a stretch of a few identical residues within the DNA binding domains of the two proteins (Höög *et al.*, 1991). P3A2 was recovered by affinity chromatography from embryo nuclear extract, and partial amino acid sequences obtained from purified protein were used to clone the cDNA (Calzone *et al.*, 1991). Two gene regulatory proteins that display significant homology with P3A2 have subsequently appeared in the data bases. These are the nuclear proteins encoded by the *Drosophila erect wing* gene (DeSimone and White,

1993), and a human transcription factor called NRF-1, which regulates nuclear genes encoding certain mitochondrial proteins and some other widely expressed genes (Virbasius *et al.*, 1993). These three proteins constitute a new transcription factor family.

Here we report measurements of the quantities of P3A1 and P3A2 proteins in eggs, and in nuclei of embryos of various stages. Our measurements were obtained by quantitation of reactions with antibodies that had been raised against recombinant P3A1 and P3A2 encoded by full-length sea urchin cDNA clones (Höög *et al.*, 1991; Calzone *et al.*, 1991). Some relevant information that we use in the following was available at the outset: (*i*) Only P3A2 can be purified from 24 hr nuclear extract on affinity columns presenting oligomerized P3A sites, as established by immunological identification. Though P3A1 protein is detectable at low levels in this extract, it is not concentrated in the same runs that yield several hundred-fold concentration of P3A2 (Calzone *et al.*, 1991; unpublished data). (*ii*) Measurements of the molar quantities of P3A2 DNA-binding activity in cleavage- and blastula-stage nuclei were available for comparison with our immunological measurements, from the gel shift titrations of Calzone *et al.* (1988). These measurements demonstrate significant P3A2 DNA-binding activity in blastomere nuclei as early as mid-cleavage (7 hr post-fertilization). (*iii*) Maternal mRNAs encoding both P3A1 and P3A2 are present in the unfertilized egg, and the levels of these increase only modestly during embryogenesis (Cutting *et al.*, 1990). Though relatively rare, we show below that the P3A2 mRNA must be taken into account in order to explain the measured prevalence of this factor after the earliest stages of embryogenesis. Our objectives were to determine, if possible, whether P3A2 or P3A1 is the transcription factor responsible for controlling the oral/aboral ectoderm expression of the *CyIIIa* fusion genes; and also to explore further the hypothesis put forward by Höög *et al.* (1991) that P3A1 and P3A2 might act as antagonistic elements of a regulatory switch. This idea of course requires that the nuclear ratio of the two factors change sharply with respect either to embryonic territory or stage (or both).

The following measurements demonstrate that both P3A1 and P3A2 are maternal transcription factors. In early embryos both are present in the nuclear compartment, but as development proceeds P3A1 essentially disappears from nuclei, and only P3A2 remains as a potentially significant regulatory factor.

MATERIALS AND METHODS

Recombinant protein production

Recombinant P3A1 and P3A2 proteins were purified from bacterial cultures expressing complete or truncated coding sequences, as desired, which had been subcloned into pRSET expression vectors (InVitrogen) in the appropriate reading frame. These expression constructs were derived from the cDNA clones originally described by Höög *et al.* (1991) and Calzone *et al.* (1991). A six-histidine tag is present at the amino terminus of the pRSET fusion proteins, which allows affinity purification by means of nickel-agarose chromatography (Hochuli *et al.*, 1987). For each protein, four 250 ml cultures of bacteria expressing these constructs were grown at 37°C until log phase ($OD_{600} \sim 0.6$). The cultures were then induced with IPTG at a final concentration of 1 mM and grown for an additional 4 hr at 37°C. The bacteria were harvested by centrifugation and initially processed by lysozyme lysis: The bacterial pellet was resuspended in 8 ml of 50 m*M* Tris, pH 8.0, 50 m*M* NaCl, 5 m*M* EDTA, 7.5 m*M* β-mercapto-ethanol, and 800 µl of 10 mg/ml lysozyme was added and thoroughly mixed. After incubating at room temperature for 10 min, the protein solution was frozen at -70°C. On thawing at room temperature, 0.7 vol of urea lysis buffer (8 M urea, 1 *M* NaCl, 50 m*M* Tris, pH 8.0, 7.5 m*M* β-mercapto-ethanol) were added and mixed by vortexing. The mixture was centrifuged for 1 hr at 35,000 RPM to remove insoluble matter, in preparation for affinity purification. A 50% solution of nickel-agarose resin (Qiagen) was

equilibrated with buffer 1 (3 *M* urea, 50 m*M* Tris, pH 8.0, 100 m*M* KCl, 7.5 m*M* β-mercapto-ethanol). The crude protein extract was batch loaded with 8 ml of the equilibrated resin solution for 1 hr at room temperature. The loaded resin was then poured into a Bio-Rad Econo-Prep column, allowed to settle, and washed with 250 ml of buffer 1, then 250 ml of buffer 2 (3 *M* urea, 10 m*M* Tris, 100 m*M* potassium phosphate, pH 6.3, 100 m*M* KCl, 7.5 m*M* β-mercapto-ethanol) and 50 ml buffer 3 (3 *M* urea, 50 m*M* Tris, pH 7.4, 100 m*M* KCl, 10% glycerol, 7.5 m*M* β-mercapto-ethanol). The purified protein was eluted with buffer 4 (3 *M* urea, 300 m*M* imidazole, pH 6.3, 10 m*M* Tris, 100 m*M* KCl, 10% glycerol, 7.5 m*M* β-mercapto-ethanol) and frozen at -70°C. The P3A1 and P3A2 protein fragments used for production of antibodies were purified in a similar manner, except: (*i*) buffers 1-3 contained 8 *M* rather than 3 *M* urea; (*ii*) pelleted bacterial cells were lysed in 8-10 ml buffer 1, then batch loaded to the nickel resin; (*iii*) the urea was removed before elution from the column with a 300 ml linear gradient in which the urea concentration (in buffer 3) was decreased from 8 to 0 *M*; and (*iv*) buffer 4 contained no urea.

Antibody production and immunological blot reactions

The P3A1 rabbit polyclonal antibody was generated against a gel purified fragment of the P3A1 protein consisting of only the zinc finger region (residues 103-274; Höög *et al.*, 1991). The serum was used diluted 1/500-1/1000 for blotting. A second P3A1 rabbit antibody was generated by Cocalico Biochemicals. The antigen was a carboxy terminal fragment of P3A1 (amino acids 211-387), purified as described in the previous section. It showed a pattern of staining similar to the Zn finger antibody. The Caltech Monoclonal Antibody Facility generated a series of monoclonal antibodies to the P3A2 protein. The antigen was the full-length recombinant protein purified as described above. Tissue culture supernatant from clone 7B12-2E8 was diluted 1:1 for blotting. For

antibody blotting, a standard SDS-polyacrylamide protein gel (3 or 5% stacking, 10 or 12% separating) was loaded with samples dissolved in SDS sample buffer (final concentration 10% glycerol, 50 m*M* Tris, pH 6.8, 2% SDS, 1% β-mercapto-ethanol, 0.025% bromophenol blue) heated immediately before loading. After electrophoresis, the gel was electroblotted to a nitrocellulose membrane using standard conditions. The blotted membrane was typically incubated overnight with 5% non-fat dry milk in PBS (pH 7.0) at 4°C. Antibodies were diluted in this same milk solution. The secondary antibody was a goat-anti-(mouse or rabbit)-HRP antibody diluted 1/1000-1/3000. Antibody incubations were generally for 1-2 hr at room temperature. The bound antibodies were detected by chemiluminescence using the Amersham ECL reagents mixed according to the manufacturer's instructions (Amersham). Several film exposures of the blots were made to ensure that an appropriate range of densities was registered on the film. The amounts of purified recombinant protein standards and nuclear samples were optimized for each particular experiment. Protein concentrations were measured by the standard Bradford assay (Bradford, 1976).

Isolation of nuclei, preparation of lysates, and quantitation of samples

Strongylocentrotus purpuratus embryos were grown and harvested for preparation of nuclear extracts as described by Calzone *et al.* (1991). Samples to be collected before hatching were first treated with amino triazole and passed through Nytex filters to remove the fertilization envelopes. Eggs or appropriately staged embryos were collected and were frozen in liquid nitrogen. After thawing, the nuclei were collected by centrifugation at 3000*g*, resuspended in buffer C (10 m*M* HEPES, pH 7.9, 1 m*M* EDTA, 1 m*M* EGTA, 1 m*M* spermidine-Tri-HCl, 1 m*M* DTT) and frozen at -70°C. Nuclear samples were examined under phase contrast microscopy to verify the presence of nuclei. To prepare lysates that were loaded directly on the gels, the following procedures were used: The

lysate was sheared several times with a syringe fitted with a 26-gauge needle followed by 30-60 sec of sonication with a Branson Sonifier 450 with micro-tip. After lysis, the nuclear solution was mixed with 2X SDS sample buffer (Laemmli, 1970), heated briefly to 80°C and loaded immediately onto the SDS polyacrylamide gel. Egg lysates were prepared in a similar manner. Samples were quantitated by the DAPI fluorescence method of Brunk *et al.* (1979), which permits the measurement of minute DNA concentrations in the presence of large amounts of protein. A mass of 1.78 pg per genome (Hinegardner, 1974) was used to convert DNA concentration to the number of nuclei per μl.

Analysis of data

The film recording each experiment was analyzed with a Molecular Dynamics scanning laser densitometer. Standard curves were generated from known concentrations of the respective recombinant proteins that were included on each blot for use in estimating molar quantities of the sea urchin proteins.

RESULTS

Experimental approach

Our objective required the determination of absolute quantities of P3A1 and P3A2 proteins in the nuclear compartments from early to late embryogenesis, and also in unfertilized egg cytoplasm and pronuclei. We found that it was difficult to obtain accurate or reproducible measurements of the cytoplasmic or whole embryo contents of these factors with the methods in hand, for two reasons. First, we were unable to convincingly control cytoplasmic proteolysis of P3A1 and P3A2; and second, one of

these factors is so dilute in the cytoplasm of later embryos that we could not load sufficient lysate to resolve detectable quantities of the factors by gel electrophoresis. However, as will become evident in the following, the nuclear measurements alone are informative.

Nuclei were isolated by a procedure similar to that utilized in this laboratory for the preparation of nuclear extract (Calzone *et al.*, 1988, 1991) in order to facilitate direct comparison with earlier measurements of P3A2 DNA-binding activity. However, the nuclei (or whole eggs), were lysed directly and the P3A1 and P3A2 contents determined, omitting the fractionation procedures used to prepare the nuclear extracts used for gel shift analysis and affinity chromatography (Calzone *et al.*, 1988, 1991). We expressed the number of nuclei represented in each observation in terms of "embryo equivalents," based on the DNA content of given volumes of the nuclear lysates (see Materials and Methods). The P3A1 and P3A2 contents of these samples were estimated immunologically by reference to standards prepared with known amounts of recombinant P3A1 and P3A2. Measurements of P3A1 and P3A2 proteins in nuclear and egg lysates were made on separate gels; several different samples were examined for each developmental time point. For comparative purposes, a complete time course was examined on a single gel along with a set of the appropriate standards. The gels were blotted and the reactions of the antibodies measured by a chemiluminescent detection system, which could be quantitated by densitometry of appropriately exposed samples. Examples of the immunoblots are shown in Fig. 1. It should be noted that adjacent lanes on the blots do not necessarily contain equivalent amounts of material (see legend); as for each stage these had to be adjusted in order to obtain reactions that fell within the linear densitometric reaction range afforded by the corrected standards.

Maternal P3A1 and P3A2

Figure 1 shows that the pronucleus of the unfertilized egg contains both P3A1 and P3A2. A qualitative difference is apparent, however: Though the mobility of P3A2 is the same at all stages and in all preparations, a more slowly migrating maternal form of P3A1 is present in the egg, together with the form present in embryo nuclei throughout development. Both forms are clearly seen in Fig. 1A ("egg nuc"), but mainly the more slowly migrating form is seen in the total egg extracts in Fig. 1B ("egg 1.0, 2.5"). This implies that the more slowly migrating form is mainly cytoplasmic, and could be present in the pronuclear fraction as a contaminant. In fact an egg *cytoplasmic* extract (data not shown) also contains mainly the more slowly migrating form. The functional significance of this cytoplasmic maternal version of P3A1 remains entirely obscure. It could be a product of alternative splicing occurring during oogenesis, although only a single sized P3A1 mRNA was observed in mature egg or embryo RNA preparations (Cutting *et al.*, 1990). More likely, it is a post-translationally modified form of P3A1. As we show below, there is no quantitative requirement that any of the maternal cytoplasmic P3A1 serve as a source of embryo nuclear P3A1. Thus whatever it may be, the role of the more slowly migrating form of P3A1 could be entirely cytoplasmic.

The number of molecules of P3A1 and P3A2 per egg, per pronucleus, and by subtraction, per egg cytoplasm, are shown in Table 1, calculated as we describe from data such as illustrated in Fig. 1. We see that only about 1% of each factor is present in the pronucleus. However, the absolute number of molecules of each factor is quite sufficient to imply significant target site occupancy, given the equilibrium constants measured for these factors (Höög *et al.*, 1991; *cf.* calculations in Calzone *et al.*, 1991). Table 1 also shows that the egg cytoplasm contains about 10X more P3A1 than P3A2. The low mobility form must account for the great majority of this large maternal stockpile of P3A1 protein.

Nuclear content of P3A1 in the embryo

Figure 2 displays the prevalence of P3A1 in the total nuclear compartment of embryos up to the post-gastrula stage (circles), based on the average per nucleus measurements obtained as described above (squares). The measured per nucleus content in the lysate would represent the amount of P3A1 present in all embryo nuclei, were the factor equally distributed to all nuclei (for which there is no evidence, pro or con; thus we can only refer to this as the *average* per nucleus content *in vivo*). At the top of Fig. 2 the maternal content of P3A1 is indicated, merely to facilitate comparison, though again, we have no evidence that this maternal component remains present throughout embryogenesis. The time course depicted in Fig. 2 shows that within a few hours after fertilization there is an initial rise in nuclear P3A1 content from about 1% to 10% of the maternal level. However, considered on a per nucleus basis, this just suffices to keep the average intranuclear content of P3A1 at the same level as in the pronucleus of the unfertilized egg. This rise could be due to transit of maternal P3A1 into the nuclear compartment. However, this is not a necessary presumption, because the amount of P3A1 mRNA present (~2000 molecules/embryo; Cutting *et al.*, 1990) is probably sufficient to account for the increase in nuclear P3A1 over the first 10 hr of development. Thus at approximately two protein molecules min^{-1} $mRNA^{-1}$ (i.e., given a translation rate of ~1.8 codons sec^{-1} and inter-ribosome spacing in sea urchin embryo polyribosomes of 135 nt; reviewed by Davidson, 1986, pp. 75-79) about $2x10^6$ newly synthesized molecules of P3A1 could have accumulated over 10 hr, if the newly synthesized P3A1 were stable. This is just the increase in nuclear compartment content of P3A1 between fertilization and the 10 hr cleavage stage. Figure 2 also shows that after 5 hr (8-cell stage) the average per nucleus quantity of P3A1 begins to *decline* sharply, and by 10 hr it is only about 10% of its initial level. This decline continues, and by the late gastrula

stage P3A1 has virtually disappeared from the nuclear compartment. Nuclear P3A1 is thus most likely to be occupying *cis*-regulatory target sites within the first few hours of development, and it can be of little consequence in late development.

Nuclear content of P3A2 in the embryo

Quantitation of P3A2 is shown in Fig. 3, where the symbolism is the same as in Fig. 2. The most important conclusion is that in contrast to P3A1 the level of P3A2 per average embryo nucleus remains significant at least through the late gastrula period, even in the limit assumption that it is distributed equally to all embryo nuclei (otherwise it would be even higher in some nuclei). Another difference with respect to P3A1 lies in the relations between nuclear P3A2, maternal P3A2, and the P3A2 that could be newly synthesized in the embryo, given the relatively rare mRNA level reported by Cutting *et al.* (1990; a recalculation from their data yields a value of about 1400 molecules per embryo rather than the 700-1000 originally reported). That is the amount of P3A2 in the nuclear compartment of the embryo soon exceeds the maternal component by a factor of about three, so at least two-thirds of the nuclear P3A2 present later in embryogenesis *must* be the product of new translation. In fact, it all could be newly synthesized. This is shown by the dotted line indicating the accumulation of P3A2 on the basis of the number of mRNA molecules per embryo, the translation rate as above, and the assumption of protein stability.

DISCUSSION

Interpretation of prior measurements

We can now better understand several puzzling earlier observations. We could not purify P3A1 from 24 hr nuclear extract by affinity chromatography (see Introduction), because there are over six times more molecules of nuclear P3A2 than of nuclear P3A1 in the embryo at this stage. Since the equilibrium constant of P3A2 for the *SM50* target site used in the affinity purification of P3A2 in our earlier work is ~50X higher than that of P3A1 for this site, no significant P3A1 reaction could have occurred at the ratios of P3A1/P3A2 measured here. We can also compare the amounts of nuclear P3A2 protein that we have obtained here by direct measurement, to the amounts of nuclear P3A2 active in DNA-binding that were calculated from probe excess titration of nuclear extract by Calzone *et al.* (1988). For this it is necessary to take into account that P3A2 most likely binds DNA as a dimer (C. Wilson and F. Calzone, pers. commun.). For 7 hr mid-cleavage embryos, the active P3A2 bound 1.9×10^5 molecules of probe per embryo equivalent of nuclear *extract*, thus indicating ~3.8×10^5 molecules of P3A2 monomer. Extrapolating from Fig. 3, we see that at 7 hr there are indeed about 3×10^5 molecules of P3A2 present per embryo equivalent of nuclear *lysate*. This rather remarkable agreement shows that the nuclear extract of Calzone *et al.* (1988) retained essentially all the P3A2 inactive form in the 7 hr nuclei. The agreement is not as good for 24 hr nuclear extract. Here only about 13% of the amount of P3A2 present, i.e., 7.2×10^7 molecules per embryo equivalent of nuclear lysate (Fig 3), retained activity. Thus, according to Calzone *et al.* (1988) there are 5.6×10^6 molecules of P3A2 active in DNA binding per embryo equivalent of 24 hr nuclear extract. It is possible that the extract contained some denatured or partially proteolyzed P3A2, and hence we believe the conservative direct measurement of the mass of P3A2 shown in Fig. 3 for 24 hr is more

reliable than our earlier estimates based on factor activity. Of course the lower ratio of P3A2 active in DNA binding to P3A2 present could also be due to a biologically significant, covalent inactivation of the factor at the later stage, rather than to a nuclear extract preparation artifact. There are in fact at least five different phosphorylated charge variants of P3A2 in the 24 hr nuclear extract (M. Harrington, R. Zeller and E. Davidson, unpublished data), and the relative activities of these remain to be established.

Developmental changes in the P3A1:P3A2 ratios and a speculation on the regulatory roles of these factors

The quantities of P3A1 (from Fig. 2) and P3A2 (from Fig. 3) per average nucleus, and their molar ratio, are co-plotted in Fig. 4. The nuclear ratio of these factors passes through the value 1.0 soon after cleavage ends, i.e., by the early blastula stage. After this nuclear P3A1 continues to fall, while nuclear P3A2 remains at more or less the same level. While there may be sufficient P3A1 to ensure significant target site binding after the 16 hr cross-over point, and perhaps until late gastrulation *were P3A2 not present*, in any particular nucleus in which the P3A1:P3A2 ratio is much below unity, only P3A2 is likely to be significant with respect to interaction with their common target sites. This argument is based on the equilibrium constant measurements that we reported earlier. Briefly, the relevant results are as follows: Calzone *et al.* (1991) found that the equilibrium constant for the reaction of P3A2 with four different "weak" target sites is within a factor of two of 1.3×10^7 M^{-1}; for certain preferred "strong" sites it is about 10X this; and for adjacent double sites it is 30-50X this. However, Höög *et al.* (1991) showed that P3A1 does not discriminate between the strong or weak sites distinguished by P3A2, or between double or single sites. Thus the equilibrium constants for the reaction of P3A1 with a weak P3A2 site or a strong P3A2 double site are both $1.6\text{-}3 \times 10^7$ M^{-1}. Therefore, even for the weak P3A2 sites, as the P3A1:P3A2 ratio falls below unity, P3A2

will proportionately replace P3A1 in complex, and for the strong P3A2 sites, P3A1 will be replaced by P3A2 at any ratios of P3A1:P3A2 below ten. For multiple sites P3A2 may be bound preferentially at all ratios measured. This leads to a very interesting biological conclusion, *viz.* that unless they are sequestered differentially in the nuclei of the early embryo, P3A1 is likely to be of functional importance *only* in early to mid-cleavage. This corresponds to the period *before* the three known genes, i.e., *SM50*, *Spec1a*, and *CyIIIa* containing P3A sites become transcriptionally active. Thus we propose that P3A1 is a general maternal repressor, which is replaced by P3A2 as development proceeds.

P3A2 is the factor that is directly responsible for negative spatial control of *CyIIIa* expression in the embryonic ectoderm. The two *CyIIIa* target sites through which P3A2 interactions are required to repress ectopic oral ectoderm expression were among those specifically characterized as weak sites by Calzone *et al.* (1991). Thus throughout the period when the negative control function is mediated by these *CyIIIa* P3A sites, only P3A2, and not P3A1, will be relevant. This period extends into late embryogenesis, since even in pluteus-stage embryos, after P3A1 has disappeared, *CyIIIa·CAT* constructs lacking either or both P3A sites display ectopic oral ectoderm expression (C. Kirchhamer and E. Davidson, in preparation). Why then would one negatively functioning factor, i.e., P3A2, replace another, i.e., P3A1, in the early to mid-cleavage embryo? We believe the answer to this apparent paradox is that P3A2 is capable of being facultatively modified so that its repressive activity is confined to the oral blastomeres, wherever these arise. In the aboral ectoderm where the *CyIIIa* gene is expressed, P3A2 may be present but inactive as a repressor, or may be present in a form that is not bound at all. Thus, as the gene is activated, a general repressor (P3A1) is replaced by a factor that exercises a particular spatial control function (P3A2).

There remains the mystery of the enormous maternal (and possibly embryonic) cytoplasmic concentration of P3A1. Most of this must be the low mobility variant of

P3A1 displayed in Fig. 1. One interesting possibility is that P3A1 functions in a manner analogous to the maternal "Y-box" transcription factors of the *Xenopus* oocyte (Wolffe *et al.*, 1992; Tafuri and Wolffe, 1992). These factors (which are not Zn finger proteins) may both regulate genes encoding maternal mRNAs, and also sequester maternal mRNA by binding to it in the egg cytoplasm. Were P3A1 also an RNA-binding protein, this might account for the very large quantity of P3A1 in the egg. A specific prediction and easily testable prediction would be that the P3A1 protein of the egg would display a nonspecific affinity for single-stranded DNA, as do the Y-box factors.

REFERENCES

Bradford, M. (1976). A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal. Biochem.* **72**:248-254.

Brunk, C.F., Jones, K.C. and James, T.W. (1979). Assay for nanogram quantities of DNA in cellular homogenates. *Anal. Biochem.* **92**:498-500.

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

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

Cameron, R.A., Fraser, S.E., Britten, R.J. and Davidson, E.H. (1989). The oral-aboral axis of a sea urchin embryo is specified by first cleavage. *Development* **106**:641-647.

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

Cutting, A.E., Höög, C.H., Calzone, F.J., Britten, R.J. and Davidson, E.H. (1990). Rare maternal mRNAs code for regulatory proteins that control lineage specific gene expression in the sea urchin embryo. *Proc. Natl. Acad. Sci. USA*. **87**:7953-7959.

Davidson, E.H. (1986). *Gene Activity in Early Development*. Academic Press: Orlando, Florida.

DeSimone, S.M. and White, K. (1993). The *Drosophila erect wing* gene, which is important for both neuronal and muscle development, encodes a protein which is similar to the sea urchin P3A2 DNA-binding protein. *Mol. Cell. Biol.* **23**:3641-3649.

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

Hinegardner, R. (1974). Cellular DNA content of the Echinodermata. *Comp. Biochem. Physiol. B.* **49B**:219.

Hochuli, E., Dobeli, H. and Schacher, A. (1987). New metal chelate adsorbant selective for proteins and peptides containing neighboring histidine residues. *J. Chrom.* **411**:177-184.

Hough-Evans, B.R., Franks, R.R., Zeller, R.W., Britten, R.J. and Davidson, E.H. (1990). Negative spatial regulation of the lineage specific CyIIIa actin gene in the sea urchin embryo. *Development* **110**:41-50.

Höög, C.H., Calzone, F.J., Cutting, A.E., Britten, R.J. and Davidson, E.H. (1991). Gene regulatory factors of the sea urchin embryo II. Two dissimilar proteins, P3A1 and P3A2, bind to the same target sites that are required for early territorial gene expression. *Development* **112**:351-364.

Laemmli, U.K. (1970). Cleavage of structural proteins during the assembly of the head of the bacteriophage T4. *Nature* **227**:680-685.

Tafuri, S.R. and Wolffe, A.P. (1992). DNA binding, multimerization, and transcription stimulation by the *Xenopus* Y box proteins *in vitro*. *The New Biologist* **4**:349-359.

Thiebaud, P., Goodstein, M., Calzone, F.J., Thézé, N., Britten, R.J. and Davidson, E.H. (1990). Intersecting batteries of differentially expressed genes in the sea urchin embryo. *Genes Dev.* **4**:1999-2010.

Virbasius, C.A., Virbasius, J.V. and Scarpulla, R.C. (1993). NRF-1, an activator involved in nuclear-mitochondrial interactions, utilizes a new DNA-binding domain conserved in a family of developmental regulators. *Genes Dev.* **7**:2431-2445.

Wolffe, A.P., Tafuri, S., Ranjan, M., and Familari, M. (1992). The Y-box factors: A family of nucleic acid binding proteins conserved from *Escherichia coli* to man. *The New Biologist* **4**:290-298.

Table 1: P3A1 and P3A2 protein molecules per egg.

Stage	**P3A1**	**P3A2**
Whole eggs	2.16×10^7	2.19×10^6
Egg pronuclei	1.46×10^5	1.71×10^4
Egg cytoplasm	2.14×10^7	2.17×10^6

Fig. 1. Examples of quantitative antibody blots from various developmental stages. In each blot, a series of protein standards and nuclear samples were run on a standard SDS protein gel and blotted to a nitrocellulose membrane. Specific antibody binding was detected using a chemiluminescent substrate. Several different exposures were made for each blot to ensure that a proper range of intensities was present for subsequent densitometry scanning: (a) P3A1 and P3A2 protein levels in staged nuclear preparations. The P3A2 standards in this experiment were 3.68 (A), 1.84 (B), 0.92 (C) and 0.37 (D) ng of nickel-purified recombinant protein. The P3A1 standards were 5 (A), 3.5 (B), 2 (C) and 1 (D) ng of nickel-purified recombinant protein. The number of embryo equivalents of nuclei or pronuclei loaded in each P3A2 sample were $1.82x10^5$ (egg pronuclei), $3.90x10^5$ nuclei (5 hr), $3.00x10^4$ nuclei (10 hr), $7.92x10^3$ nuclei (24 hr), $6.47x10^3$ nuclei (36 hr) and $6.34x10^3$ nuclei (48 hr). For P3A1, the embryo equivalents were $1.45x10^5$ egg pronuclei, and $2.08x10^4$ nuclei (5 hr), $9.59x10^3$ nuclei (10 hr), $2.53x10^4$ nuclei (24 hr), $2.07x10^4$ nuclei (36 hr) and $2.05x10^4$ (48 hr) nuclei. The *large* arrowheads show the nuclear forms of P3A1 and P3A2, and the *small* arrow indicates the low mobility maternal form of P3A1 (see text). (b) P3A1 protein levels in whole sea urchin eggs and staged nuclear preparations. The standards were 10, 5, 2.5 and 1 ng of recombinant P3A1 protein. For the whole egg samples, 1081 and 2703 eggs were loaded. The 5 hr time points contained $3.25x10^4$ and $6.49x10^4$ embryo equivalents of nuclear sample. The 10 hr time points contained $1.5x10^4$ and $3x10^4$ embryo equivalents. (c) P3A2 protein levels in whole eggs. The standards in this experiment were 3.68, 1.84, 0.92, 0.37 and 0.19 ng of recombinant P3A2. The number of eggs loaded were $6.06x10^3$, $1.21x10^4$, $2.42x10^4$, $4.85x10^4$ and $9.69x10^4$.

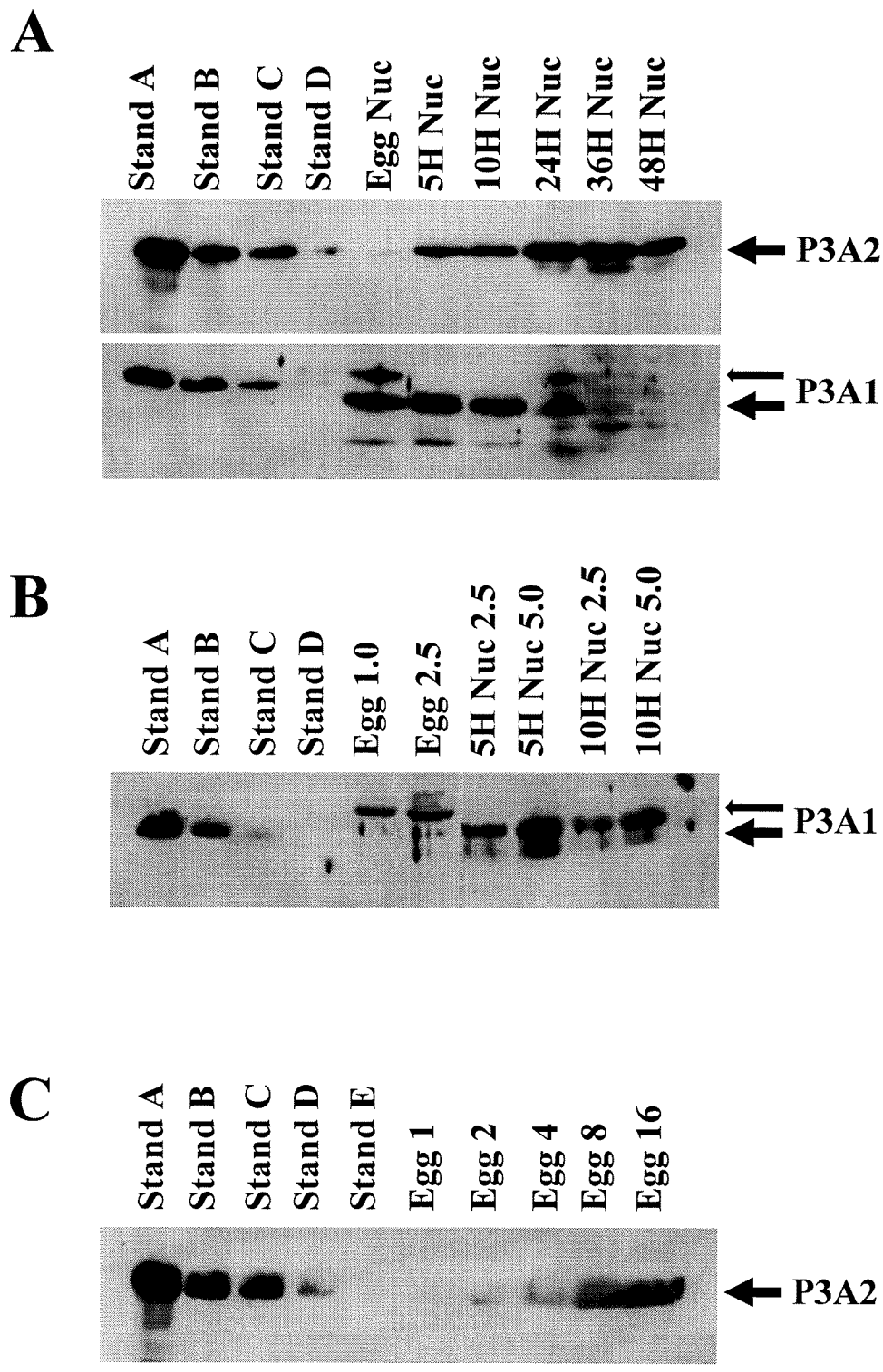

A
Stand A
Stand B
Stand C
Stand D
Egg Nuc
5H Nuc
10H Nuc
24H Nuc
36H Nuc
48H Nuc
P3A2
P3A1
B
Stand A
Stand B
Stand C
Stand D
Egg 1.0
Egg 2.5
5H Nuc 2.5
5H Nuc 5.0
10H Nuc 2.5
10H Nuc 5.0
P3A1
C
Stand A
Stand B
Stand C
Stand D
Stand E
Egg 1
Egg 2
Egg 4
Egg 8
Egg 16
P3A2

Fig. 2. Graphical representation of P3A1 protein levels in sea urchin embryo nuclei. Data points for these graphs were generated from laser densitometer scanning of exposed film from the chemiluminescent antibody detection as in Fig. 1. In this graph, the amount of protein present per nucleus (filled squares) or in the nuclear compartment of a single embryo (filled circles) is plotted on the ordinate and time post-fertilization is plotted on the abscissa. The amount of protein present in the nuclear compartment of an embryo was calculated by multiplying the factor concentration within a nucleus by the total number of cells at the particular developmental stage. The dotted line shows the level of maternal P3A1 protein present in the egg (see Table 1). If all nuclei in an embryo contain the same amount of P3A1, the *measured* amount of P3A1 per nucleus is also the *average* amount per embryo nucleus. The error bars around the measurements represent standard deviations (for these measurements n varies from 3 to 7). The level of P3A1 protein falls from 5 hr (8-cell stage when there are about 1.5×10^5 molecules P3A1/nucleus) to 36 hr (600-cell stage) when there are about 1.7×10^3 molecules/nucleus. The protein is undetectable in this experiment by 48 hr (800-cell stage).

P3A1 Protein in the Embryo

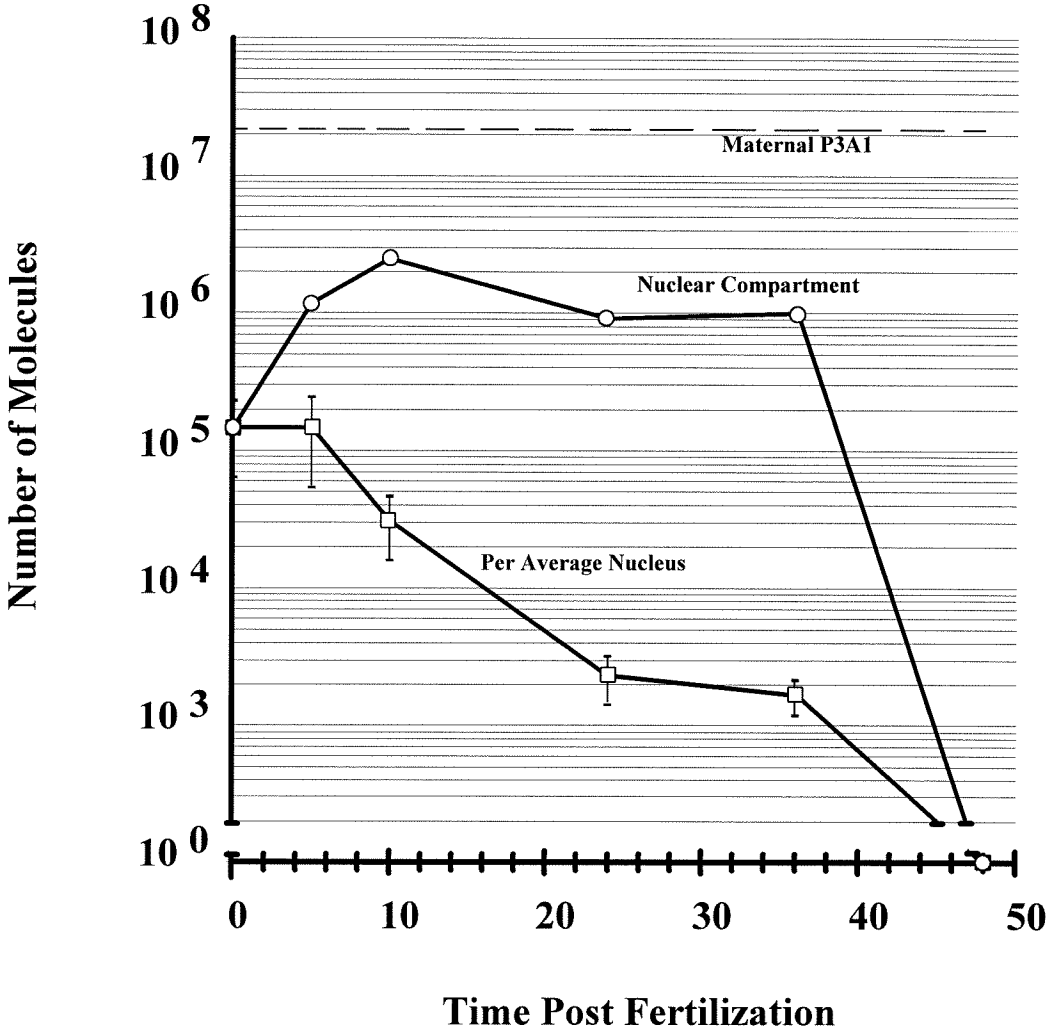


Fig. 3. Graphical representation of P3A2 protein levels in the *S. purpuratus* embryo. The data for this graph were calculated as described for Fig. 2, with standard errors (n = 3 to 7) as shown. The dashed line shows the level of maternal P3A2 protein present in the egg (2.2×10^6 molecules). An estimate of the amount of new protein that could be synthesized by the embryo is also shown (filled triangles). For this calculation, it was assumed that the embryo is able to synthesize two molecules of protein $mRNA^{-1}$ min^{-1} with a lag time of 1 hr to allow for polyribosome loading (Davidson, 1986). The amount of mRNA is recalculated from Cutting *et al.* (1990) and is taken to be 1400 molecules $embryo^{-1}$. The egg pronucleus contains about 1.7×10^4 P3A2 protein molecules. By 24 hr (400-cell stage) the level of protein has peaked at 1.4×10^4 molecules per nucleus, and within a factor of two to three, this level is maintained through 48 hr.

P3A2 in the Nuclear Compartment

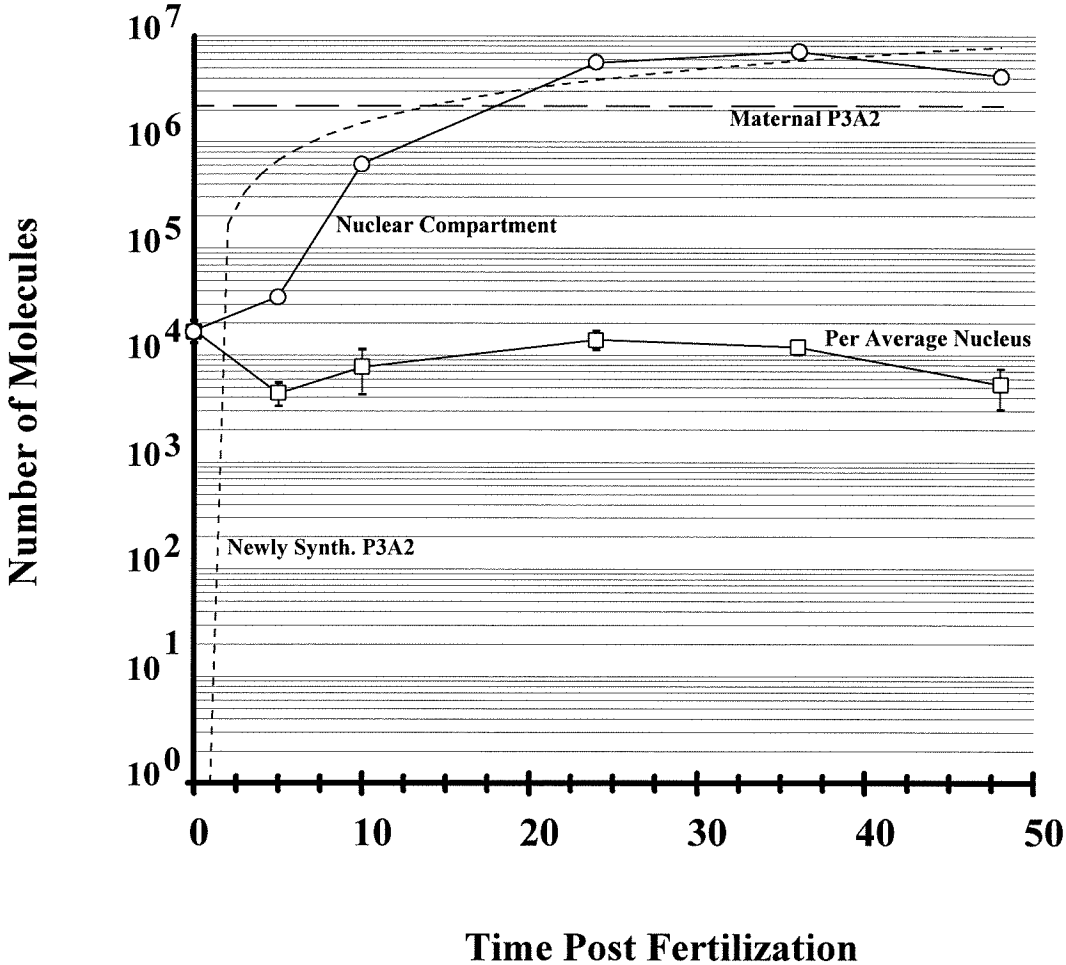
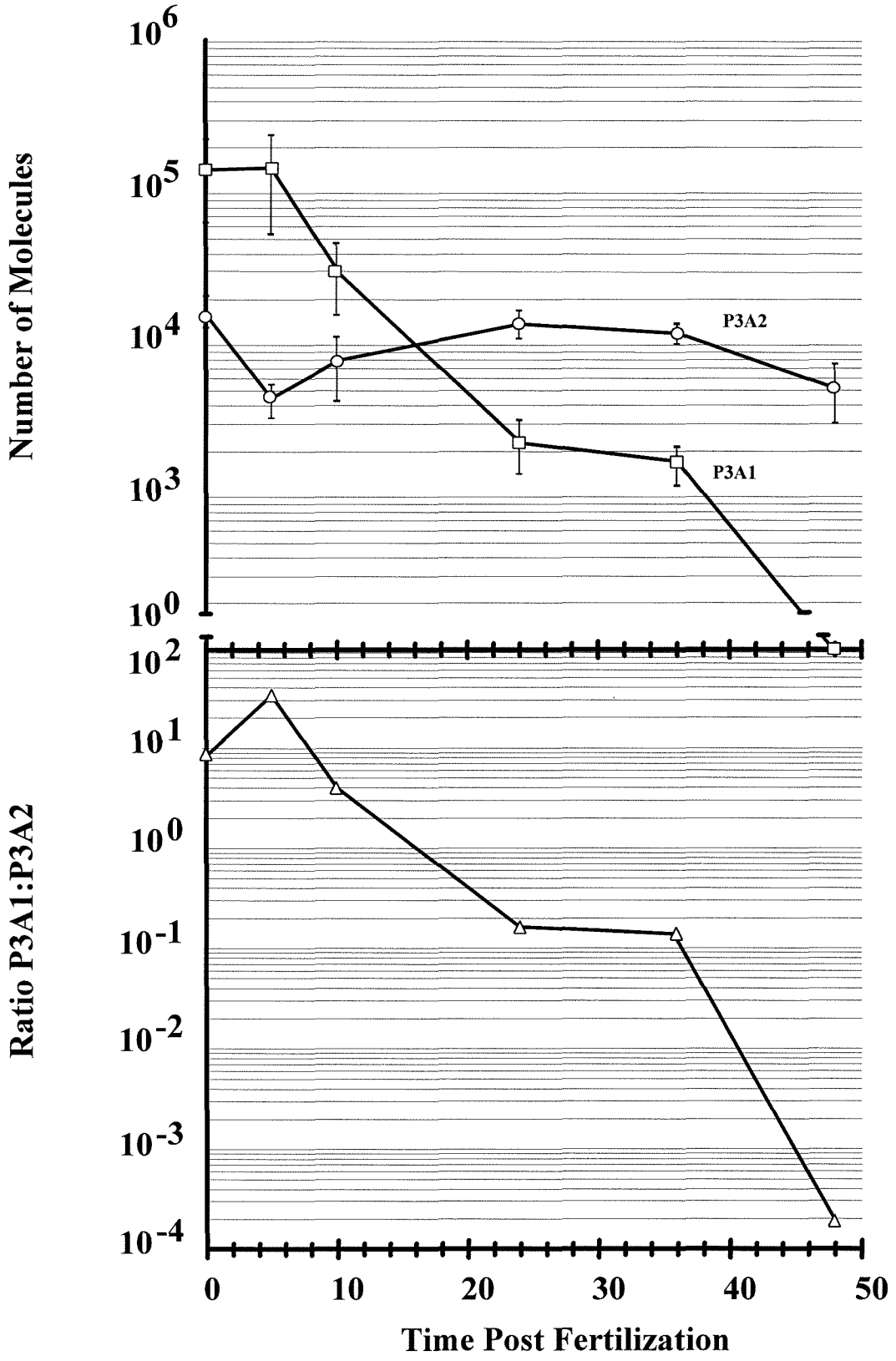


Fig. 4. Graphical representation of P3A1 and P3A2 protein levels in sea urchin embryo nuclei. (a) Levels of P3A1 (filled squares) and P3A2 (filled circles) within the average cell nucleus during development. As described in the legends to Figs. 2 and 3, the level of P3A2 protein remains roughly constant throughout early development while P3A2 protein declines until it is undetectable by 48 hr. (b) Ratio of P3A1 to P3A2 protein in the cell nucleus during development.

P3A1 and P3A2 in the Average Cell Nucleus



Chapter 5

SUMMARY

This thesis has reported the characterization, *in vitro*, of three different transcription factors which interact specifically with the regulatory domains of several genes expressed in the early sea urchin embryo. While much can be learned from the biochemical characterization of DNA-binding proteins, a thorough understanding of the transcriptional control of gene expression in the early embryo requires an in-depth analysis, *in vivo*, of *cis*-regulatory interactions with transcription factors. Since a genetics based analysis of transcription factor function *in vivo* is impractical in the sea urchin, alternative methods for studying *in vivo* function must be developed. Here I will describe several approaches which may provide the information we need to unravel the regulatory network of transcriptional control in sea urchin embryos.

The examination of transcription factor function *in vivo* will involve a variety of analyses that range from observations of morphological changes to modulations of reporter gene activity. A number of morphological landmarks during development can serve as guideposts for effects caused by changes in gene transcription. These include, but are not limited to, the timing and positions of cleavage planes, the ingression of the skeletogenic mesenchyme cells and the subsequent formation of the embryonic spicules, and the process of archenteron invagination. At a molecular level, many antibodies and/or nucleic acid probes exist for territory-specific genes or their products (Chapter 1, table 1 and references therein), which can be used as markers for alterations in both temporal and spatial patterns of endogenous gene expression. Such examination at the cellular level may reveal differences which are not readily apparent from a morphological examination of the entire embryo. In addition, several reporter genes, also expressed in a territory-specific manner (Chapter 1), are available to both qualitatively and quantitatively monitor transcriptional control. An examination of a number of these characteristics will most likely reveal the functions of transcription factors which have been modified by the methods described below.

The synthesis of specific proteins in the embryo may be modified through the use of (1) antisense oligonucleotide technologies and (2) the microinjection of mRNA into zygotes. The introduction of antisense oligonucleotides into a cell allows the investigator to disrupt or prevent the synthesis of a particular protein from its mRNA. When antisense oligonucleotides form hybrids with mRNA molecules, RNAse H, or a similar endonuclease present in embryonic cells, specifically cleaves the RNA/DNA hybrids resulting in the destruction of mRNA (Minshull and Hunt, 1986; Dash *et al.*, 1987). Recently Char *et al.* (1994) reported on the use of antisense oligonucleotide technologies in sea urchin embryos to disrupt the synthesis of the SpOct protein, an octamer like transcription factor present in the early embryo. When sea urchin zygotes were injected with solutions of containing antisense SpOct oligonucleotides, most zygotes arrested before first cleavage. Injections of random sequence oligonucleotides or sense oligonucleotides did not have this effect. The arrested zygote phenotype could be rescued by the co-injection of SpOct mRNA lacking the sequence complementary to the antisense oligonucleotide suggesting that SpOct translation is required for early sea urchin development.

The antisense approach will only be successful for examining transcription factor function, *in vivo*, when new protein synthesis is required. However, Calzone *et al.* (in preparation) has shown that all of the factors which bind specifically to *cis*-elements in the *CyIIIa* regulatory domain are present maternally. Quantitative measurements of the P3A1 factor (Chapter 4) show that there is a large maternal pool of P3A1 protein (3×10^7 molecules per egg), which is more than enough to account for the nuclear levels of protein through the first 48 hr of development. While these experiments did not rule out the possibility of new P3A1 synthesis, it seems unlikely that new synthesis is required. For this reason, one would not expect antisense oligonucleotide injections to

disrupt P3A1 protein levels. The P3A2 protein is also present maternally but at a much lower level than P3A1 (2 X 10^6 molecules per egg, Chapter 4). Measurements of the nuclear levels of P3A2 indicate that new protein synthesis is required; therefore, P3A2 antisense experiments may prove useful.

Specific proteins can be synthesized from mRNAs injected directly into sea urchin zygotes or embryos. This was recently demonstrated by injecting mRNAs encoding either the mouse serotonin or rat muscarinic acetylcholine receptors into zygotes (Cameron *et al.* 1994). When challenged with the appropriate ligands, only injected embryos showed developmental defects indicating that functional protein is produced in the embryo. The successful expression of functional receptors suggests that zygotic mRNA injections may be a useful method for the synthesis of DNA-binding proteins in early embryos. Pilot experiments have been initiated to examine the P3A2 transcription factor, *in vivo* (C. Wilson, pers. commun.). A number of mRNAs encoding both native and mutant P3A2 proteins have been injected into zygotes, and the effects on embryonic development and reporter gene expression are currently being examined.

The mRNA injection experiments are subject to the same difficulties mentioned for the antisense experiments, namely that many sea urchin transcription factors are present maternally. Proteins produced from injected mRNAs must therefore be synthesized in sufficient amounts to "swamp out" the maternal protein pool. DNA-binding proteins present in large maternal pools, for example P3A1, may be relatively "immune" to effects caused by injected transcription factor mRNAs. Additionally, since it takes some time to load polyribosomes onto mRNA (Davidson, 1986), maternal transcription factors would have the opportunity to occupy *cis*-regulatory sites before any new protein synthesis is initiated. By the time sufficient quantities of protein have been produced from injected mRNAs, gene transcription may have been irreversibly initiated

by maternal DNA-binding proteins. While transcription factor mRNA injections may be informative, the technique, like the antisense technology, will probably have a limited usefulness.

Two techniques which may prove effective in overcoming the obstacles presented by maternal pools of sea urchin transcription factors are now discussed. Using either method, transcription factor activity is modulated in the embryo as a whole or in a restricted subset of cells. A number of antibodies to sea urchin transcription factors are becoming available. In some cases, the binding of an antibody to a transcription factor has little effect on DNA binding (e.g., the supershift experiments of Chapter 2). In other cases, the binding of an antibody to a transcription factor can abolish DNA binding altogether presumably by interfering with the DNA binding domain. If such antibodies are injected into sea urchin zygotes, they might also be expected to inhibit the appropriate transcription factors from binding DNA. In addition to zygotic injections, introduction of the antibody to single cells of the cleavage stage embryo may allow the examination of transcription factor function in different territories. While such an approach has not been demonstrated in the sea urchin, several examples in other systems exist. One example, from tissue culture cells, has been reported for the serum response factor $p67^{SRF}$ (Vandromme *et al.*, 1992). The ubiquitous $p67^{SRF}$ protein recognizes the serum response element present in a variety of gene regulatory regions. The protein is constitutively expressed in the nucleus of rat L6 and mouse C2 myogenic cell lines during differentiation. When antibodies directed against the $p67^{SRF}$ protein were injected into these myogenic cells, a number of effects were observed including the failure of myoblasts to form myotubes and the inhibition of myogenin expression. Similar approaches may be made with sea urchin embryos as antibodies specific for sea urchin transcription factors become available.

A second approach which may prove useful in the examination of transcription factor function *in vivo* is the direct injection of DNA-binding proteins into sea urchin embryos or zygotes. This technique offers several advantages over mRNA injections. For example, the delay in protein synthesis caused by polyribosome loading of the mRNA is avoided. Assuming injected factors and maternal factors are equally transported to the nucleus, the exogenous protein has an excellent chance to compete with the endogenous protein for target site occupation. Protein activity can also be examined in vitro, before injection. Such examination of protein activity is not directly possible with the mRNA injection technique where an aliquot of mRNA used as template to synthesize protein *in vitro*, is, at best, an indirect assay of protein activity. Coupled with the antibody technique, DNA-binding protein injections are likely to provide informative data in situations where results from antisense or mRNA injections remain obscure.

Sea urchin embryos are an important system for the biochemical isolation and characterization of transcription factors and will remain so, much as they have been a favorite subject of embryologists for over a century. As new techniques for the examination of transcription factor function *in vivo* are applied, the sea urchin embryo should provide a well-rounded system in which to examine the transcriptional regulatory networks.

REFERENCES

Cameron, R.A., Smith, C.S., Britten, R.J. and Davidson, E.H. (1994). Ligand-dependent stimulation of introduced mammalian brain receptors alters spicule symmetry and other morphogenetic events in sea urchin embryos. *Mech. Dev.* **45**:31-47.

Char, B.R., Tan, H. and Maxson, R. (1994). A POU gene required for early cleavage and protein accumulation in the sea urchin embryo. *Development* **120**:1929-1935.

Dash, P., Lotan, I., Knapp, M., Kandel, E.R. and Goelet, P. (1987). Selective elimination of mRNAs in vivo: Complementary oligodeoxynucleotides promote RNA degradation by an RNAse H like activity. *Proc. Natl. Acad. Sci. USA.* **84**:7896-7900.

Davidson, E.H. (1986). *Gene Activity in Early Development.* Third Edition. Academic Press, Orlando, Florida.

Minshull, J. and Hunt, T. (1986). The use of single-stranded DNA and RNAse H to promote quantitative "hybrid arrest of translation" of mRNA-DNA hybrids in reticulocyte lysate cell-free translations. *Nuc. Acids Res.* **14**:6433-6451.

Vandromme, M., Gauthier-Rouvière, C., Carnac, G., Lamb, N. and Fernandez, A. (1992). Serum response factor $p67^{SRF}$ is expressed and required during myogenic differentiation of both mouse C2 and rat L6 muscle cell lines. *J. Cell Biol.* **118**:1489-1500.

Appendix A

Negative spatial regulation of the lineage specific CyIIIa actin gene in the sea urchin embryo

Barbara R. Hough-Evans, Roberta R. Franks, Robert W. Zeller, Roy J. Britten and Eric H. Davidson

Development 110, 41–50 (1990)
Printed in Great Britain © The Company of Biologists Limited 1990

Negative spatial regulation of the lineage specific CyIIIa actin gene in the sea urchin embryo

BARBARA R. HOUGH-EVANS, ROBERTA R. FRANKS, ROBERT W. ZELLER, ROY J. BRITTEN and ERIC H. DAVIDSON

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

Summary

The CyIIIa · CAT fusion gene was injected into *Strongylocentrotus purpuratus* eggs, together with excess ligated competitor sequences representing subregions of the CyIIIa regulatory domain. In this construct, the chloramphenicol acetyltransferase (CAT) reporter gene is placed under the control of the 2300 nucleotide upstream regulatory domain of the lineage-specific CyIIIa cytoskeletal actin gene. CAT mRNA was detected by *in situ* hybridization in serial sections of pluteus stage embryos derived from the injected eggs. When carrier DNA lacking competitor CyIIIa fragments was coinjected with CyIIIa · CAT, CAT mRNA was observed exclusively in aboral ectoderm cells, i.e. the territory in which the CyIIIa gene itself is normally expressed (as also reported by us previously). The same result was obtained when five of seven different competitor subfragments bearing sites of DNA–protein interaction were coinjected. However, coinjection of excess quantities of either of two widely separated, nonhomologous fragments of the CyIIIa regulatory domain produced a dramatic ectopic expression of CAT mRNA in the recipient embryos. CAT mRNA was observed in gut, mesenchyme cells and oral ectoderm in these embryos. We conclude that these fragments contain regulatory sites that negatively control spatial expression of the CyIIIa gene.

Key words: sea urchin embryo, lineage specific, CyIIIa actin gene, negative spatial regulation.

Introduction

The CyIIIa cytoskeletal actin gene of the sea urchin *Strongylocentrotus purpuratus* is expressed during early embryogenesis only in the eleven cell lineages that constitute the aboral ectoderm of the pluteus (Cox *et al.* 1986; Cameron *et al.* 1989*a*; reviewed by Davidson, 1989). Control of CyIIIa transcription depends on the interaction with endogenous factors of *cis*-regulatory sequences located within a 2300 nucleotide 5′ regulatory domain, as shown by gene transfer experiments carried out with a CyIIIa · CAT fusion gene construct (Flytzanis *et al.* 1987; Hough-Evans *et al.* 1987). Detailed mapping studies have revealed about 20 sites within this domain where nuclear proteins bind with high specificity *in vitro* (Calzone *et al.* 1988; Thézé *et al.* 1990). These fall into 13 distinct sequence classes, suggesting that this number of different factors is required *in vivo*. As part of an effort to determine the biological significance of the individual elements of this complex regulatory system, we have determined the effect on CyIIIa · CAT expression *in vivo* of competition for individual factors or sets of factors. Individual subfragments of the regulatory domain, each containing one or a few factor binding sites, were ligated together with carrier DNA, and coinjected in excess with respect to the CyIIIa · CAT fusion into unfertilized eggs. In the accompanying paper, Franks *et al.* (1990) show that individual subfragments including the specific binding sites for the factors P7I, P5 or P4, each competitively depress CyIIIa · CAT expression, as does a subfragment that includes sites for two different factors called P2I and P2II, and a subfragment that includes three other sites, for factors P8I, P8II and P8III. Interactions with these factors (or a least one of the P2 factors and at least one of the P8 factors) evidently are *required* for normal levels of CyIIIa · CAT expression. Thus these interactions mediate positive regulatory functions. Of the remaining factors, interaction with P3B is necessary for embryonic viability (Franks *et al.* 1990), while no significant competitive depression of the level of CAT expression was observed for subfragments bearing the P7II, P3A, P1 or P6 binding sites.

In this paper, we describe the competitive effects of the CyIIIa regulatory subfragments on the *spatial* rather than the quantitative expression of CyIIIa · CAT. The cell types in which CAT mRNA is expressed in embryos bearing excess competitor fragments were determined by *in situ* hybridization, carried out on serial sections made at an advanced stage of embryonic development.

No qualitative effect on spatial expression was observed with any of the subfragments found by Franks *et al.* (1990) to compete for positively acting factors, *viz.* those listed above. However, coinjection of two wholly nonhomologous competitor subfragments, one containing binding sites for the factors P3B and P3A, and the other binding sites for factor P7II, resulted in a remarkable derangement of the spatial pattern of CyIIIa · CAT expression.

Materials and methods

Eggs and embryos

Gametes were collected and eggs injected as described previously (McMahon *et al.* 1985; Franks *et al.* 1990). The eggs were microinjected with approximately 2 pl of DNA solution, either before or after fertilization, in place in the Petri dish. Embryo samples were taken after hatching (which occurs at 18 h postfertilization) for CAT assays, and at 72 h for *in situ* hybridization.

Microinjection solutions

Solutions for injection contained 40 % glycerol, 250–1500 molecules of a CyIIIa · CAT fusion gene per pl, and competitor and/or sea urchin carrier DNA, exactly as described in detail in the accompanying paper (Franks *et al.* 1990).

CAT assays

CAT enzyme content per embryo was estimated on samples of about 50 hatched blastulae, by a method derived from that of Gorman *et al.* (1982), as described by McMahon *et al.* (1984) and by Franks *et al.* (1990).

Cytological preparation

Preparation of embryo sections and microscope slides were as described previously (Hough-Evans *et al.* 1987, based on the method of Cox *et al.* 1984). Embryos that developed from eggs that had been microinjected with CyIIIa · CAT or with CyIIIa · CAT plus competitor subfragments were fixed at the desired stage (in most cases 72 h pluteus) in 1 % glutaraldehyde (or in 2 % paraformaldehyde, 0.5 % glutaraldehyde) in buffer containing 2.5 % NaCl, 25 mM phosphate buffer pH 7.5. Embryos enclosed in agarose boxes were embedded in Paraplast (Monojet) or Tissueprep 2 (Fisher Scientific) and sectioned at 5 μm, and serial sections were placed on polylysine-coated slides.

In situ *hybridization*

The single-stranded antisense CAT RNA hybridization probes used here were transcribed as described previously (Hough-Evans *et al.* 1987) from a fragment of CAT coding sequence inserted into the polylinker of the pSP65 vector. Either ^{3}H- or ^{35}S-labeled nucleotides were incorporated, as indicated. *In situ* hybridization with tritiated probes was carried out as in earlier experiments by the procedures of Angerer and Angerer (1981) and Cox *et al.* (1984). These procedures were modified for ^{35}S-labeled probes by the addition of 5 mM dithiothreitol and 1 % β-mercaptoethanol to the hybridization solution, hybridization in an N_2 atmosphere, and posthybridization washes that included 1 % β-mercaptoethanol. The photographic emulsion (Kodak NTB-2) was diluted with three parts distilled water in ^{3}H-probe experiments, and a 1:1 dilution was used in experiments carried out with ^{35}S-labeled probe.

Results

Our object in these experiments was to identify regions of the CyIIIa regulatory domain that might be required in normal embryos to *prevent* expression of the CyIIIa gene (i.e. of CyIIIa · CAT), in any but the aboral ectoderm cell lineages. The consequence of effective competition for a regulatory factor that mediates negative control of spatial CyIIIa expression would be to induce ectopic expression in other lineages. Effective competition of course requires that the competitor sequences be present in sufficient quantity within the relevant nuclei to titrate out a biologically significant fraction of the respective factor molecules. Franks *et al.* (1990) showed that this condition can indeed be met for many different positively acting factors. In the experiments described here, embryos taken from the same competitor injection series as studied quantitatively for CAT enzyme expression by Franks *et al.* (1990) were allowed to develop to the 70–72 h pluteus stage, at which stage serially sectioned material provides an unequivocal orientation and identification of anatomical features and tissues.

An extensive series of control experiments confirmed that for the protocols, materials and procedures utilized in this work, CyIIIa · CAT expression is normally confined to cells of the aboral ectoderm. A quantitative summary of the results of all the available CyIIIa · CAT controls is given in Table 1. Using the criterion that label must be present in cells of *two or more adjacent sections* (to avoid artifacts of occasional random sticking of labeled particulates), we found that in this recent set of experiments (Table 1, expts 4–6), 81 % of all serially sectioned embryos display cells containing significant detectable CAT mRNA. *All* of these displayed CAT mRNA in aboral ectoderm, furthermore. Three embryos, representing 2.5 % of those studied in the

Table 1. *Expression of CyIIIa · CAT in pluteus-stage embryos developed from injected eggs*

Experiment no.	Isotope	Labeled/ Scored*	Labeled in aboral ectoderm	Ectopically labeled†
1	^{3}H	6/26	6	0
2	^{3}H	28/50	28	0
3	^{3}H	7/7	7	0
4	^{3}H	9/9	9	0
5	^{35}S	93/118	93	1
6	^{3}H	16/19	16	2
Total:		159/229		3

Embryos labeled in aboral ectoderm/Σ labeled=159/159=100 %.
Embryos labeled ectopically/Σ labeled=3/159=2 %.

Notes:

* Only embryos represented by 9 or more sections were scored. Embryos were scored as labeled only if *labeled cells appeared in two or more adjacent sections*. In later experiments (experiments 1 and 2 are from Hough-Evans *et al.* 1987) a larger proportion of injected embryos were labeled, probably due to improved injection and hybridization procedures.

† Embryos were scored as being ectopically labeled if labeled cells were seen in tissue other than aboral ectoderm, in more than one section.

recent experiments (Table 1), also displayed CAT mRNA in a few cells of other types.

Examples of control embryo sections, in which significant fractions of the visible aboral ectoderm can be seen to contain CAT mRNA, are shown in Fig. 1. Gut, mesenchyme and oral ectoderm are unlabeled in these *in situ* hybridizations. The single injection gene transfer methodology used in these experiments results in mosaic DNA incorporation patterns. Hough-Evans *et al.* (1988) deduced from direct observations made by DNA *in situ* hybridization to the exogenous sequences, and other prior evidence, that most probably, in about three-fourths of recipient embryos, a single concatenate of the injected DNA is stably incorporated in a single blastomere nucleus after 1st, 2nd, 3rd or 4th cleavage. However, exogenous DNA incorporation is random with respect to cell type and cell lineage (Sucov *et al.* 1988; Franks *et al.* 1988*a*,*b*; Livant *et al.* 1988). In some cases, the labeling patterns shown in Fig. 1 of this paper permit us to identify those blastomeres of the early cleavage embryo that in fact incorporated the exogenous DNA, given knowledge of the regional cell lineage of the aboral ectoderm (Cameron *et al.* 1987; 1989*a*,*b*). For example, perusal of other serial sections (not shown) of the embryo illustrated in Fig. 1a, upper right, indicate that the incorporation appears to have occurred in one of the two first cleavage blastomeres, *viz.*, that giving rise to the 3rd cleavage blastomeres **VA**, $\mathbf{VL_R}$ and the overlying **Na** and $\mathbf{NL_R}$ blastomeres, while in the Fig. 1a lower embryo, incorporation appears to have been confined to the single 3rd cleavage **VA** blastomere. This evidence is important in a specific way for our present concerns, since the **VA** progeny that give rise to aboral ectoderm do not segregate from those giving rise to gut until 6th cleavage, and from those giving rise to skeletogenic mesenchyme until 4th cleavage (Cameron *et al.* 1987). Since all aboral ectoderm regions of the **VA** domain express CAT mRNA in this embryo, the exogenous DNA must also have been present in gut (and probably mesenchyme cells). In these locations, however, the CyIIIa·CAT gene is manifestly silent, an excellent demonstration of differential expression. Similar analyses of the lineage affiliations of labeled cells in other embryos shown are given in the legend to Fig. 1. These observations leave no doubt that, as reported earlier, CyIIIa·CAT expression is repressed in gut and mesenchyme, in the same embryos in which it is active in aboral ectoderm. Furthermore, embryos in which large **NL** aboral ectoderm domains are labeled, but the adjacent ciliary band regions of the oral ectoderm do not express CAT mRNA (e.g. Fig. 1e) show that CyIIIa·CAT is not expressed in the progeny of the 5th cleavage **NL1u** and **NL2u** blastomeres, while it is expressed in the progeny of their sister cells, **NL1l** and **NL2l** (Cameron *et al.* 1989*a*). However, this form of evidence cannot be applied to the lack of expression of CyIIIa·CAT in the main facial and stomodeal region of the oral ectoderm, as illustrated clearly in the embryos shown in Fig. 1a–b (lower), 1f, 1i, and 1j. These regions descend from the single 3rd cleavage blastomere **No**, which gives rise only to facial and stomadeal oral ectoderm progeny. Thus, were incorporation of any of the exogenous CyIIIa·CAT DNA to occur in the **No** blastomere (or its progeny), the resulting pluteus would be scored in these experiments as unlabeled. For all incorporations in the **No** domain at or after 3rd cleavage (i.e. the majority; Hough-Evans *et al.* 1988) the expectation would thus be that 12.5% of the successfully injected embryos would display no CAT mRNA, and perhaps this factor largely accounts for the observation that in Table 1 (expts 3–6) 19% of the embryos were scored as silent with respect to CAT mRNA expression.

In vivo *competitions that do not affect spatial expression*

Spatial expression of CyIIIa·CAT was assessed in those samples of embryos that contained the maximum amounts of competitor DNA in the quantitative competition studies of Franks *et al.* (1990). Table 2 shows that in most cases, whether or not CAT expression was depressed by the introduction of competitor fragments, the spatial pattern of CyIIIa·CAT expression was not affected. That is, CAT mRNA was confined to aboral ectoderm just as in the control series summarized in Fig. 1 and Table 1. We conclude that there is no evidence from these experiments that the DNA–protein interactions mediated by the subfragments, which include the sites for factors P8I, II and III; P4; P5; P2I or P2II, affect spatial expression of CyIIIa·CAT *negatively*, because no significant ectopic expression occurred. In each of these cases, we know that the *in vivo* competition was effective (Franks *et al.* 1990). Nor was spatial expression of CyIIIa·CAT affected in the P6 competition sample; however, 'competition' with this subfragment also failed to depress the level of CyIIIa·CAT expression in the experiments of Franks *et al.* (1990). Therefore, for this case we lack independent evidence that the competition was effective. A null control subfragment (designated '0'; Thézé *et al.* 1990) which contains no known sites of DNA–protein interaction, also failed to alter spatial expression when introduced in excess as competitor. Competition with the whole regulatory domain, while stoichiometrically depressing CAT expression (Livant *et al.* 1988; Franks *et al.* 1990), similarly left the spatial pattern of expression unchanged. An obvious interpretation is that titration onto competitor sequences has sequestered required positive regulatory factors, in-

Table 2. *DNA fragments of the CyIIIa control region not affecting spatial regulation*

Fragment	Total no. scored	Embryos labeled in aboral ectoderm*	No. of embryos with (possible) ectopic expression
P2 (I+II)	59	20	0
P4	48	11	1
P5	113	60	0
P6	70	57	0
P8 (I+II+III)	36	25	1

* As in Table 1; see note (*).

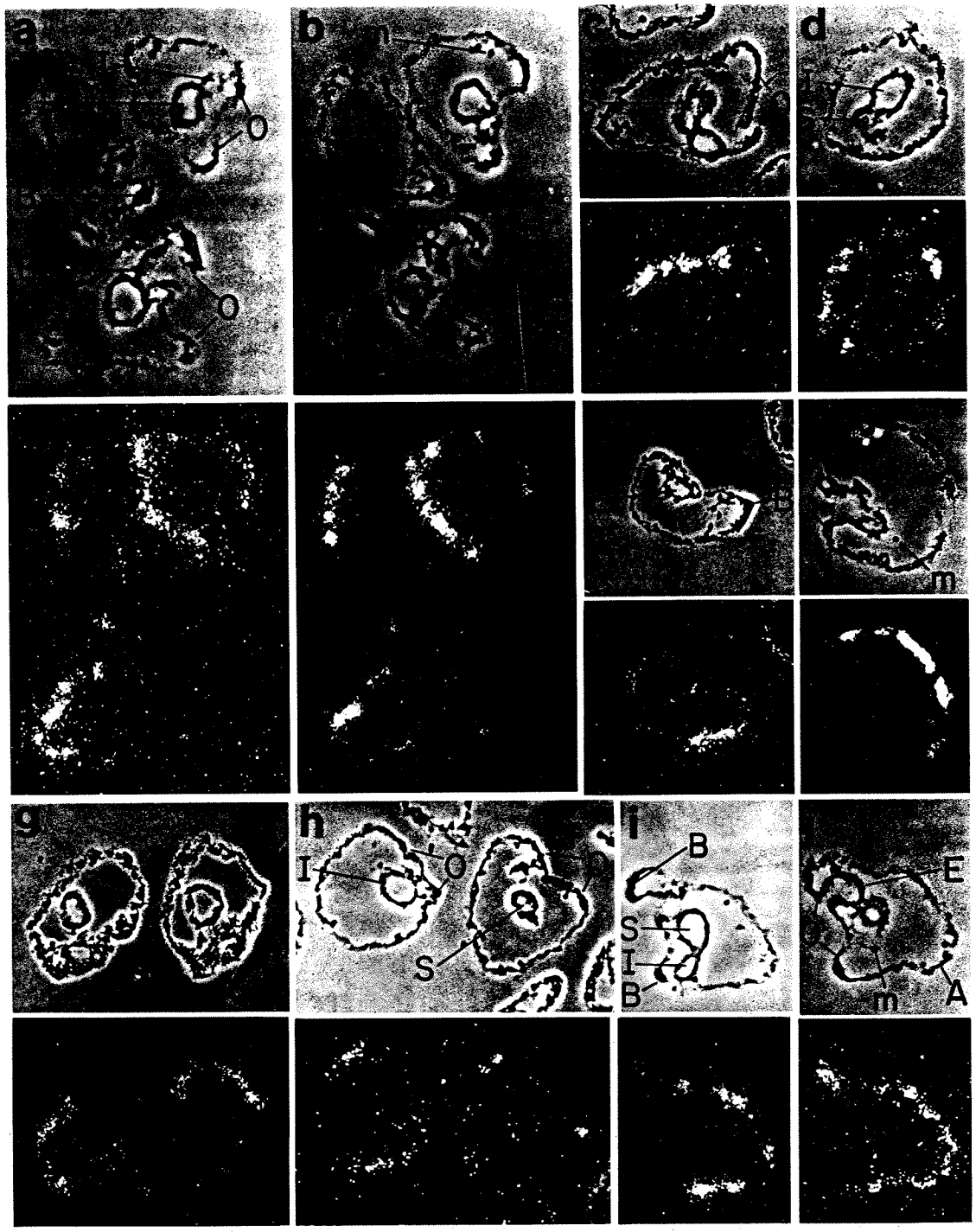

a
O
O
b
m
d
m
g
h
I
O
S
i
B
S
I
B
E
m
A

Fig. 1. Correct lineage-specific expression of the fusion gene CyIIIa · CAT in pluteus embryos that developed from injected eggs. *In situ* hybridization to 5 μm thick embryo sections was carried out using a single-stranded antisense RNA copy of the CAT gene. The probe was labeled with ^{35}S-UTP in panels a, b, and e–j, and with ^{3}H-UTP in panels c and d. Embryo sections are shown under phase-contrast and dark-field illumination. (a and b) Adjacent sections of three embryos that expressed CAT mRNA in many cells of the aboral ectoderm. The upper and lower right embryos are cut approximately in sagittal section; stomach, intestine and esophagus are visible, and the oral 'face' is toward the right. Analysis of serial sections of the upper right embryo establishes that label is found in cells deriving from 8-cell blastomeres **Na**, **VA**, **NL$_R$**, and **VL$_R$** (Cameron *et al.* 1987, 1989*a*). Both the **VA** and **VL** blastomeres also contribute cells to the gut, here completely unlabeled. The lower right embryo appears to be labeled only at the apex, whose cells are clonal descendants of the **VA** blastomere. The remaining embryo (upper left) is cut so that the only tissue seen is aboral ectoderm, except for two small groups of ciliated band cells. (c) Embryo cut lengthwise through stomach and intestine. Oral ectoderm is to the right. The remaining ectoderm, including the only labeled portion, is aboral. The labeling pattern in this embryo indicates that the **Na** blastomere incorporated the exogenous CyIIIa · CAT gene. (d) Section through aboral ectoderm, intestine and part of the stomach, on the left side of the embryo. This side of the ectoderm (derived from blastomeres **NL$_L$** and **VL$_L$**) is extensively labeled. **VL$_L$** also contributes to the gut, which is not labeled. Both mouth (and oral ectoderm) and anus are several sections away on the slide. (e) This embryo is cut in cross-section through the esophagus and coelomic pouches. A small amount of oral ectoderm is present, adjacent to the esophagus, toward the upper right. The only labeled cells are in the patch of aboral ectoderm on the lower right. (f) Approximately longitudinal cut through the open mouth and esophagus (oral face, left). The aboral ectoderm is almost completely labeled on one side, in cells derived from **NL$_R$**. (g) Two embryos cut in cross section. The sections illustrated are toward but not in the oral face, and all the ectoderm seen is aboral. In each embryo one side of the ectoderm is labeled rather extensively, probably in derivatives of an **NL** daughter cell of the 16-cell stage. (h) Two embryos in which only a few cells of the aboral ectoderm are labeled. The embryo on the left is cut in longitudinal section through the intestine. Some **NL$_R$** and **NL$_L$** (lateral ectoderm) descendants are labeled. The embryo on the right is cut sagittally and includes a tangential section of the stomach. In this embryo also some but not all **NL$_R$** and **NL$_L$** derivatives are labeled. The oral face of each embryo is toward the upper right of the figure. (i and j) Approximately midsagittal sections of an embryo that had expressed CyIIIa · CAT in a majority of cells of its aboral ectoderm (as observed in other serial sections not shown). This implies that **Na**, **VA**, **NL$_R$**, **VL$_R$**, **NL$_L$** and **VL$_L$** cells all contain the injected gene. Oral face, left. Note that the labeling stops at the ciliated band (thickened cells at the upper and lower edges of the oral face). The sections shown are 5 μm apart. Particular regions of the embryo sections shown are indicated as follows: A, apex; B, ciliated band; C, coelomic pouch; E, esophagus; I, intestine; M, mouth; m, mesenchyme cell; O, oral ectoderm; S, stomach. Oral ectoderm lies between the two lines labeled 'O'; all the ectoderm not labeled as oral is aboral ectoderm.

Table 3. *Ectopic expression of CyIIIa · CAT when competing P3 or P7II DNA fragments are injected in molar excess*

Fragment (expt)	Total no. embryos*	Total labeled†	Embryos with ectopic label‡	Ectopic as % of labeled embryos
P3 (1)	47	39	20	51
P3 (2)	12	9	2	22
P3 (3)	13	11	6	55
P3 (4)	42	35	8	23
ΣP3	114	94 (82 %)	36	38
P7II (1)	36	35	20	57
P7II (2)	46	43	31	72
P7II (3)	114	110	40	36
ΣP7II	194	188 (97 %)	91	48
CyIIIa · CAT alone§				
(4)	9	9	0	0
(5.6)	137	109 (80 %)	3	3

* Complete embryos and embryos of which 9 or more serial sections were scored.

† Embryos labeled in two or more sections.

‡ Embryos with ectopic labeling in two or more sections.

§ These results also listed in Table 1, as expts 4, 5 and 6 of that Table. Expts 5, 6 of this series were carried out at the same time and on the same batch of eggs as P3 (4) and P7II (3); and expt 4 similarly served as the internal control for expts P3 (3) and P7II (1).

cluding any that might have promoted ectopic expression had they been present.

Ectopic spatial expression of CyIIIa · CAT caused by competition for P3A and P7II factors

The key result shown in this paper is illustrated in Figs 2 and 3, and summarized in Table 3. This is that *in vivo* competition with two subfragments of the CyIIIa regulatory domain results in a striking ectopic spatial expression of CyIIIa · CAT. CAT mRNA in these samples is detected in lineages that normally never express the CyIIIa gene itself, and in which the CyIIIa · CAT construct remains silent in the absence of competitor (except for 2.5 % or less of possible aberrant cases). One of these subfragments, here referred to as the P3 subfragment, contains two known sites for high specificity DNA-binding proteins, *viz.* the P3A and the P3B sites (Calzone *et al.* 1988; Thézé *et al.* 1990). The P3B site is an octamer protein binding site, and since competitive titration with this site results in embryonic lethality (Franks *et al.* 1990), the endogenous octamer protein is obviously a necessary factor. The site responsible for the specific spatial control function that is perturbed in the P3 competition experiment is most probably the P3A site. Neither a subfragment containing only this site nor the P7II site causes significant competitive decrease in CyIIIa · CAT expression even when present in what should be sufficient excess (Franks *et al.* 1990), and thus by this test neither functions positively *in vivo*. The *in situ* hybridization results obtained with the P3 and P7II competitions

indicate that in fact both interactions must function *negatively* in normal undisturbed embryos, presumably *preventing* CyIIIa expression in other than the aboral ectoderm cell lineages. When excess molar quantities of either of these binding sites are introduced into the embryos, the result is a failure of these normal repressive interactions, as the factors become sequestered on the competitor fragments. Table 3 shows that from 22% to 55% of embryos in the P3 competition series, and 36% to 72% in the P7II competition series, have at least some cells that display illegitimate expression of CAT mRNA.

The effects of competition with the P3 and P7II subfragments are subtly different. Fig. 2 shows representative examples of ectopic expression with the P3 subfragment. In most cases the ectopic expression is confined to mesenchyme cells, and in some examples a significant fraction of the mesenchyme cells present in the section display CAT mRNA (e.g. Fig. 2d and h). Unfortunately we cannot clearly distinguish skeletogenic mesenchyme from secondary mesenchyme, since the mineral elements of the skeleton are dissolved during preparation. From their positions, however, both would seem to be expressing CAT mRNA: thus for example the active mesenchyme cell along the aboral wall of the embryo shown in Fig. 2h is likely to be skeletogenic, while those applied to the wall of the gut as in Fig. 2c, d, f and g are more probably secondary, but these assignments are certainly not secure. Fig. 2 also includes some very clear examples of labeling in the facial oral ectoderm, which descends entirely from the progeny of the 3rd cleavage **No** blastomere (e.g. in Fig. 2a, b and e). We did not see any convincing cases among the P3 competition embryos of labeling in the wall of the gut, however. Possible labeling in either stomach or intestine was confined to occasional single cells that could not be verified or supported by labeling in the adjacent sections. Expression occurred in the aboral ectoderm in almost all the P3 competition embryos, whether or not they displayed ectopic CAT mRNA as well, as can be seen, for example in the sections shown in Fig. 2d, e and h. We conclude that the P3 subfragment contains a regulatory element the function of which is required to repress CyIIIa expression in mesenchyme cells and oral ecoderm.

The effects of competition with the P7II fragment, as shown in Fig. 3, are even more dramatic. The central **No** domain of the oral ectoderm is strongly labeled in Fig. 3a and b (the same embryo seen in adjacent sections), and in Fig. 3c–f; and one side of the ciliary band, which derives from an NL blastomere (Cameron *et al.* 1989*a*) is strongly labeled in the embryos shown in Fig. 3h and j. Many of the mesenchyme cells to be seen, for example, in the embryo shown in Fig. 3g, are also labeled. Intense labeling of cells in the wall of the gut can be seen in the embryos shown in Fig. 3c, d and i. Again, almost all of the labeled embryos included in the P7II experiments of Table 3 display aboral ectoderm labeling as well, as can be seen in Fig. 3a–h. We conclude that the P7II interaction is required to repress CyIIIa · CAT expression (and thus presumably CyIIIa expression) in *all* major embryonic lineages except aboral ectoderm. Competitive interference with this interaction produces a catastrophic spatial derangement of expression, in which expression in other lineage element(s) wherein the exogenous DNA resides is added to the normal pattern of aboral ectoderm expression.

Discussion

Negative spatial regulatory functions appear to be confined to the P3A and P7II sites of the CyIIIa gene

Franks *et al.* (1990), in the accompanying study, failed to detect any competitive decrease in expression of CAT mRNA from the CyIIIa · CAT fusion when the P3A or P7II subfragments were coinjected in molar excess. From the quantities of these factors estimated in the embryo nuclei *in vivo* (Calzone *et al.* 1988), it appeared unlikely that this negative result could be due to insufficient excess of the competitor sequence. The results reported here (Table 3 and Figs 2 and 3), demonstrating a frequent, and often extensive ectopic spatial expression of CAT mRNA in the P3 and P7II competition samples, prove that the quantities of stably incorporated competitor sequence indeed sufficed to titrate out the respective factors, at least in some cells of many embryos. It follows that by the competition test these interactions cannot function positively in regulating CyIIIa · CAT expression. Present evidence suggests that the sole function of these interactions is indeed negative, *viz.*, *repression* of CyIIIa expression except in the embryonic territory composed of the eleven aboral ectoderm cell lineages.

We have in this work and that reported by Franks *et al.* (1990) examined subfragments of the CyIIIa regulatory domain that include all of the known sites of high-specificity DNA–protein interaction detected *in vitro* (Calzone *et al.* 1988; Thézé *et al.* 1990). Only one subfragment, that including the two sites bound by the factor P6, failed to result *either* in ectopic spatial expression *or* decrease in CAT expression when coinjected in molar excess with the CyIIIa · CAT fusion. We did not examine the effect of P1 site competition on spatial expression in this work, and it is not excluded that this interaction might also exercise negative spatial control. However, this seems unlikely since, as Franks *et al.* (1990) found, P1 acts as a weak positive regulator of CyIIIa · CAT expression.

Spatial regulation of lineage-specific CyIIIa gene expression

Both of the sites that according to these results would mediate negative spatial regulation of the CyIIIa gene are located close to one or more positively acting sites. This is particularly clear in the case of the P7II site, which is only about 100 nucleotides to the 3′ side of the P7I site, that was demonstrated in the accompanying paper to function positively. This is less than the amount of DNA bound by a single nucleosome. Thus as in other systems in which negative regulatory interac-

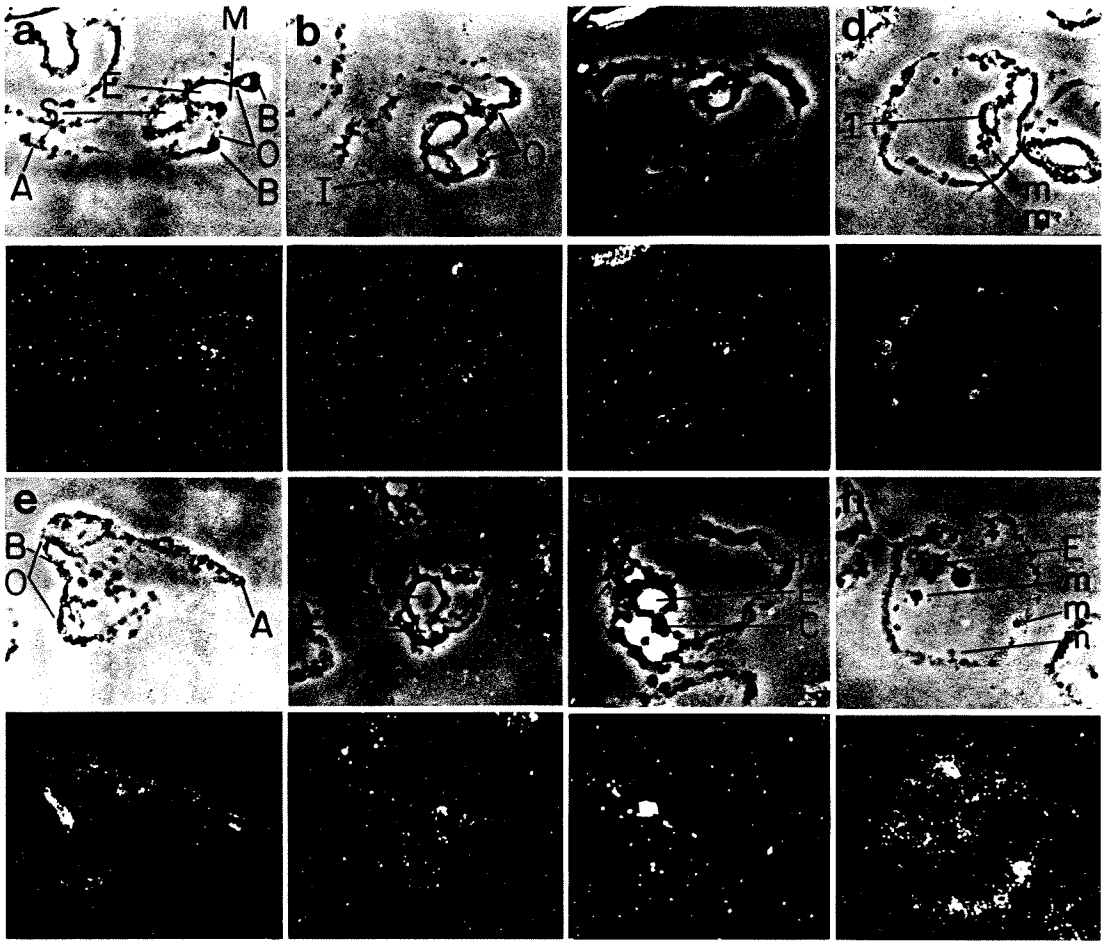
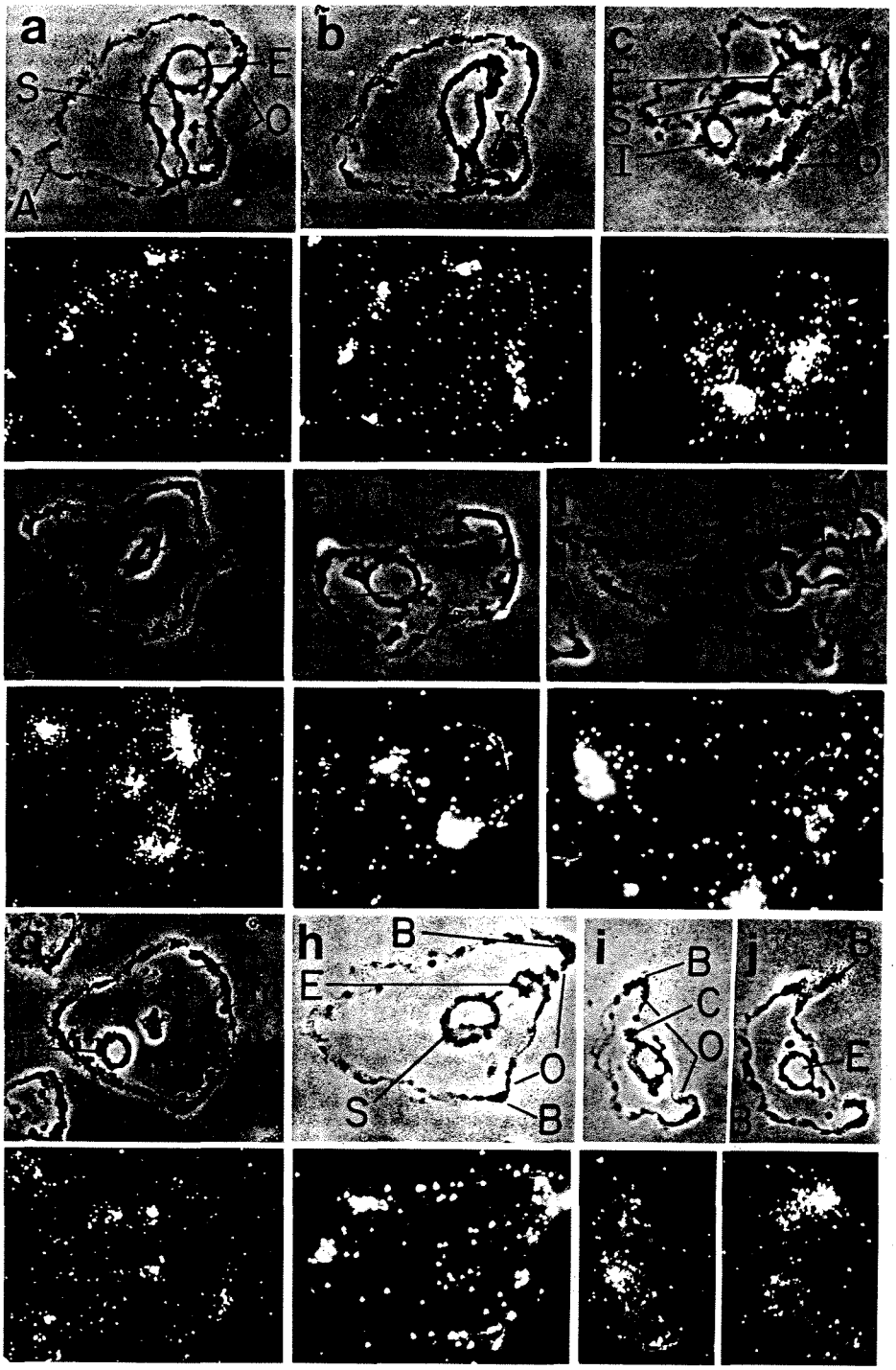


Fig. 2. *In situ* hybridization showing ectopic CAT expression in embryos from eggs injected with P3 DNA fragments in competition with the CyIIIa · CAT fusion gene. All the panels except h show hybridization with tritium-labeled CAT probes; in h the probe was labeled with ^{35}S-UTP. Embryos shown were collected in three completely independent competition injection experiments: (1) a–d; (2) e and h; (3) f and g. Embryo regions shown: A, apex; B, ciliated band; C, coelomic pouch, E, esophagus; I, intestine; M, mouth; m, mesenchyme cell; O, oral ectoderm; S, stomach. (a) Midsagittal section of an embryo labeled in oral ectoderm (near the open mouth) as well as aboral ectoderm (labeling seen in one cell just aboral (left) of the upper ciliated band; in other sections more of the aboral ectoderm was labeled). Esophagus and stomach are visible in this section. (b) Same embryo as shown in panel a, second section further along in the series. The oral ectoderm is labeled in several contiguous sections of this pluteus. (c) Longitudinal section through intestine, oral (upper) ectoderm, and aboral ectoderm. The only labeling is in a mesenchyme cell attached to the gut. (d) Mesenchyme cell labeling in an embryo cut in cross section. The intestine is cut tangentially; all the ectoderm in this section is aboral ectoderm. Two labeled mesenchyme cells are shown. In addition, this embryo is labeled in some cells of the aboral ectoderm. (e) Tangential section through the right side of an embryo in which the ciliary band bordering the oral face is shown on the left, heavily labeled. The remaining tissue in this section is aboral ectoderm. (f) This embryo was sectioned through the stomach and esophagus. A single labeled mesenchyme cell is shown. (g) Longitudinal section (oral face to the left) through esophagus and the wall of one coelomic pouch. A mesenchyme cell above the esophagus is labeled. (h) Embryo similar to that shown in d. Some of the mesenchyme cells seen in the phase photograph are labeled, while others are not. All of the ectoderm in this cross section is aboral. A tangential section of part of the esophagus is completely unlabeled.

tions play an important role, a likely mechanism would be that the negative function operates by interfering with or modifying a physically contiguous positive interaction. The P7I interaction might thus serve as the essential (positive) element of the repression submechanism mediated by the P7II factor. Similarly, the P3A binding site is but 13 nucleotides distant from the P3B octamer site on the 3′ side, and only 90 nucleotides



distant from the potent, positively functioning P5 binding site (Calzone *et al.* 1988; Thézé *et al.* 1990; Franks *et al.* 1990) on the 5′ side.

We cannot state on the basis of this or other extant data whether there are positive *as well as* negative lineage-specific controls on spatial expression of these

Fig. 3. *In situ* hybridization of sections of pluteus stage embryos that developed from eggs injected with CyIIIa · CAT and P7II competitor. An antisense RNA probe to CAT mRNA was labeled with ^{3}H-UTP (panels a, b and g) or ^{35}S-UTP (c, d, e, f, h and i). Photographs taken under phase-contrast (upper) and dark-field (lower) illumination. Embryos are labeled as in Figs 1 and 2. (a and b) Adjacent 5 μm sections of an embryo labeled in both oral and aboral ectoderm. This labeling pattern implies that clonal descendants of **Na** and **No** blastomeres incorporated the CyIIIa · CAT gene. The apex ectoderm and gut are unlabeled; therefore no cells derived from **V** blastomeres contain exogenous DNA. Oral ectoderm is to the right; the three parts of the gut, esophagus, stomach and intestine, are clearly distinguishable. (c) Oblique section through a pluteus in which the intestine and oral ectoderm are labeled, as well as a portion of the aboral ectoderm adjacent to the anus. These regions are derived from the **No** and **VO** blastomeres, indicating that the CyIIIa · CAT DNA was probably incorporated into the oral precursor cell at the 4-cell stage. Additional labeling in the right lateral ectoderm (sections not shown) may mean that some $\mathbf{VL_R}$ (right anal plate) descendants also contain CyIIIa · CAT. (d) Gut (intestine) and oral ectoderm labeling in an embryo cut in oblique cross section. The oral face is up; aboral (lateral) ectoderm is also labeled. Our interpretation of the labeling is that some **No** and $\mathbf{NL_L}$ daughter cell clones contain the injected gene. (e) Similar section (of a different embryo) cut through the esophagus. The oral face is up. Oral ectoderm (**No**), and aboral (**Na**) ectoderm as well, are labeled. (f) Sagittal section showing aboral ectoderm labeling near the embryo apex. Since the gut is not labeled in this embryo, only apical ectoderm-forming **VA** clones contain CyIIIa · CAT. The open mouth, part of the gut and labeled oral ectoderm (**No**) are to the right. In later sections the right lateral ectoderm ($\mathbf{NL_R}$ derivatives) is also labeled. (g) Labeled mesenchyme cells, in an embryo cut in cross section. Aboral ectoderm is also labeled, in a few cells, evidence of a second incorporation event. (h) Pluteus labeled in the upper part of the ciliated band (an oral ectoderm, **No**, derivative) and a few cells of aboral ectoderm. (i and j) Two adjacent longitudinal sections through upper gut (mouth, esophagus and coelomic sac) and arms. Labeling is confined to the right side of the embryo, and is absent from gut and apex; the precursors are descendants of $\mathbf{NL_R}$ and $\mathbf{VL_R}$. The oral face is to the right, and the tips of the arms carry cells of the ciliated band. Mouth and the ciliated band of the upper left arm are labeled. In other sections of the same embryo (not shown) the lower left ciliated band, but not the aboral ectoderm, was labeled.

aboral ectoderm-specific genes. Were such to exist, competitive titration *in vivo* (or deletion, or *in vitro* mutations) would merely have decreased or abolished the signal observed, as in the cases recorded in Table 2, without altering the spatial pattern of expression. However, that same result would be expected from competition for factors that have no spatial regulatory functions whatsoever, but serve as enhancers or positive temporal regulators, etc. Even if further experiments reveal that there are indeed lineage-specific, positively acting factors, the negative functions displayed in this study are a dominant regulatory feature, so that loss of these functions ruins the lineage specificity. This is also seen in the interspecific gene transfer experiment (Franks *et al.* 1988*b*). The heavily ectopic pattern of CyIIIa · CAT expression observed in *Lytechinus variegatus* embryos in that study strongly resembles that produced in the P7II competition series of Fig. 3 of this paper, in that expression occurs in all major lineage elements. The direct prediction follows that in *L. variegatus* nuclei the P7II factor is either missing or is evolutionarily divergent, so that it can no longer function on the P7II site of the *S. purpuratus* CyIIIa gene. It appears unlikely that failure of the P3A interaction alone to occur would suffice to explain the phenotype of CyIIIa · CAT expression in *L. variegatus* embryos since competitive titration of P3A in *S. purpuratus* does not generate ectopic expression in the gut (Fig. 2). The fact that competitive interference with the P7II interaction alone derepresses CyIIIa · CAT expression in all major lineages suggests that the regulatory system logic is of the form that repression in oral ectoderm and mesenchyme requires *both* the P3A and P7II interactions, while repression in gut requires only the P7II interaction. Alternatively, these two interactions could have similar functions, but there could be such large concentrations of P3A factor in gut cells in particular that the incorporated competitor sequences were quantitatively insufficient to titrate them out.

In summary, the major result of this work is the discovery of lineage-specific, negative regulatory interactions in the sea urchin embryo. These interactions are an essential feature of the mechanism by which is achieved the initial differential spatial expression of a structural gene, that in normal development is transcribed exclusively in the aboral ectoderm lineages of the early embryo.

This research was supported by a grant from the National Institutes of Health [HD-05753]. RWZ is supported by an NIH Training Grant [GM-07616].

References

ANGERER, L. M. AND ANGERER, R. C. (1981). Detection of poly(A)$^{+}$ RNA in sea urchin eggs and embryos by quantitative *in situ* hybridization. *Nuc. Acid Res.* **9**, 2819–2840.

CALZONE, F. J., THÉZÉ, N., THIEBAUD, P., HILL, R. L., BRITTEN, R. J. AND DAVIDSON, E. H. (1988). Developmental appearance of factors that bind specifically to *cis*-regulatory sequences of a gene expressed in the sea urchin embryo. *Genes & Dev.* **2**, 1074–1088.

CAMERON, R. A., FRASER, S. E., BRITTEN, R. J. AND DAVIDSON, E. H. (1989*a*). Segregation of oral from aboral ectoderm precursors is completed at 5th cleavage in the embryogenesis of *Strongylocentrotus purpuratus*. *Devl Biol.* **137**, 77–85.

CAMERON, R. A., FRASER, S. E., BRITTEN, R. J. AND DAVIDSON, E. H. (1989*b*). The oral-aboral axis of a sea urchin embryo is specified by first cleavage. *Development* **106**, 641–647.

CAMERON, R. A., HOUGH-EVANS, B. R., BRITTEN, R. J. AND DAVIDSON, E. H. (1987). Lineage and fate of each blastomere of the eight-cell sea urchin embryo. *Genes & Dev.* **1**, 75–85.

COX, K. H., ANGERER, L. M., LEE, J. J., BRITTEN, R. J., DAVIDSON, E. H. AND ANGERER, R. C. (1986). Cell lineage-specific programs of expression of multiple actin genes during sea urchin embryogenesis. *J. molec. Biol.* **188**, 159–172.

COX, K. H., DELEON, D. V., ANGERER, L. M. AND ANGERER, R. C. (1984). Detection of mRNAs in sea urchin embryos by *in situ*

hybridization using asymmetric RNA probes. *Devl Biol.* **101**, 485–502.

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

FLYTZANIS, C. N., BRITTEN, R. J. AND DAVIDSON, E. H. (1987). Ontogenic activation of a fusion gene introduced into the sea urchin egg. *Proc. natn. Acad. Sci. U.S.A.* **84**, 151–155.

FRANKS, R. R., HOUGH-EVANS, B. R., ANDERSON, R. F., BRITTEN, R. J. AND DAVIDSON, E. H. (1990). Competitive titration in living sea urchin embryos of regulatory factors required for expression of the CyIIIa actin gene. *Development* **110**, 00–00.

FRANKS, R. R., HOUGH-EVANS, B. R., BRITTEN, R. J. AND DAVIDSON, E. H. (1988*a*). Direct introduction of cloned DNA into the sea urchin zygote nucleus, and fate of injected DNA. *Development* **102**, 287–299.

FRANKS, R. R., HOUGH-EVANS, B. R., BRITTEN, R. J. AND DAVIDSON, E. H. (1988*b*). Spatially deranged though temporally correct expression of a *Strongylocentrotus purpuratus* actin gene fusion in transgenic embryos of a different sea urchin family. *Genes & Dev.* **2**, 1–12.

GORMAN, C. M., MOFFAT, L. F. AND HOWARD, B. H. (1982). Recombinant genomes which express chloramphenicol acetyltransferase in mammalian cells. *Mol. Cell Biol.* **2**, 1044–1051.

HOUGH-EVANS, B. R., BRITTEN, R. J. AND DAVIDSON, E. H. (1988). Mosaic incorporation and regulated expression of an exogenous gene in the sea urchin embryo. *Devl Biol.* **129**, 198–208.

HOUGH-EVANS, B. R., FRANKS, R. R., CAMERON, R. A., BRITTEN, R. J. AND DAVIDSON, E. H. (1987). Correct cell type-specific expression of a fusion gene injected into sea urchin eggs. *Devl Biol.* **121**, 576–579.

LIVANT, D. L., CUTTING, A. E., BRITTEN, R. J. AND DAVIDSON, E. H. (1988). An *in vivo* titration of regulatory factors required for expression of a fusion gene in transgenic embryos. *Proc. natn. Acad. Sci. U.S.A.* **85**, 7607–7611.

MCMAHON, A. P., FLYTZANIS, C. N., HOUGH-EVANS, B. R., KATULA, K. S., BRITTEN, R. J. AND DAVIDSON, E. H. (1985). Introduction of cloned DNA into sea urchin egg cytoplasm: Replication and persistence during embryogenesis. *Devl Biol.* **108**, 420–430.

MCMAHON, A. P., NOVAK, T. J., BRITTEN, R. J. AND DAVIDSON, E. H. (1984). Inducible expression of a cloned heat shock fusion gene in sea urchin embryos. *Proc. natn. Acad. Sci. U.S.A.* **81**, 7490–7494.

SUCOV, H. M., HOUGH-EVANS, B. R., FRANKS, R. R., BRITTEN, R. J. AND DAVIDSON, E. H. (1988). A regulatory domain that directs lineage-specific expression of a skeletal matrix protein gene in the sea urchin embryo. *Genes & Dev.* **2**, 1238–1250.

THÉZÉ, N., CALZONE, F. J., THIEBAUD, P., HILL, R. L., BRITTEN, R. J. AND DAVIDSON, E. H. (1990). Sequences of the CyIIIa actin gene regulatory domain bound specifically by sea urchin embryo nuclear proteins. *Mol. Reprod. and Dev.* **25**, 110–122.

(*Accepted 1 June 1990*)

Appendix B

Gene regulatory factors of the sea urchin embryo I. Purification by affinity chromatography and cloning of P3A2, a novel DNA-binding protein

Frank J. Calzone, Christer Höög, David B. Teplow, Ann E. Cutting, Robert W. Zeller, Roy J. Britten and Eric H. Davidson

Development 112, 335–350 (1991)
Printed in Great Britain © The Company of Biologists Limited 1991

Gene regulatory factors of the sea urchin embryo

I. Purification by affinity chromatography and cloning of P3A2, a novel DNA-binding protein

FRANK J. CALZONE*, CHRISTER HÖÖG†, DAVID B. TEPLOW, ANN E. CUTTING, ROBERT W. ZELLER, ROY J. BRITTEN and ERIC H. DAVIDSON

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

*Present address: Department of Developmental and Cell Biology, University of California, Irvine, Irvine, CA 92717, USA
†Present address: Department of Molecular Genetics, Karolinska Institutet, S-10401 Stockholm, Sweden

Summary

The P3A2 regulatory protein interacts with specific sites in the control region of the *CyIIIa* actin gene. Previous studies showed that this interaction is required to confine expression of a *CyIIIa · CAT* fusion to the aboral ectoderm, the embryonic territory in which *CyIIIa* is normally utilized. P3A2 also binds specifically to similar target sites located in the regulatory region of the *SM50* gene, which is expressed only in skeletogenic mesenchyme lineages. The P3A2 factor was purified by affinity chromatography from nuclear extracts of 24 h sea urchin embryos, and partial peptide sequences were used to isolate a cDNA clone encoding the complete protein. There are no significant similarities between P3A2 and any other protein in existing sequence data bases. P3A2 thus includes a novel type of DNA-binding domain. To examine the differential utilization of P3A2 in *CyIIIa* and *SM50* genes, we measured the specific affinity of this protein for the various target sites in the regulatory DNAs of each gene, and identified the core target site sequences. The stability of P3A2 complexes formed with *SM50* target sites is 50–100 times greater than that of the complexes formed with *CyIIIa* target sites, though the factor binds to very similar core sequence elements. P3A2 is one of at least twelve different proteins whose interaction with *CyIIIa* regulatory DNA is required for correct developmental expression. The results reported demonstrate that it might be possible to purify most of these regulatory proteins, or any other specific DNA-binding proteins of the sea urchin embryo, by using the simple procedures described for P3A2.

Key words: regulatory protein, embryonic gene regulation, *cis*-regulatory target site.

Introduction

The early sea urchin embryo develops in a relatively simple manner in which lineages descendant from invariant cleavage-stage founder cells clonally express specific sets of genes. These lineages construct five 'territories,' defined in terms of cell fate and patterns of macromolecular expression (Davidson, 1989). At present the two of these territories for which most molecular data exist are those giving rise to the skeletogenic mesenchyme, and to the aboral ectoderm of the embryo. Founder cells for the skeletogenic mesenchyme are apparently specified autonomously, i.e. by the action of factors present in the cytoplasm of the polar region of the egg, which they inherit, while founder cells for the aboral ectoderm and other territories are apparently specified conditionally, i.e. at least in part by interblastomere interaction, following an initial polarization of the egg cytoplasm in the future oral–aboral axis (Hörstadius, 1939; Davidson, 1986, 1989). To approach the mechanisms by which the diverse genetic programs of the polyclonal territories of the embryo are set up, we have undertaken to characterize the regulatory factors that direct expression of marker genes activated specifically in these territories.

This report concerns a factor called P3A2, which appears to participate in regulation of marker genes that are expressed in different territorial domains of the early embryo. P3A2 target sites appear in the regulatory domain of the *CyIIIa* cytoskeletal actin gene, productive expression of which is confined to embryonic and larval aboral ectoderm (Cox *et al.* 1986; Cameron *et al.* 1989; Davidson, 1989). The *CyIIIa cis*-regulatory system includes sites for eleven other DNA-binding factors as well (Calzone *et al.* 1988; Thézé *et al.*

1990). Competitive interference *in vivo* with DNA–protein interactions at P3A2 sites and at the target site of another of the *CyIIIa* factors, P7II, causes ectopic expression of a *CyIIIa·CAT* reporter construct (Hough-Evans *et al.* 1990). P3A2 and P7II are thus thought to exert negative control on the *CyIIIa* gene in cells of other territories, *viz* gut, oral ectoderm and mesenchyme. Thiebaud *et al.* (1990) showed that both factors also recognize with high affinity sites present in the upstream region of the *SpecI* gene, another gene expressed exclusively in aboral ectoderm (Lynn *et al.* 1983; Hardin *et al.* 1985; Gan *et al.* 1990). However, the P3A2 factor (though not the P7II factor) also binds very tightly to sites located in the control region of a gene called *SM50* (Thiebaud *et al.* 1990). This gene codes for a matrix protein that is expressed exclusively in the lineages of the skeletogenic territory (Benson *et al.* 1987; Sucov *et al.* 1987, 1988). These lineages arise from entirely different founder cells than those specified as aboral ectoderm precursors (Davidson, 1989). The P3A2 factor is encoded by a rare maternal message (Cutting *et al.* 1990), and the active factor can be extracted from mid-cleavage nuclei, i.e. it is present during the initial processes of territorial founder cell specification. P3A2 may serve as an element in several different early 'specification switches', in which the sense of the specification is defined by the combination of factors engaged.

A natural advantage of sea urchin embryos for the molecular analysis of territorial gene regulation is the enormous amount of biological material available. We describe herein relatively simple procedures by which several billions of synchronous embryos were harvested, the nuclei separated, and active nuclear extract prepared. Though present at about a thousand molecules per nucleus at the stage from which the extract was obtained (Calzone *et al.* 1988), the P3A2 factor could be purified by site-specific affinity chromatography, and cloned. This interesting factor turns out to be unrelated in sequence to any known DNA-binding proteins. A feature of likely biological interest revealed by measurements carried out with affinity-purified P3A2 is that this protein discriminates sharply amongst the different marker genes, in respect to the stability of its interactions with the various target sites. It follows that different concentrations of P3A2 *in vivo*, in space or time, would be predicted to result in very different patterns of P3A2 regulatory interactions.

Materials and methods

Large-scale embryo cultures

Standard methods for culture of *S. purpuratus* sea urchin embryos are well known. Here we address the special problems associated with culture and harvest of unusually large numbers of embryos. Eggs from 200–400 gravid *S. purpuratus* females were collected over beakers of ice-cold, filtered sea water after intercoelomic injection with 0.5 M KCl. The settled eggs were combined and diluted to a 10–15 % suspension in fresh, ice-cold filtered sea water, and allowed to settle at 4 °C in 1 l beakers. One billion settled *S. purpuratus* eggs occupied a volume of about 1 l. The washing procedure was repeated 4–5 times. Fertilization efficiency and viability were reduced if the eggs were not washed sufficiently or remained packed too long in larger pellets. The washed eggs were stirred gently in a 10–15 % suspension, while counting an aliquot and making preparations for fertilization. The eggs were transferred at a concentration of 20 000–30 000 per ml at 16 °C to 4, plastic 20-gallon containers (Rubbermaid Garbage Cans) in a total volume of about 80 l. The eggs were then fertilized by addition of dry sperm freshly diluted into sea water. The cultures were vigorously aerated and stirred at 60 revs min^{-1} with a large paddle. Embryos cultured beyond the late gastrula stage should be diluted to about 5000–10 000 per ml. The embryos were harvested after hatching at the mesenchyme blastula stage (21–24 h), in portions of about 1.25×10^9. About 60 l of embryo suspension were concentrated by filtration with a 51 micron Nitex filter to about 4 l, transferred to 6 flat-bottomed 1 l bottles, and pelleted by centrifugation at about 1200 ***g***. The embryo pellets in each tube were resuspended in about 900 ml of ice-cold 1 M glucose, and collected by centrifugation at 1800 ***g***. The washed embryo pellets were then resuspended in about 10× the pellet volume with Buffer A [10 mM Tris (pH 7.4), 1 mM EDTA, 1 mM EGTA, 1 mM Spermidine–Tris–HCl, 1 mM dithiothreitol (DTT), 0.36 M sucrose]; and frozen in 1 l aliquots in heavy duty freezer bags (Ziplock) in liquid nitrogen. Frozen embryos were stored at −70 °C.

Preparation of nuclear protein extracts

Nuclear protein extracts were prepared as described previously (Calzone *et al.* 1988) with several modifications. The frozen embryos were crushed and uniformly thawed in batches of about 2.5×10^9. After mixing with a high speed overhead stirrer, the nuclei were collected by centrifugation in 1 l flat-bottomed bottles (Nalgene), at about 2500 ***g*** for 40 min at 4 °C. The pelleted nuclei were washed by resuspension in about 2–4 l of ice-cold Buffer A and recentrifuged. If nuclei trapped in fertilization envelopes were observed in the supernatant, they were released by mixing at moderate speed two times for 30 s in a Waring blender. The Buffer A washes were repeated two times, and were followed by 2–3 washes with Buffer A containing 0.1 % NP-40. By the end of these washes the nuclei were contained in a single 1 l bottle. According to Bradford (1976) measurements, the nuclear pellets include about 2.5 % of total starting embryo protein. The nuclei were resuspended in 3.0–3.7× the pellet volume with Buffer D [10 mM Hepes (pH 7.9), 1 mM EDTA, 1 mM EGTA, 1 mM Spermidine–Tris–HCl, 1 mM DTT, and 10 % glycerol]. The nuclei were transferred to 35 ml Oak Ridge tubes, and proteins were released from chromatin by addition of one-tenth vol of 4 M ammonium sulfate (pH 7.9), with vigorous mixing, followed by incubation on ice for 1 h, with occasional gentle mixing, and centrifugation at 35 000 revs min^{-1} in Beckman 50 Ti.2, 60 Ti or 70 Ti rotors for 3–6 h at 4 °C (Parker and Topol, 1984). Additional supernatant was recovered by recentrifugation of the chromatin pellets for 6 h. Roughly 60–70 % of the total protein in the nuclear pellet typically remained associated with chromatin. Approximately 50 % of the solubilized proteins were precipitated from the supernatant by addition of 0.25 gm of ammonium sulfate per ml and incubation overnight on ice. These precipitated proteins were collected by centrifugation at 10 000 ***g***, resuspended in about 20 ml of Buffer C [20 mM Hepes buffer (pH 7.9); this pH value refers to a 10× stock solution of Buffer C salts, 0.1 mM EDTA, 40 mM KCl, 1 mM DTT, 0.1 % NP-40, 20 % glycerol]. The proteins were dialyzed overnight at 4 °C against Buffer C. About 40 to 70 % of the protein in the

fraction precipitates at this step, and was removed by centrifugation at 35 000 revs min^{-1} in a 60 Ti rotor for 15 min. The samples were stored at −70 °C. The final yield of protein was about 10 % of the starting protein in the nuclei.

Site-specific DNA affinity chromatography

Site-specific DNA affinity columns were prepared essentially as described by Kadonaga *et al.* (1987). The oligonucleotides used to prepare the two synthetic binding sites used for purification of P3A2 are as follows. For the *CyIIIa* gene, complementary oligonucleotides 11/12 were: (11), GAAGCGCAAACAAACTTTATTAAGC; (12), CTTCGCTTAATAAAGTTTGTTTGCG; for the *SM50* gene, oligonucleotides 25/26 were: (25), GCTTCTGCGCACACCCCACGCGCATGGGGCGT; (26), AAGCACGCCCCATGCGCGTGGGGTGTGCGCAG. Columns were stored at 4 °C. To purify P3A2, the nuclear protein extracts were applied to site-specific DNA affinity columns without prior purification. For each affinity-purification, a quantity of nuclear extract estimated to contain enough factor to saturate approximately 30 % of the total specific sites coupled to the column resin was applied to each column. Application of less extract reduced the ratio of P3A2 to nonspecific proteins in the bound fractions. A 0.2 ml *SM50* column was used for the pilot experiment described in text. A 1.0 ml *CyIIIa* and a 1.5 ml *SM50* columns were used for the large scale purification. The volume of nuclear extracts applied to affinity columns was 3.3 ml for the pilot experiments, and 52 ml for the large scale preparation. When loading large volumes, the extracts were passed through a 0.5 ml Sepharose CL-4B pre-column that did not contain DNA to prevent accumulation of any insoluble debris in the site-specific DNA affinity column. Affinity chromatography was carried out at 4 °C. Before loading, the concentration of KCl in the nuclear extracts was increased to 0.1 M. All loading, washing and elution buffers contained 20 mM Hepes (pH 7.9), 0.1 mM EDTA, 1 mM DTT, 0.1 % NP-40, 20 % glycerol, and the concentrations of KCl indicated here and in Results. The loading rate was approximately 0.1–0.2 column volumes per min. After loading the column was washed with 30–50 ml of column buffer containing 0.1 M KCl at a rate of 1 ml per min. Bound proteins were eluted with a KCl step gradient (3 column vol per step; an increase of 0.1 M KCl per step) at a rate of about 0.5 column vol per min. For the second cycle of affinity chromatography, the appropriate fractions were pooled (see Results), dialyzed against 0.1 M KCl in column buffer, loaded and eluted as described above.

DNA-binding assays

DNAase I footprinting, methylation interference and gel-retardation assays were performed according to standard procedures, which have been described previously (Calzone *et al.* 1988; Thézé *et al.* 1990). The amount of protein, probe, nonspecific DNA, and DNAase I had to be optimized for each experiment. The *CyIIIa* gene probe in the purification and characterization of purified P3A2 was an 86-bp *Hin*dIII–*Hin*fI fragment of the *CyIIIa* gene regulatory region (−244 to −157), or a *Hin*dIII–*Dde*I fragment (−244 to −114); (see Thézé *et al.* 1990, for the sequence). The *SM50* gene probe for similar experiments was an *Hpa*II–*Bst*NI fragment of the regulatory region (−49 to −200) subcloned into Bluescript KS (Strategene; kindly provided by K. Whittaker) and released from vector by digestion with *Eco*RI and *Hin*dIII. A binding site for the CTF transcriptional activator is included in the *Hpa*II–*Bst*NI fragment. The following oligonucleotide gene probes were used to compare the equilibrium binding constants of P3A2 interactions with *SM50* and *Spe*I P3A sites. For the *SM50* gene, oligonucleotides 50/51: (*50*), GATCTTTTCGGCTTCTGCGCACACCCCACGCGCATGGGGC; (*51*), GATCGCCCCATGCGCGTGGGGTGTGCGCAGAAGCCGAAAA. For the *Spe*I gene, oligonucleotide 52/53: (*52*), GATCATCTGCGCATGCACAGATCAATCCGCGCATGCTCAG; (*53*), GATCCTGAGCATGCGCGGATTGATCTGTGCATGCGCGAT. The procedure for detection of P3A2 DNA-binding activity in nitrocellulose filter blots of one-dimensional SDS protein gels (Laemmli, 1970) was similar to the protocol described by Vinson *et al.* (1988) for screening cDNA expression libraries.

Gel purification and renaturation of P3A2

A gel slice containing approximately 1 μg of P3A2 in a SDS protein gel was located using parallel marker lanes, excised, mashed in 0.5 ml of elution buffer [50 mM Tris (pH 7.6), 0.1 mM EDTA, 0.1 % SDS, 5 mM DTT, 150 mM NaCl, 0.1 mM PMSF] and incubated at room temperature for 6 h. Gel pieces were removed by centrifugation and washed twice with 0.2 ml of water. The supernatants were combined and protein was precipitated by addition of 5 vol of acetone, followed by incubation overnight at −20 °C. The protein precipitate was collected by centrifugation, washed with fresh acetone–water (5:1), briefly dried and resuspended in 0.1 ml of denaturation buffer [20 mM Hepes (pH 7.9), 0.1 mM EDTA, 1 mM DTT, 6 M guanidinium–HCl, 1 mM $MgCl_2$]. The sample was passed over a 0.5 ml BioRad P6 column equilibrated with 10 mM Hepes (pH 7.9), 0.1 % NP-40, 1 mM DTT, 100 mM KCl, 10 % glycerol. The fractions containing DNA-binding activity were located by the gel retardation method. Compared to the starting material, the efficiency of recovery of specific DNA-binding activity after renaturation was less than 1 %.

Protein sequencing

Protein sequencing of affinity-purified P3A2 was accomplished as described by Aebersold *et al.* (1989). Approximately 15–30 μg of P3A2, purified by two cycles of chomatography on an *SM50*-P3A site-specific DNA affinity column, were separated from remaining protein contaminants by electrophoresis on a 7 % SDS gel (3–5 μg per lane); and electrophoretically transferred to nitrocellulose (S&S, 0.45 μm). Tryptic fragments of P3A2 separated by HPLC, and sequenced on polybrene-coated glass fiber supports using an Applied Biosystems 477A/120A pulsed liquid sequenator, essentially as recommended by the manufacturer. Phenylthiohydantoin-amino acid analysis was performed on-line and sequence assignments were made by visual inspection of the chromatograms. The protein sequences obtained are listed below. Numbers within parentheses specify the HPLC fractions from which the peptides came. One-letter codes are used for the amino acids. Lower case letters signify uncertain assignments. At position 10 in fraction 29.2, two signals were observed, neither of which could unambiguously be assigned to the major sequence. (*14*), ATIDEYATR; (*17.1*), SFETNPSIR; (*17.2*), KVSLA; (*19.1*), VGQQAVV; (*19.2*), LKATIYELVLKkgk; (*19.3*), SSVIND; (*20*), TIVINCYK; (*25*), QTVVAGDGQPIQIANVNIAQQSGxxgTMAAIK; (*29.1*), VFGAAPLENIMR; (*29.2*), SQVLPvFIe(N,R); (*30.1*), GIVLQDLDNASAQrk; (*31.1*), APQPSNENSDSYELPPLVIDGIHDxh; (*31.2*), NAVMQSQPIPLQVATLVVNAASPTQVNK.

Isolation and sequencing of P3A2 cDNA clones

The partial protein sequences of P3A2 were used to construct the oligonucleotide probes shown below (I indicates inosinic acid). (*P25*), CTGCTGIGCAATGTTIACATTIGCAATCTGGATIGGCTGTCCATC; (*P31.1*), GTCATGGATACCAT-

CAATGACIAGTGGTGGIAGCTCATAIGAGTCIGAGTTCTCATT; (*P31.2*), GTGGCCACCTGIAGIGGGATIGGCTGIGACTGCATGACAGCATT; (*P30.1*), CTGGGCIGAIGCATTATCIAGATCCTGIAGGACAATTCC; (*P14*), TGIGC(G,A)TA(T,C)TC(G,A)TCIAT. A set of duplicate filters of a λ ZAP (Stratagene) cDNA library of 14 h embryo poly(A) RNA was reacted with a hybridization mixture containing 0.5 picomole of each oligonucleotide, labeled with ^{32}P at the 5′ terminus, in 6×SET [1×SET is 0.15 M NaCl, 10 mM Tris (pH 8.0), 1 mM EDTA], 50 mM sodium-phosphate (pH 7.4), 5×Denhardt's solution (Denhardt, 1966), 0.1 mg ml^{-1} salmon sperm DNA at 37°C for 18 h. The filters were washed twice at room temperature for 10 min each in 2×SSC (1×SSC is 0.15 M NaCl, 15 mM sodium citrate), 0.2 % SDS twice at room temperature in 1×SSC, 0.2 % SDS, twice in 3 M tetramethylammonium chloride (TMA), 50 mM Tris (pH 8.0), 2 mM EDTA, 0.1 % SDS at 37°C for 15 min, and once in the same TMA buffer at 48°C for 15 min. Seven positive clones were detected in a total of 6×10^5 phage. The two longest clones, p21 and p25, were selected for further analysis. cDNA inserts were sequenced using standard procedures for the Bluescript system (Strategene).

Preparation of recombinant P3A2

A T7 RNA polymerase based expression system in *E. coli* developed by Studier and Moffat (1986) was used to prepare recombinant P3A2. The plasmid construct used to express P3A2 was an *Eco*RV–*Dra*I cDNA fragment, joined using *Bam*HI linker into the *Bam*HI site of pET3c. *Eco*RV cuts in codon 6 of the P3A2 message sequence. The *Dra*I site is located in the 3′ noncoding trailer of the message. The host for P3A2 expression was *E. coli* strain (lys S). To prepare P3A2, a 0.5 l culture of cells containing the expression plasmid was grown to an OD_{600} of 0.7 in NZ medium containing 50 μg ml^{-1} ampicillin and 20 μg ml^{-1} chloramphenicol NZ [10 g l^{-1} NZ amine (Sheffield), and 5 g l^{-1} NaCl]. The T7 RNA polymerase promoter was induced by addition of 0.5 mM IPTG and incubation of the culture was continued for 4 h. The cells were collected by centrifugation, resuspended in 50 ml of fresh NZ medium, pelleted again by centrifugation, and resuspended in 10 ml of 50 mM Tris (pH 7.6), 50 mM NaCl, 1 mM EDTA, frozen in liquid nitrogen, and stored at −70°C. After thawing, addition of 2 mg ml^{-1} lysozyme, incubation on ice for 2 h, and addition of PMSF to 1 mM, lysis was completed by sonication (4 to 5 high intensity bursts). NP-40, sucrose and DTT were then added to concentrations of 0.5 %, 5 % and 1 mM, respectively. Insoluble material was removed by centrifugation at 10 000 *g*. Greater than 90 % of the recombinant P3A2 was recovered in the soluble supernatant. After addition of one-tenth volume of 4 M ammonium sulfate and glycerol to a final concentration of 20 %, ribosomes and other particles were removed from the supernatant by centrifugation at 100 000 *g*. Recombinant P3A2 constituted at least 1 % of the total protein in the postribosomal supernatant. The proteins were diluted 100- to 1000-fold for DNA-binding protein assays; dialysis to remove ammonium sulfate was not necessary.

Antibody blotting

All procedures were carried out at room temperature. Nitrocellulose blots were incubated for 3 h in 'Blotto' to block nonspecific binding sites (Blotto is 5 % nonfat milk, 0.01 % AntifoamA, 0.0001 % merthiolate in PBS, pH 7.5). Whole rabbit serum containing P3A2 antibodies was diluted 1/400 in 15 ml Blotto and incubated with the blot for 1 h. The primary antiserum was removed and the blot was washed with four changes of Blotto (15 min each). After a 10 min wash in PBS, the blot was incubated for 10 min in 0.1 % glutaraldehyde in PBS followed by two additional washes in PBS (Ikegaki and Kennet, 1989). Secondary antibody, goat anti-rabbit conjugated to horseradish peroxidase (EY Labs), was diluted 1/1000 in 15 ml of Blotto and subsequently incubated for 1 h. The antibody was removed with four 15 min washes of Blotto and the blot was stained with a DAB-metal substrate for 15 min (0.1 % 3,3′ diaminobenzidine, 0.03 % cobalt chloride, 0.05 % H_2O_2 in PBS, pH 7.5). The stained blot was photographed with Kodak Plus-X film.

Results

Purification of P3A site binding factors

A sea urchin blastula contains about 5×10^5 molecules of a factor that produces a footprint over the P3A site in the *CyIIIa* and *SM50* genes (Calzone *et al.* 1988; and data shown below in Table 1). The factor was purified by directly applying crude nuclear protein extract from about 2.5×10^9 embryos to site-specific DNA-affinity columns on which were mounted the *CyIIIa* or the *SM50* P3A target binding sites. The specific sites were those located at positions −169 to −193 in the regulatory domain of the *CyIIIa* gene (the sequence is presented by Thézé *et al.* 1990); and the double P3A site at position −98 to −129 in the *SM50* gene (Sucov *et al.* 1988; Thiebaud *et al.* 1990). Details including column oligonucleotide sequences and operating protocols are given in Materials and methods. The elution of specific DNA-binding proteins from each column was monitored by gel-retardation assays and by SDS protein gel electrophoresis. Results from the first rounds of affinity-purification are shown in Fig. 1. P3A activity can be seen to elute from both columns with a protein of approximately 62×10^3 M_r that was easily detected in the silver-stain pattern of the SDS gels. Note that a higher concentration of salt (≥0.7 M KCl) was required for elution of the activity from the *SM50* column than from the *CyIIIa* column (0.3–0.4 M KCl). A consequence was that when purified on the *SM50* column the P3A activity was relatively free of nonspecific proteins (see Table 1). As shown below, a large preference in the specific binding to the *SM50* P3A site over the *CyIIIa* site underlies the differential elution of P3A activity from these two columns. DNAase I footprint and methylation interference patterns obtained after reaction of the proteins purified on the *SM50* column are shown in Fig. 2. These patterns are identical to those obtained with crude extracts, using the same double P3A target site of the *SM50* gene (data not shown).

The protein preparation obtained after two cycles of affinity chromatography with the *SM50* column (Fig. 1B) contained two major species, having apparent relative molecular masses of 62 and 110×10^3, respectively, though as can be seen in Fig. 1A the 110×10^3 species is also present in column fractions that display no P3A activity. In Fig. 3 we show experiments demonstrating that the specific DNA-binding activity recovered from the *SM50* column is entirely due to the 62×10^3 protein, hereafter called P3A2. Thus the 62×10^3 protein reacted with a probe containing

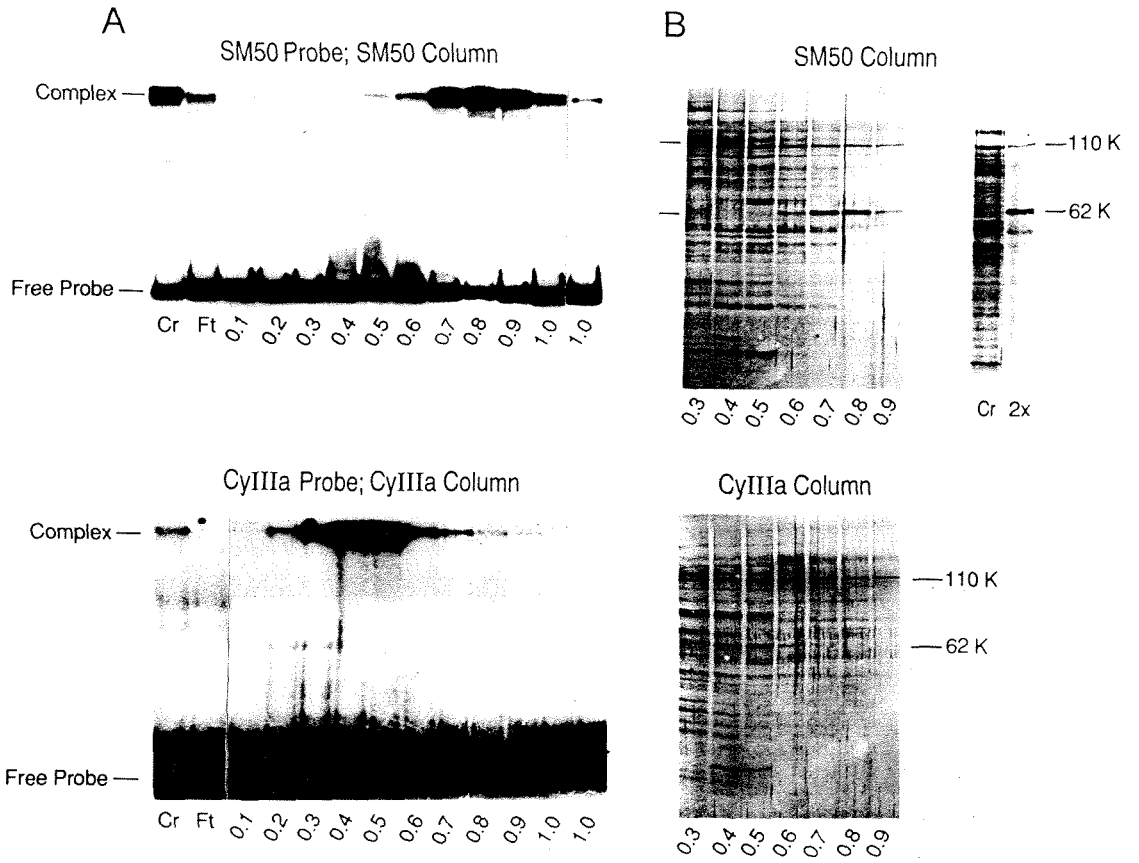


Fig. 1. Affinity purification of P3A2. Blastula nuclear extracts representing about 2.5×10^9 embryos (10^{12} nuclei) were passed sequentially over a *CyIIIa* and a *SM50* site-specific column (see Materials and methods for details). After washing extensively the bound proteins were eluted with a step gradient of KCl. Approximately 7×10^{-9} moles of specific factor were applied to the affinity columns. The *CyIIIa* column bound about 50% of the P3A2 in the extract. The majority (90%) of the remaining factor was bound by the *SM50* column (it is likely that the capacity of the *CyIIIa* column for P3A2 was exceeded by a factor of about two). The elution of bound proteins was monitored by gel retardation assay or SDS gel electrophoresis as shown; the concentration of KCl in the bound fractions assayed is indicated below each lane. (A) Gel retardation assays. The probe used to monitor the elution of specific proteins from the *SM50* column was the *Hpa*II–*Bst*NI fragment of the *SM50* regulatory DNA shown in Fig. 2. Protein binding to the *CyIIIa* column was assayed with the *Hin*dIII–*Dde*I fragment of the *CyIIIa* gene shown in Fig. 5. The binding reactions (10 μl) contained 8.3×10^{-11} M of the *SM50* probe or 6.8×10^{-11} M of the *CyIIIa* probe. Lanes labeled 'Cr' contain crude, or unfractionated nuclear extract, and lanes marked 'Ft' contain samples of the flow-through fractions from the affinity columns. For Cr and Ft reactions 8.8 μg of poly (dI/dC) (*SM50* probe) or 5.6 μg of poly (dAT) (*CyIIIa* probe) were added. Appoximately 10^{-5} of the total protein in the Cr and Ft fractions was used for a given assay. The amount of poly (dI/dC) or poly (dAT) in reactions carried out with higher KCl fractions eluted from the affinity columns was reduced by 70%. About 10^{-4} of the total protein recovered in each KCl fraction was used per reaction. (B) SDS protein gels. Approximately 3 μl of each column fraction indicated was analyzed by electrophoresis on a 10% SDS–polyacrylamide gel. Proteins were detected by silver stain. The proteins present in the preparation of P3A2 obtained after two cycles of affinity-purification with an *SM50* column are shown in the *SM50* panel, lane 9, marked '2x'.

concatenated P3A binding sites (oligonucleotides 25/26) in protein gel blots, but not with a probe containing *CyIIIa* binding sites for a factor we call P1 (Fig. 3A). The same negative results (not shown) were obtained with probes representing binding sites for other factors, P6 and P7I. However, the 110×10^3 protein reacted equally well with all of these unrelated binding sites, and thus is to be considered a nonspecific DNA-binding protein. The 62×10^3 protein was gel purified, renatured, and shown to form a specific complex with a probe containing the *SM50* P3A site by the gel shift method (Fig. 3C), and by DNAase I footprinting (Fig. 3B). As Fig. 3C shows, the 110×10^3 protein did not reveal any specific DNA-binding activity.

Table 1. *Pilot scale purification of P3A2 from blastula nuclear extracts*

(1) Fraction	(2) Factor[a] (moles)	(3) Total protein[b] (g)	(4) Sp. Act.[c] (moles/g protein)	(5) Recovery[d] (%)	(6) Purification[e] (fold)
Nuclear extract	3.0×10^{-10}	9.2×10^{-2}	3.3×10^{-9}	100	1
First *SM50* site column[f] (0.7–1.0 M KCl)	2.3×10^{-10}	2×10^{-3}	1.1×10^{-7}	76	33
Second *SM50* site column (0.5–0.9 M KCl)	1.6×10^{-10}	3×10^{-5}	5.3×10^{-6}	54	1600

[a] The amount of P3A2 factor in blastula nuclear extracts was determined as described previously (Calzone *et al.* 1988; see Fig. 7 of this paper) using the *SM50* probe shown in Fig. 2, oligonucleotide 25/26 concatenate as a specific competitor, and poly dAT as a nonspecific competitor. The oligonucleotide 25/26 competitor contains the tandem P3A sites in the *SM50* gene used to prepare the *SM50* affinity column (see Materials and methods for details and Fig. 2). The extract used in the purification represented 1.6×10^{8} embryos (6.3×10^{10} nuclei).

[b] The amount of protein in each fraction was determined by the method of Bradford (1976).

[c] The specific activity of P3A2 at each step was calculated as [moles of factor (column(2)/total protein (3))].

[d] Recovery was calculated as [moles of factor (2)/3.3×10^{-10}].

[e] Purification was calculated as [specific activity (4)/3.3×10^{-9}].

The information obtained in the pilot purification summarized in Table 1 allowed us to estimate the relative concentration of P3A2 in blastula nuclear extracts in two different ways. A conservative measure of the purity of the P3A2 preparation, after two cycles of affinity chromatography with an *SM50* column, based on the silver-stained SDS gels, indicated that P3A2 constitutes about 30 % of the total protein (30 μg) in the bound fraction. The final efficiency of recovery of P3A2 was about 54 %, and thus the total amount of P3A2 in the crude extract was approximately 17 μg ((0.30×30 μg)/0.54). Second, according to a gel retardation measurement, the quantity of nuclear extract used in the pilot purification of Table1 contained about 300 picomoles of the factor that interacts with the P3A probes. Assuming the protein binds specific DNA as a monomer and using 49×10^{3} as the relative molecular mass for the factor (established by sequencing the cloned P3A2 mRNA; see below), we calculated that the crude extracts of blastula nuclei contained about 24 μg of P3A2. The good agreement of these two different estimates of P3A2 yields achieved in the pilot purification should not be taken as proof that P3A2 binds DNA in monomeric form. Thus differential silver-staining of the nonspecific proteins in the affinity-purified preparation could easily have led us to underestimate the purity of P3A2 by a factor of two. The mass ratio of P3A2 to total blastula nuclear extracts was estimated to be about 2×10^{-4} (20 μg/92 mg; see Table 1). Of the total enrichment required for purification of P3A2 from the crude nuclear pellet, a factor of about 10 was achieved by elimination of 90 % of the total protein in the nuclear pellet by the various steps in the preparation of the nuclear extracts (see Materials and methods), and the remainder was accomplished by affinity chromatography. The results of the P3A2 purification suggest that the isolation of other DNA-binding factors of similar prevalence should be relatively straightforward by the same procedures. Aebersold *et al.* (1988) have shown that for the purpose of protein sequencing a purity of about 5 % is generally sufficient prior to protein purification on SDS protein gels. Thus, a relatively modest 250-fold purification would have been adequate for P3A2. The concentrations of the other factors in blastula stage nuclear extracts that have been shown to bind specifically to the *CyIIIa* regulatory DNA (Calzone *et al.* 1988) range from about 0.2 to 4 times the concentration of P3A2 (1×10^{5} to 2×10^{6} molecules/400 cell embryo). Thus the minimum purification required for each of these factors would range from 62- to 1250-fold. These are levels of purification easily achieved by direct site-specific DNA affinity chromatography of the crude nuclear extracts.

Cloning P3A2 mRNA

Approximately 0.5 nanomole of the 62×10^{3} P3A2 protein obtained with the *SM50* column was further purified by SDS–polyacrylamide gel electrophoresis (see Fig. 1), transferred to a nitrocellulose filter, and excised and digested *in situ* with trypsin. Peptides released from the filter were separated by HPLC and protein sequences obtained; these are shown in Materials and methods. These sequences were used to construct five oligonucleotide probes, also given in Materials and methods, with which clones encoding P3A2 were recovered from a cDNA library representing 14 h embryo poly(A) RNA. Seven cDNA clones were isolated from a total of about 6×10^{5} recombinant phage. The longest insert (clone p21) was 3664 nt, close to the full length of the P3A2 message according to RNA gel blot measurements (Cutting *et al.* 1990). Alignment of the tryptic peptide sequences of the purified P3A2 with the predicted protein encoded in the p21 cDNA insert revealed the peptide sequences, within an open reading frame beginning with an ATG at position 79, and ending at a stop codon at position 1446. The complete sequence is shown in Fig. 4, where the location of each tryptic fragment sequence in indicated. The derived protein is 49×10^{3}, i.e. significantly smaller than the estimate of 62×10^{3} provided by protein gel electrophoresis (Fig. 1). To confirm that clone p21 indeed encodes P3A2, a recombinant protein contain-

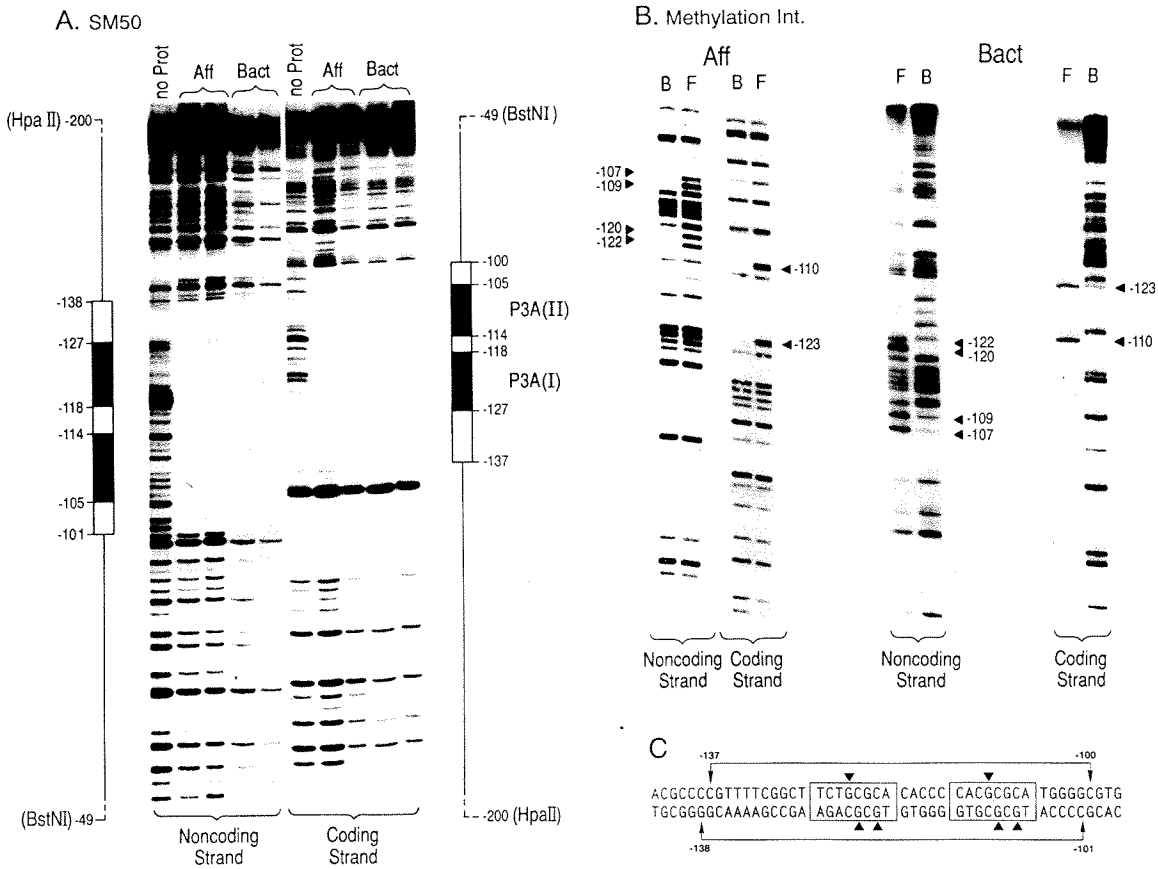


Fig. 2. DNAase I footprint and methylation interference reactions of P3A2 with *SM50* probes. (A) DNAase I footprints. Probes labeled by the kinase reaction at the *Bst*NI (noncoding) or *Hpa*II (coding) terminus were reacted with (left to right) 0 μl (no Prot), 1.0, 2.0 μl (Aff) of P3A2 purified by two cycles of affinity chromatography (first preparation, see Table 1) or 1.0, 2.0 μl of bacterial P3A2 extract and treated with 60 μg ml^{-1} of DNAase I for 1 min on ice. The binding reactions (12.5 μl) contained 5 μg of poly (dAT) and about 0.3 ng of specific probe. The preparation of the recombinant P3A2 is described in Materials and methods. Marker lanes are not shown. The regions of the probes protected from nuclease attack by the P3A2 proteins are represented by the boxes. The filled boxes within the protected sequences map the position of P3A core target site sequences. (B) Methylation interference assays. Noncoding and coding probes were separately treated with DMS and reacted with affinity-purified (Aff) or bacterial (Bact) P3A2. Specific complexes and free probe were separated by gel electrophoresis, eluted and cleaved at residues with piperidine. The cleavage products for each probe derived from specific complex are in lane B, free probe products are in lane F. The solid triangles identify the sites of strong methylation interference in the binding of P3A2 with each strand-specific probe. The probe for the reactions with affinity-purified protein was labeled by the end-fill reaction at the *Hpa*II end (noncoding) or *Bst*NI end (coding). Reactions with bacterial protein used the same probes shown in (A). (C) Sequence map of DNAase I footprint and methylation interference patterns. The sequence in the region of *SM50* regulatory DNA which was found to specifically interact with affinity-purified P3A2 is shown. The DNAase I footprints on each strand are represented by the brackets. The solid triangles locate positions of strong methylation interference. The boxes map the tandem P3A target site core sequences.

ing amino acid residues 7 to 459 (the C terminus), fused to a short length of vector and linker N-terminal residues, was expressed in *E.coli*, and assayed for DNA-binding activity. The DNAase I footprint and the methylation interference patterns obtained for the recombinant protein with an *SM50* probe were found to be identical to those observed for affinity-purified P3A2. These results are included in Fig. 2.

Characteristics of the P3A2 protein sequence

A search of current data bases revealed no significant similarity between P3A2 and any previously reported protein sequence. A low resolution N-terminal and C-terminal deletion analysis (see accompanying paper, Höög *et al.* 1991) has shown that the sequences essential for the DNA-binding activity of P3A2 span a broad region, the N terminus of which occurs between amino

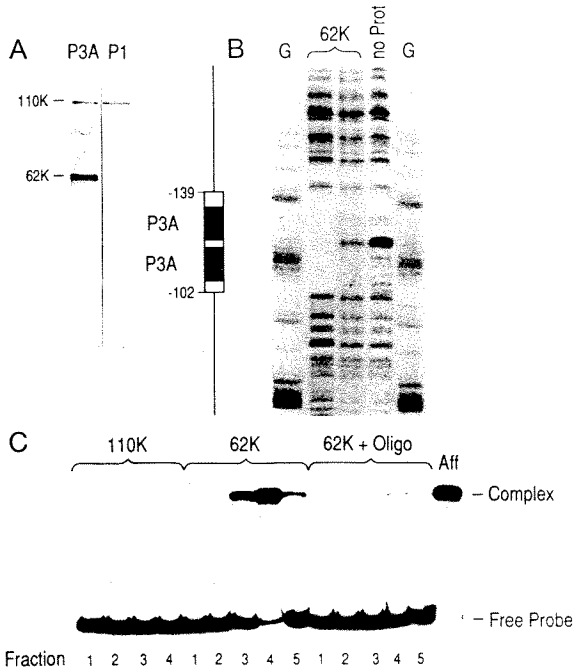


Fig. 3. DNA-binding reactions with gel eluted, affinity-purified proteins. (A) Protein gel blot reactions. Proteins purified by two cycles with an *SM50* column (second preparation, Table 1) were separated by electrophoresis in an 8% SDS–polyacrylamide gel, transferred to nitrocellulose and reacted separately with labeled P3A or nonspecific P1 binding sites (see Materials and methods for details). The P3A probe was a concatenate of oligonucleotides 25/26; The P1 probe was the concatenated binding site 37/38 (Thézé *et al.* 1990). Other oligonucleotide sites tested which generated a reaction pattern exactly similar to P1 (not shown here) were P6 and P7I (oligonucleotides 5/6 and 19/20 of Thézé *et al.* 1990). (B) DNAase I footprints. The 62×10^3 protein was gel-purified and renatured as described in Materials and methods. The renatured protein (80 μl and 40 μl, left to right) was reacted with a coding-strand probe including the whole *SM50* regulatory region and treated with DNAase I. G indicates a G ladder sequence reaction. (C) Gel retardation assays. Gel-purified 62×10^3 and 110×10^3 proteins were assayed for DNA-binding activity by the gel retardation method using an *SM50* probe containing the tandem P3A target sites (see Fig. 2). The specificity of the complex detected with the 62×10^3 protein was demonstrated by competition with concatenated binding site 25/26 (lanes marked 62 k+oligonucleotide). The lane marked Aff shows the complex present in affinity-purified P3A2 before gel-purification. 'Fraction' refers to eluates from the Biogel P6 column used to remove guanidium HCl (see Materials and methods).

acid positions 25–90, and the C-terminal boundary between residues 222–358, i.e. a DNA-binding domain of at least 130 amino acids is necessary. Nor has the sequence of this DNA-binding domain *per se* yet revealed similarity to any previously characterized DNA-binding motif. P3A2 is thus a novel, sequence-specific recognition factor. Two regions of P3A2 are relatively rich in serine and threonine, and may include sites of phosphorylation detected in the affinity-purified protein (F.Calzone and M. Harrington, unpublished observations). Near the N terminus, the sequence spanning residues 4 to 28 is 36% serine plus threonine (S+T); residues 266–290 are 28% (S+T) and residues 364–389 are 26.9% (S+T). For comparison, the remaining sequences in P3A2 average 11.7% (S+T). P3A2 also has a very glutamine-rich region, similar to those that have been detected in several other DNA-binding proteins, e.g. Oct-1 and Zeste (Pirrotta *et al.* 1987; Sturm *et al.* 1988). Thus the P3A2 sequence between residues 308 and 336 is 37.9% glutamine, compared to 6.7% glutamine in the remainder of the protein.

P3A2 interacts with CyIIIa *regulatory DNA*

Several lines of evidence indicate that P3A2 recognizes regulatory sites in the *CyIIIa* gene, as well as in the *SM50* gene. The *CyIIIa* site reproduced in oligonucleotides 11/12 (sequence given in Materials and methods) and mounted on the *CyIIIa* affinity column is also protected from DNAase I when reacted with P3A2 purified with the *SM50* column. This footprint is shown in Fig. 5A, and Fig. 5B demonstrates that the same essential G residues that are contacted by P3A2 in the *SM50* P3A target site are contacted in the *CyIIIa* target site used for these experiments, although additional G contacts were detected outside of the core sequence. These data are summarized in Fig. 5C, which should be compared with Fig. 2C. In addition, P3A2 factor purified using the *CyIIIa* affinity column produced the same DNAase I footprints and methylation interference patterns noted for P3A2 purified on the *SM50* column (not shown). Finally, as shown in Fig. 6, an antibody raised against *recombinant* P3A2 detected protein of apparent mass 62×10^3 in the specific protein fractions eluted from both the *CyIIIa* and *SM50* columns. Taken together with results of the gene transfer studies summarized in Introduction, these observations suggest that P3A2 participates in regulation of mutually exclusive patterns of gene transcription, in developmentally unrelated cell types, as exemplified by the territory-specific *CyIIIa* and *SM50* genes.

The organization of P3A2 target sites in the CyIIIa *and* SM50 *genes*

DNAase I footprint and gel retardation assays performed with affinity-purified P3A2 revealed multiple binding sites for the protein in the extensive regulatory domains of both the *CyIIIa* and *SM50* genes (data not shown). The coordinates of each site are listed in Table 2. The *CyIIIa* regulatory region contains a total of three P3A2 binding sites. The binding site at position −182 to −195 includes the consensus P3A target sequence identified by Thiebaud *et al.* (1990), i.e. $^{C}/_{T}X^{C}/_{T}GCGC^{A}/_{T}$, while those at positions −807 to −820, and −101 to −114 differ respectively in two and

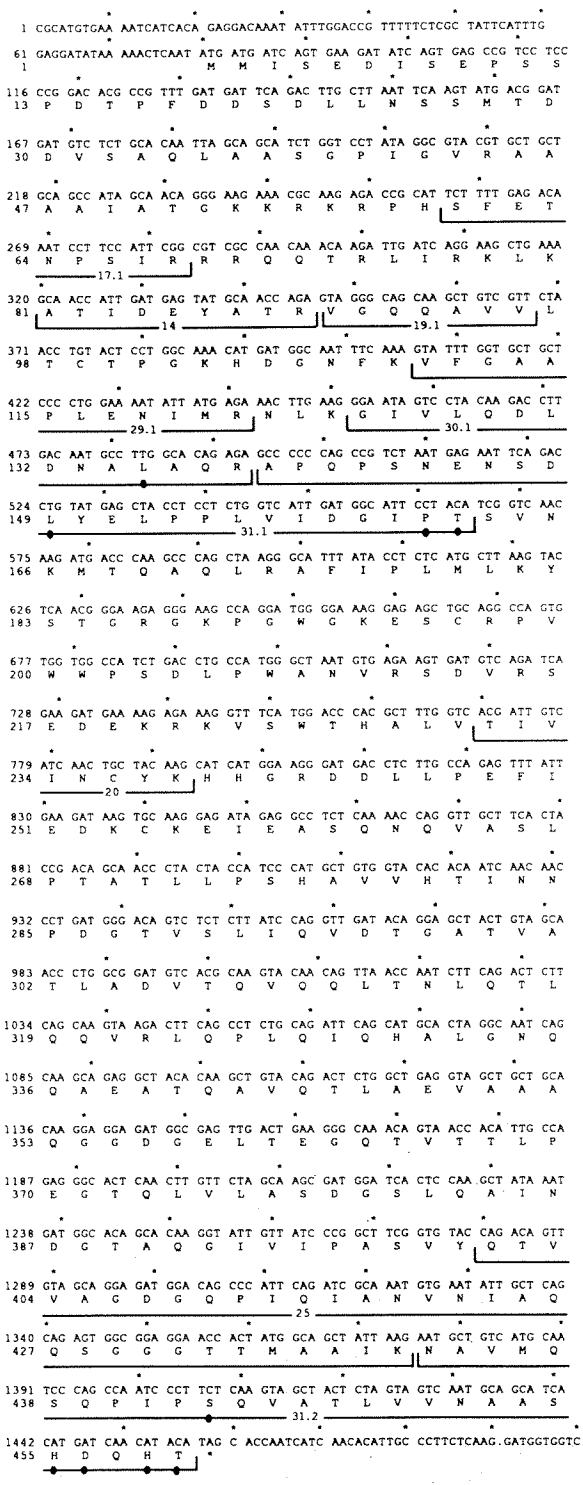
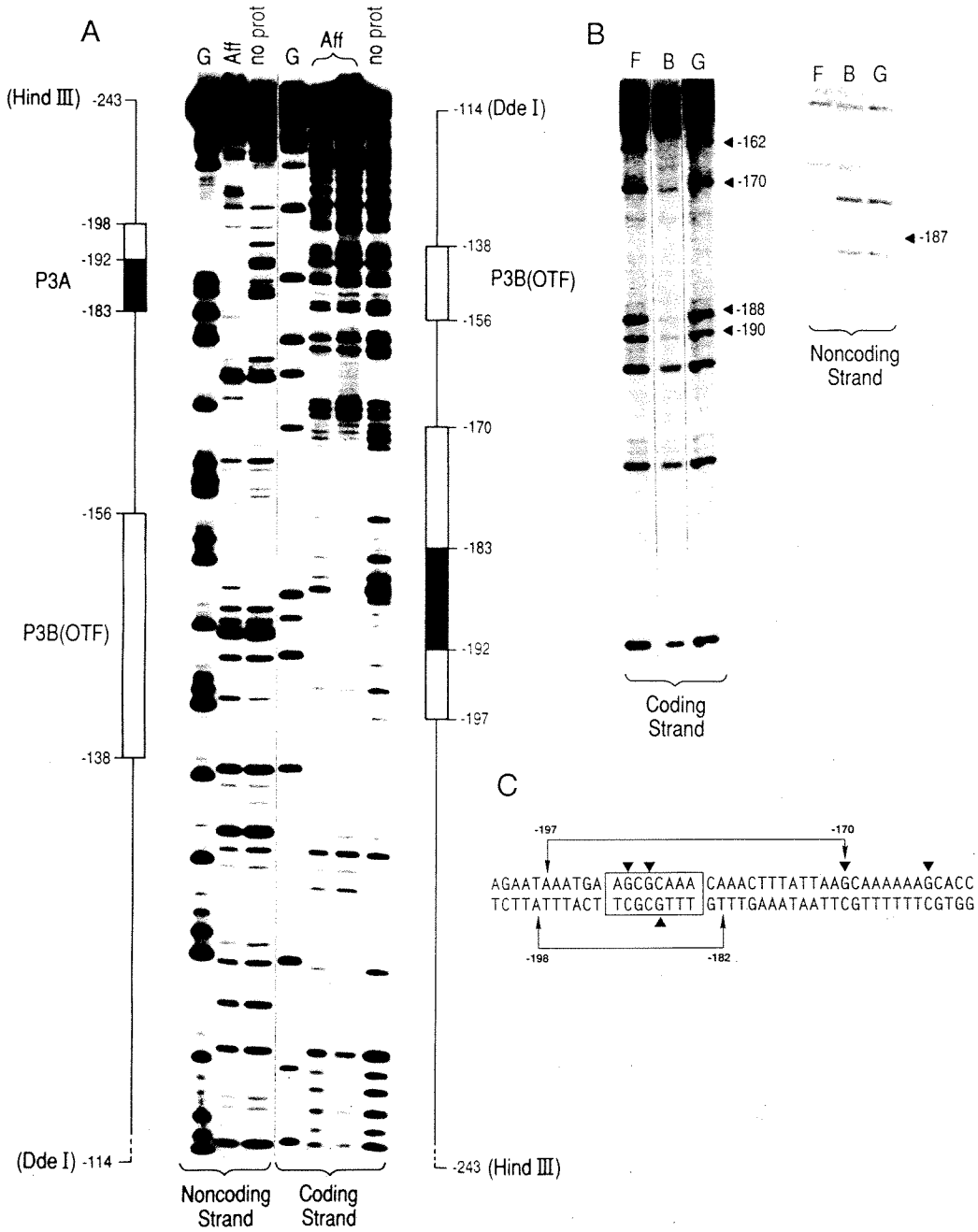


Fig. 4. Sequence of P3A2 mRNA, and of encoded protein. The sequence may be missing up to about 50 nt from the 5′ leader (see text). Brackets indicate locations of predicted peptides obtained by direct sequencing of tryptic peptides from affinity-purified P3A2, as described. All peptides shown in Materials and methods are identified and are indicated, except for several that are redundant, i.e. included within the sequences bracketed. Agreement between the sequenced peptides and the translated sequence of the cDNA was perfect, except where marked with dots. The misidentifications occasionally occurred as confusions between S and L, and at the C-terminal ends of the peptides where the sequence is more difficult to interpret. The sequence has been entered in the Genbank/EMBL database, accession no. X57014.



three nucleotides out of the seven constrained positions (marked by dots in Table 2). Correspondingly, the interaction of P3A2 with the site at −101 to −114 is relatively weak (Table 2), and can only be observed after occupation of the strong site at −182 to −195, and only if both sites are present on the probe fragment. Five P3A sites occur in the regulatory region of the *SM50* gene. The two closely spaced sites at positions −102 to −128 (Table 2) are those that were utilized for the purification of P3A2 described above. As shown in Fig. 2, the pattern of essential G contacts in these sites indicates that P3A2 binds similarly to both of them. Another strong P3A site occurs in the *SM50* regulatory region close to the transcription start, at position −23 to

Fig. 5. P3A2 interactions with *CyIIIa* regulatory DNA. (A) DNAase I footprints. The *CyIIIa* regulatory DNA fragment indicated in the Figure was labeled by the polynucleotide kinase method, and reacted with P3A2 obtained from an *SM50* affinity column ('Aff' on Figure); or with no protein, as indicated. 'G' indicates marker G ladder sequencing reactions. For the noncoding strand reaction shown, 10 μl of the P3A2 preparation were used, and for the coding strand 8 μl and 2 μl were used. The reaction mixtures were treated with 80 μg ml^{-1} DNAase I for 40 s on ice. The reactions (12.5 μl) contained about 0.015 picomole of each probe. This *CyIIIa* probe also contains an octamer binding site indicated on the Figure ('P3B (OTF)'). (B) Methylation interference assays. The *CyIIIa* probes were treated with DMS and separately reacted with P3A2 purified on an *SM50* column. Free probe ('F') was separated from probe bound to P3A2 ('B') by polyacrylamide gel electrophoresis. Both probe fractions were eluted and cleaved with piperidine. (C) Sequence map of P3A2 target site at positions −198 to −170 in the *CyIIIa* gene. The sequence in the region of the probes covered by DNAase I footprints in A is shown. DNAase I footprints on each strand are indicated by boxes. The shaded region of each box covers the P3A core target site sequence. Solid triangles mark positions of strong methylation interference, from the experiments shown in B.

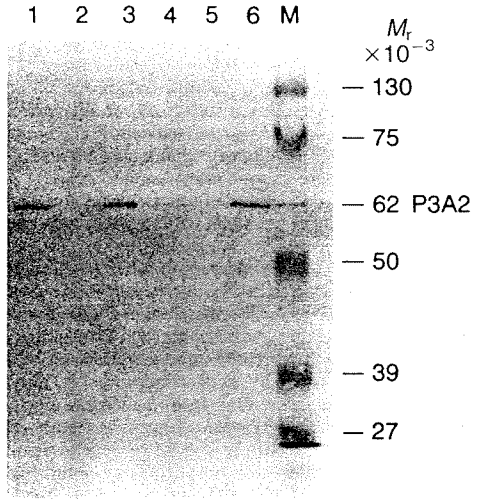


Fig. 6. Identification of P3A2 by reaction with rabbit antibody against recombinant P3A2 protein. Proteins were displayed by gel electrophoresis, transferred to filter paper, and reacted with the antibody, as described in Materials and methods. The control is shown in lane M, where the recombinant P3A2 protein used as antigen was mixed with the size markers. The reactive band is indicated. Lane 1, crude nuclear extract, prior to fractionation; lane 2, flow through fraction from *SM50* affinity column; lane 3, high salt eluate from *SM50* column, and lane 4, preceding low salt eluate from the same column (cf. Fig. 1); lane 5, high salt eluate from *CyIIIa* column and lane 6, preceding low salt eluate from the same column. Note that P3A2 is released from the *SM50* column only in the high salt fraction, but from the *CyIIIa* column in the low salt fraction, a result consistent with the observations reproduced in Fig. 1, and also with the equilibrium dissociation measurements described in text. The sensitivity of this assay is limited (compared, e.g. to gel shift reactions), and residual small amounts of P3A2 could have been present in lanes where no signal is observed.

−36, and two other sites, also agreeing perfectly with the consensus, are identified further upstream (see Table 2). If P3A2 also acts in the *SM50* gene as a negative regulator, as the evidence suggests that it does for the *CyIIIa* gene (Hough-Evans *et al.* 1990; Höög *et al.* 1991; Davidson, 1989), it may function by antagonizing the effects of a variety of *different* positive regulators. Thus P3A2 target sites are juxtaposed to the recognition sites of several transcriptional activators in the regulatory domains of both *CyIIIa* and *SM50* genes (see Thézé *et al.* 1990 and Thiebaud *et al.* 1990). Both regulatory domains contain consensus sites for the positive transcription factor CTF/NF-1 (in these references designated P4), that are in each case located close to a P3A2 target site. In addition, P3A2 sites in the *CyIIIa* regulatory region flank a recognition sequence for an octamer protein (designated P3B), and sites for other positively acting transcription factors in the *CyIIIa* domain including the enhancer factors P2I, P2II, and the temporal activator P5, are also adjacent to the P3A sites listed in Table 2.

Specific binding affinities of P3A2

One path that we have taken to investigate differential utilization of P3A2 by the *CyIIIa* and *SM50* genes is to examine the specific affinity of the protein for the various target sites in each regulatory DNA. In the experiment shown in Fig. 7, fixed amounts of affinity-purified P3A2 and labeled *SM50* probe containing the tandem P3A sites were mixed with increasing quantities of concatenated *SM50* (oligonucleotides 25/26) or *CyIIIa* (oligonucleotides 11/12) target site DNA. The probe bound in specific complex was separated from the free probe by gel electrophoresis, and the ratio of bound to free probe was used to estimate the concentrations of P3A target sites engaged in specific complex, and remaining unbound. The value for the equilibrium dissociation constant, K_D, of the interaction was extracted from these gel shift competition data as described in Fig. 7 and by Höög *et al.* (1991). An indistinguishable measurement was obtained with recombinant P3A2 so it is unlikely that any minor contaminant present in the preparation could have affected the determination. Fig. 7A shows that the K_D value for the interaction of P3A2 with the *SM50* probe (i.e. the sites shown in Fig. 2C) was 1.7×10^{-9} M. We then measured the *relative* affinity of P3A2 for the various *CyIIIa* and *SM50* target sites, as described in the legend to Fig. 7, and these results, for all the known P3A2 sites in the *SM50* and *CyIIIa* genes, are listed in Table 2. The example shown in Fig. 7B indicates that the affinity of P3A2 for the double sites included in the *SM50* probe is 58× the affinity of this factor for the *CyIIIa* sites included in the standard competitor oligonucleotides (11/12), i.e. the *CyIIIa* site at pos-

Table 2. *Relative affinities of P3A2 target sites*

Gene[a]	Sequence[b]	Relative equilibrium constant[c]
SM50	-128 CTTCTGCGCACACCCACGCGCATGGG -102 GAAGACGCGTGTGGGGTGCGCGTACCC	43 (58)[d]
Spec 1	-84 TATCTGCGCATGCACAGATCAATCCGCGCATGCT -51 ATAGACGCGTACGTGTCTAGTTAGGCGCGTACGA	35[e]
SM50	-23 CCCTCGCGCAACAG -36 GGGAGCGCCTTGTC	8-12[f]
SM50	-356 GTCCCGCGCACACG -343 CAGGGCGCGTGTGC	1-2[f]
CyIIIa	-182 TGTTTGCGCTTCAT -195 ACAAACGCGAAGTA	1.0[g]
CyIIIa	-820 CGGCGGCGCACAAA -807 GCCGCCGCGTGTGT	1-2[h]
SM50	-249 ATTATGCGCTCATC -236 TAATACGCGAGTAG	0.5-1[f]
CyIIIa	-114 CTGAGGCGTACGAT -101 GACTCCGCATGCTA	<0.5[i]

[a] Sites listed are located within regulatory domains that according to gene transfer experiments are necessary and sufficient to promote correct expression of reporter genes to which these domains had been fused. Correct spatial expression has thus been demonstrated for *CyIIIa* gene fusions (Hough-Evans *et al.* 1987, 1988, 1990); and for *SM50* gene fusions (Sucov *et al.* 1988). Correct temporal expression has been demonstrated for the *Spec*I gene (Gan *et al.* 1990), and for the *CyIIIa* gene (Flytzanis *et al.* 1987; D. Livant, B. Haigh-Evans and E. Davidson, unpublished results). Sequences of the relevant regions of these regulatory domains have been published for *Spec*I (Hardin *et al.* 1985); for *CyIIIa* (Thézé *et al.* 1990); and for *SM50* (Sucov *et al.* 1988).

[b] Symbolism in the target sequences is as follows: The shaded boxes indicate the degenerate consensus version of the canonical P3A2 core target site, *viz* 5′ $^{C}/_{T}X^{C}/_{T}GCGC^{A}/_{T}$ (Thiebaud *et al.* 1990). Regions shaded agree perfectly with this degenerate consensus. The larger open box seen around the proximal *SM50* target site shown in the first row of the Table indicates a site composed of two half sites that are nearly perfect inverse repeats of one another. Note that as displayed some sites have been reversed from their natural orientation with respect to the transcription start site, as indicated by the sequence coordinates, so that the core site is always shown in the orientation given earlier in this Note.

[c] 'Relative equilibrium constant' is calculated as the measured equilibrium constant normalized to the equilibrium constant for the *CyIIIa* site at −182 to −195, i.e. $1.35\times10^{7}\,\text{M}^{-1}$. K_D, *equilibrium dissociation constant*, is reported in text and in Fig. 7: *equilibrium constant* is K_D^{-1}. All measurements shown were obtained using affinity-purified P3A2.

[d] The value shown is from Fig. 7A. The relative equilibrium constant obtained using the *SM50* probe and *CyIIIa*-specific competitor (oligonucleotide 11/12) (Fig. 7B) is given in parentheses.

[e] The probe in this measurement was a 55 bp *Spec*I target site constructed from oligonucleotide 52/53 (see Materials and methods). The specific competitor was the *SM50* oligonucleotide 25/26 concatenate. To control for the large difference in the lengths of the oligonucleotide 52/53 probe and the standard *SM50* probe (−50 to −201), a competition series was carried out with a 55 bp *SM50* probe (oligonucleotide 50/51).

[f] These measurements were obtained by determining the concentration of P3A2 required to occupy about 50% of the target sites in a probe containing the complete *SM50* regulatory region (Sucov *et al.* 1988). The reactions were assayed by DNAase I footprinting.

[g] The values shown were taken from an experiment similar to that shown in Fig. 7A. The probe was the 86 bp Z fragment of the *CyIIIa* gene (Thézé *et al.* 1990; this fragment contains no other factor binding site). This measurement yielded a value of $K_D=7.4\times10^{-8}$ M, using the oligonucleotide 11/12 concatenate as a specific homologous competitor.

[h] The probe used for this measurement, by the method of Fig. 7B, was fragment N of the *CyIIIa* regulatory domain (see Thézé *et al.* 1990); the specific competitor was concatenated *SM50* oligonucleotide 25/26.

[i] Reaction of P3A2 to this target was not observed unless the probe also contained the target site at position −182 to −195. The measurement was carried out by the method described in f using probe F of Thézé *et al.* (1990).

itions −182 to −195 shown in Fig. 5C. This result agrees reasonably with that obtained by direct estimation and comparison of the individual K_D values measured by the method illustrated in Fig. 7A. Thus for the same *CyIIIa* site, $K_D=7.4\times10^{-8}$, yielding a ratio of 43 for the comparison with the *SM50* sites assessed in Fig. 7A. Such comparisons represent the *observed* relative affinities, and are thus probably the relevant measure for considerations of differential P3A2 function *in vivo*. Note, however, that when the probes used contained multiple target sites the observed K_Ds may not directly indicate the *intrinsic* stabilities of the respective DNA–protein complexes. We have excluded the possibility that P3A2 binds cooperatively to the two *SM50* sites at position −102 to −128 (see Table 2) since a protein titration (not shown) provides no indication whatsoever of cooperative interaction. Nor is interaction of the two P3A2 proteins that would be bound to these two sites implied by the spacing of the sites, i.e. 14 bp, or 1.33 helical turns. Table 2 also shows a high relative equilibrium constant, about 35× that for the standard *CyIIIa* site, for the interaction of P3A2 with a second double site that is found in the *Spec*I gene. In agreement with this measurement, the results of Thiebaud *et al.* (1990) also indicated a relatively strong reaction between P3A2 (in crude extracts) and the two *Spec*I proximal sites, which are separated by 21 bp. It should be noted that there is as yet no evidence that these sites are functional and they are at least not sufficient to promote accurate expression of a *Spec*I fusion gene (Gan *et al.* 1990). Table 2 also shows that the *SM50* site at position −23 to −36 binds P3A2 about 10× more tightly than does the *CyIIIa* site used for comparison (Table 2). This is not a double site, but it differs from the majority of sites in Table 2 that also agree perfectly with the P3A2 consensus target site sequence, in that it shares three additional nucleotides with one of the strong *SM50* sites, indicated by 'x' in Table 2. Overall the strength of the P3A2 interactions varied about 100-fold. The weakest binding detected was about 50% of the value of the interaction with the standard *CyIIIa* site at position −182 to −195, while all the *CyIIIa* and *SM50* sites *other* than those just discussed bind the factor 1–2× as well as does the standard site. It follows from these measurements that

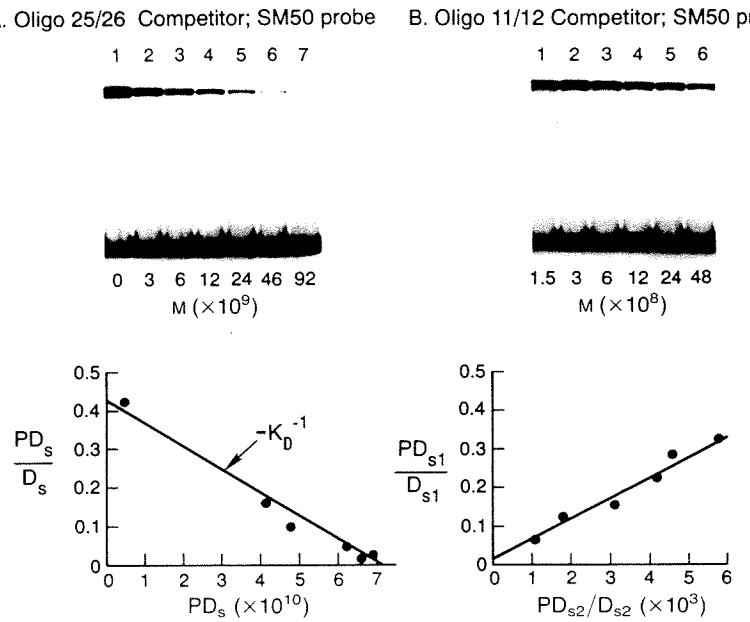


Fig. 7. Gel shift comparisons of P3A2 binding to *CyIIIa* and *SM50* P3A target sites. Representative gel retardation assays used to measure the equilibrium dissociation constants (K_D) for the interaction of affinity-purified P3A2 with *CyIIIa* and *SM50* P3A target sites are shown. These reactions contained 1.7×10^{-10} M of the *SM50* probe containing the tandem P3A consensus sites at positions −128 to −102 (see Table 2), and 0.1 μl of P3A2 purified with an *SM50* column (2 cycles; Table 1). (A) Competition series with the same binding sites, represented by concatenated oligonucleotides 25/26. Reactions were assembled that contained the amount of concatenate indicated below each lane (see Materials and methods for oligonucleotide sequences). The graph in the lower section of the figure shows a determination of K_D and P_0, the concentration of active P3A2 in the reaction, using the relation

$$\frac{PD_s}{D_s} = -\frac{1}{K_D} \times PD_s + \frac{P_0}{K_D} \qquad (1)$$

Here PD_s represents the concentration of factor: site complex, and D_s represents free probe concentration. The value of K_D determined in this experiment is 1.7×10^{-9} M. P_0 was 7.2×10^{-10} M. It was estimated that about 28% of the protein in the reaction was bound to the large excess of nonspecific poly (dAT) competitor present in the reaction (K_D for this nonspecific reaction is expected to be about 1.8×10^{-3} M [Calzone *et al.* 1988], and the poly (dAT) was present at 4.2×10^{-4} M). (B) Competition series with *CyIIIa* binding site at −182 to −195, represented by concatenated oligonucleotide 11/12 as competitor, at the concentrations indicated. Results were analyzed as follows, where the subscript 1 refers to the *SM50* site:P3A2 reaction, and the subscript 2 refers to the *CyIIIa* site:P3A2 reaction:

$$\frac{PD_{s1}}{D_{s1}} = \frac{K_{D2}}{K_{D1}} \times \frac{PD_{s2}}{D_{s2}} \qquad (2)$$

Here PD_{s1}/D_{s1} is given by the ratio of bound to free probe cts min^{-1}; K_{D1} was obtained from eq. (1);

$$PD_{s2} = P_0 - PD_{s1} - P, \quad \text{where} \quad P = \frac{K_{D1} \times PD_{s1}}{D_{s1}}$$

and D_{s2} is the concentration of oligonucleotide 11/12 in the reaction (actually it is this concentration less PD_{s2}, but the correction is insignificant (<1%)). Equations 1 and 2 follow from the definition of K_D. The value of K_{D2}/K_{D1} obtained from the experiment shown was 58.

the *effective protein concentration required for P3A2 function could vary widely for different regulatory DNAs*. Different genes would thus respond differently to given P3A2 concentrations within the same nucleus.

Discussion

In our view the sea urchin embryo presents an exciting opportunity to characterize by direct methods the major *trans*-regulatory molecules that organize the initial diversification of cell function. The lineage of the embryo is known (Cameron *et al.* 1987; Cameron and Davidson, 1991), and a specific pattern of interblastomere inductions that might account for the crucial early founder cell specifications can be inferred (Hörstadius, 1939; Davidson, 1989). A number of territory-specific genes that serve as markers for founder cell specification have been cloned. The molecular interactions that are causally responsible for territorial marker gene regulation could thus provide an image of the mechanism by which the initial territorial specifications take place. However, the usefulness of this approach depends entirely on the feasibility of cloning gene

regulatory factors, given only the target sites that constitute the *cis*-regulatory domains of key marker genes.

In this work we report what is to our knowledge the first application of site-specific affinity chromatography to cloning of regulatory factors from sea urchin embryos. The accompanying paper (Höög *et al.* 1991) describes the application of a direct ligand screening method for obtaining factor cDNA, using the same P3A target sites. Starting with target site sequences, these are the two major routes to regulatory factor characterization, i.e. other than use of homologous probes derived from preexistant clones. As an example of the latter, a heterospecific, homologous probe was used to isolate the sea urchin version of the USF factor, for which there is a target site in the *Spec*I gene (Tomlinson *et al.* 1990). It will be particularly interesting to compare sea urchin regulatory proteins isolated without any *a priori* dependence on homology, to those utilized by mammals, as the echinoderms are the only subchordate deuterostome group for which we possess significant molecular level data. Perhaps not surprisingly, the P3A2 protein described in this paper turned out to belong to no previously known class of regulatory factors. The ligand screening method applied by Höög *et al.* (1991) revealed a second factor that we call P3A1, which recognizes almost the same target sites, and is clearly a member of the Zn finger family of DNA-binding proteins. We discuss the possible interplay of P3A1 and P3A2 in the following paper.

The factor purification procedures and their practical implications

Sea urchin embryos can be raised synchronously in enormous quantities. In this work, we describe the preparation of nuclear extract from 2.5×10^9 embryos, or about 1×10^{12} nuclei. Subsequently we processed about 10× this quantity of embryos. We have found that the embryos can be frozen, and nuclear pellets and soluble protein extract obtained at later convenience. The extract itself is stable when stored frozen, and can be used for at least two years without notable loss of specific DNA-binding activity. P3A2 presented a signficant challenge for purification since it is a very low abundance factor, particularly in the 24 h embryos from which the nuclear extract was obtained. Calzone *et al.* (1988) estimated that at its peak in mid-late cleavage the average blastomere nucleus contains about 3000 molecules of P3A2, but there are only about 1200 molecules/nucleus at blastula stage. However, even at this low concentration easily available quantities of nuclear extract suffice. Thus 10^9, 24 h embryos would contain about 0.8 nanomole of P3A2, and this is more than adequate for recovery by standard methods of enough sequence to generate cloning probes. Peptide separation and mass spectrometer sequencing instrumentation now available require only 0.1–0.2 nanomoles, and current improvements in technology are rapidly lowering these limits. We have discovered, furthermore, that the same extract can be *sequentially* passed over a series of affinity columns bearing target sites for different factors. Thus a set of diverse factors can be purified from a single aliquot of nuclear extract. Recently we have succeeded in automating sequential affinity chromatography, as will be described elsewhere. The relevant point is that for the sea urchin embryo the quantities of material and the technology and methods already in hand provide ready access to any regulatory factor that in crude nuclear extracts will bind tightly and specifically to its DNA target site.

The purification of P3A2 is instructive from a quantitative point of view. The initial steps are important, since the nuclear pellet retains only 2.5 % of the total embryo protein and the nuclear extract itself includes only about 10 % of total nuclear pellet protein; hence, we estimate that a 400-fold purification has been attained even before affinity chromatography. Table 1 reports about 1600-fold purification in the affinity chromatography step (with about 50 % yield). The combined purification factor for nuclear P3A2 would thus be about 6×10^5. This was probably several-fold more purification than actually would have been required, as noted in text, and in the event we obtained far more sequence than was necessary, as illustrated in Fig. 4.

P3A2: Different affinities for target sites in CyIIIa, Spec*I and* SM50 *genes*

Although we yet lack any coherent understanding of the provenance and distribution of active form(s) of P3A2 in the cleavage-stage embryo, initial observations suggest that several different phenomena may be involved, some or all of which could be functionally important. (*i*) P3A2 exists in several different charge isoforms, as revealed by high resolution, two-dimensional gel electrophoresis (M. Harrington and F. Calzone, unpublished). Thus P3A2 is subject to modification, probably phosphorylation. Such modifications could of course affect biological activity, as predicted to account for the conditional specification of gene expression in the aboral ectoderm founder cells (Davidson, 1989). (*ii*) The same anti-P3A2 antibody used for the experiments of Fig. 6 detects P3A2 in homogenates of unfertilized eggs (R. Zeller, unpublished), though we were unable to demonstrate any maternal DNA-binding activity for P3A2 (Calzone *et al.* 1988). The maternal factor could be sequestered, complexed with an inhibitor, or modified in specific ways. (*iii*) P3A2 is also encoded by a low abundance maternal mRNA (Cutting *et al.* 1990). Whether in later embryos it is of maternal or zygotic origin, the steady state content of this mRNA per embryo remains essentially unchanged throughout early development. Cutting *et al.* (1990) showed that the amount of P3A2 that can be translated on the P3A2 mRNA could account for the total P3A2 measured in late cleavage nuclei and, of importance in this context, that the newly synthesized P3A2 would be present in much higher concentration in macromere than in micromere nuclei if the message were evenly distributed per unit volume of embryo cytoplasm. Thus active P3A2 concentration could be regulated by several different mechanisms in

the early embryo. An obvious implication is that some or all of these mechanisms are used to set up a crucial spatial distribution of active, intranuclear P3A2 concentrations, with respect to particular sets of cleavage-stage founder cells. Here the observations summarized in Table 2 of this paper are directly relevant. The relative equilibrium constants that we obtained of course represent an average for whatever P3A2 variants are present in 24 h embryo nuclei. Nonetheless, these measurements reveal sharp preferences for target sites in some genes relative to others. There are two different aspects to this phenomenon. The closely spaced double sites of both the *SM50* and *Spec*I genes display 30- to 50-fold higher affinity than the single *CyIIIa* site used as a standard in Table 2. This is unlikely to be due to interaction between P3A2 molecules since the sites are spaced differently in these two promoters, and in the *SM50* gene are not even located on the same side of the DNA helix. Perhaps each site acts as a one-dimensional diffusion 'concentrator' for the other. If it depends on one-dimensional diffusion, this effect would be expected to drop off rapidly with intersite distance (Berg and Von Hippel, 1985). However, a second explanation is required as well, *viz* that sequence elements outside of the target site core also affect the equilibrium dissociation behavior of the complexes. Thus Table 2 shows that a high relative equilibrium constant is not observed for concatenated oligonucleotide 11/12, representing the standard *CyIIIa* site, though in these concatenates there are again closely spaced target sites, here only 25 bp apart. Furthermore, the *single* proximal *SM50* target site also displays a high relative equilibrium constant, and with respect to the others listed in Table 2 this site is distinguished only by its external sequence features.

Whether the results of Table 2 are due to target site organization or to target site sequence, these features of the *cis*-regulatory DNA of different genes would give rise to different *functional interpretations of limiting intranuclear P3A2 concentrations*. If given concentrations of P3A2 are indeed specific to given sets of lineage founder cells, these different interpretations could determine which genes will be subject to P3A2 regulation in which blastomeres.

We would like to acknowledge the assistance of Patrick Leahy in the culture of the embryos. This research was supported by an NIH Grant [HD-05753] to E.H.D.; by an NSF Grant [DCB-8912530], an ACS Institutional Grant [INS-166B], and Funds from the Academic Committee on Research at UCI, to F.J.C.; and by an NSF Grant [DIR-8618937] to D.B.T. C.H. was supported by a postdoctoral fellowship from EMBO, and R.W.Z. is supported by an NIH Training Grant [GM-07616].

References

AEBERSOLD, R. H., NIKA, H., PIPES, G. D., WETTENHALL, R. E. H., CLARK, S. M., HOOD, L. E. AND KENT, S. B. H. (1989). Accelerated high sensitivity sequence analysis of proteins and peptides immobilized on chemically-modified glass fiber discs. In *Methods in Protein Sequence Analysis* (ed. B. Wittmann-Liebold), pp. 79–97. Springer-Verlag.

BENSON, S. C., SUCOV, H. M., STEPHENS, L., DAVIDSON, E. H. AND WILT, F. (1987). A lineage-specific gene encoding a major matrix protein of the sea urchin embryo spicule. I. Authentication of the cloned gene and its developmental expression. *Devl Biol.* **120**, 499–506.

BERG, O. G. AND VON HIPPEL, P. H. (1985). Diffusion-controlled macromolecular interactions. *Ann. Rev. Biophys. Biophys. Chem.* **14**, 131–160.

BRADFORD, M. (1976). A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal. Biochem.* **72**, 248–254.

CALZONE, F. J., THÉZÉ, N., THIEBAUD, P., HILL, R. L., BRITTEN, R. J. AND DAVIDSON, E. H. (1988). Developmental appearance of factors that bind specifically to *cis*-regulatory sequences of a gene expressed in the sea urchin embryo. *Genes Dev.* **2**, 1074–1088.

CAMERON, R. A., BRITTEN, R. J. AND DAVIDSON, E. H. (1989). Expression of two actin genes during larval development in the sea urchin *Strongylocentrotus purpuratus*. *Mol. Reprod. and Dev.* **1**, 149–155.

CAMERON, R. A. AND DAVIDSON, E. H. (1991). Cell type specification during sea urchin development. *Trends in Genetics* (in press).

COX, K. H., ANGERER, L. M., LEE, J. J., BRITTEN, R. J., DAVIDSON, E. H. AND ANGERER, R. C. (1986). Cell lineage-specific programs of expression of multiple actin genes during sea urchin embryogenesis. *J. molec. Biol.* **188**, 159–172.

CUTTING, A. E., HÖÖG, C., CALZONE, F. J., BRITTEN, R. J. AND DAVIDSON, E. H. (1990). Rare maternal mRNAs code for regulatory proteins that control lineage specific gene expression in the sea urchin embryo. *Proc. natn. Acad. Sci. U.S.A.* **87**, 7953–7959.

DAVIDSON, E. H. (1986). *Gene Activity in Early Development*. Academic Press: Orlando, Florida.

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

DENHARDT, D. T. (1966). A membrane-filter technique for the detection of complementary DNA. *Biochem. biophys. Res. Commun.* **23**, 641–646.

FLYTZANIS, C. N., BRITTEN, R. J. AND DAVIDSON, E. H. (1987). Ontogenic activation of a fusion gene introduced into the sea urchin egg. *Proc. natn. Acad. Sci. U.S.A.* **84**. 151–155.

GAN, L., ZHANG, W. AND KLEIN, W. H. (1990). Repetitive DNA sequences linked to the sea urchin Spec genes contain transcriptional enhancer-like elements. *Devl Biol.* **139**, 186–196.

HARDIN, S. H., CARPENTER, C. D., HARDIN, P. E., BRUSKIN, A. M. AND KLEIN, W. H. (1985). Structure of the *Spec1* gene encoding a major calcium-binding protein in the embryonic ectoderm of the sea urchin, *Strongylocentrotus purpuratus*. *J. molec. Biol.* **186**, 243–255.

HÖÖG, C., CALZONE, F. C., CUTTING, A. E., BRITTEN, R. J. AND DAVIDSON, E. H. Gene regulatory factors of the sea urchin II. Two dissimilar proteins, P3A1 and P3A2, bind to the same target sites that are required for early territorial gene expression. *Development* **112**, 000–000.

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

HOUGH-EVANS, B. R., BRITTEN, R. J. AND DAVIDSON, E. H. (1988). Mosaic incorporation and regulated expression of an exogenous gene in the sea urchin embryo. *Devl Biol.* **129**, 198–208.

HOUGH-EVANS, B. R., FRANKS, R. R., CAMERON, R. A., BRITTEN, R. J. AND DAVIDSON, E. H. (1987). Correct cell type-specific expression of a fusion gene injected into sea urchin eggs. *Devl Biol.* **121**, 576–579.

HOUGH-EVANS, B. R., FRANKS, R. R., ZELLER, R. W., BRITTEN, R. J. AND DAVIDSON, E. H. (1990). Negative spatial regulation of the lineage specific *CyIIIa* actin gene in the sea urchin embryo. *Development* **108**, 41–50.

IKEGAKI, N. AND KENNET, R. H. (1989). Glutaraldehyde fixation of the primary antibody-antigen complex on nitrocellulose paper

increases the overall sensitivity of immunoblot assay. *J. Imm. Meth.* **124**, 205–210.

KADONAGA, J. T., CARNER, K. R., MASIARZ, F. R. AND TJIAN, R. (1987). Isolation of cDNA encoding transcription factor Sp1 and functional analysis of the DNA binding domain. *Cell* **51**, 1079–1090.

LAEMMLI, U. K. (1970). Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* **227**, 680–685.

LYNN, D., ANGERER, L., BRUSKIN, A., KLEIN, W. AND ANGERER, R. (1983). Localization of a family of mRNAs in a single cell type and its precursors in sea urchin embryos. *Proc. natn. Acad. Sci. U.S.A.* **80**, 2656–2660.

PARKER, C. S. AND TOPOL, J. (1984). A *Drosophila* RNA polymeraseII transcription factor binds to the regulatory site on an hsp70 gene. *Cell* **37**, 273–283.

PIRROTTA, V., MANET, E., HARDON, E., BICKEL, S. E. AND BENSON, M. (1987). Structure and sequence of the *Drosophila* zeste gene. *EMBO J.* **6**, 791–799.

STUDIER, F. W. AND MOFFAT, B. A. (1986). Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes. *J. molec. Biol* **189**, 113–130.

STURM, R. A., DAS, G. AND HERR, W. (1988). The ubiquitous octamer-binding protein Oct-1 contains a POU domain with a homeo box subdomain. *Genes Dev.* **2**, 1582–1599.

SUCOV, H. M., BENSON, S., ROBINSON, J. J., BRITTEN, R. J., WILT, F. AND DAVIDSON, E. H. (1987). A lineage-specific gene encoding a major matrix protein of the sea urchin embryo spicule. II. Structure of the gene and derived sequence of the protein. *Devl Biol.* **120**, 507–519.

SUCOV, H., HOUGH-EVANS, B. R., FRANKS, R. R., BRITTEN, R. J. AND DAVIDSON, E. H. (1988). A regulatory domain that directs lineage specific expression of a skeletal matrix protein gene in the sea urchin embryo. *Genes Dev.* **2**, 1238–1250.

THÉZÉ, N., CALZONE, F. J., THIEBAUD, P., HILL, R. L., BRITTEN, R. J. AND DAVIDSON, E. H. (1990). Sequence of the *CyIIIa* actin gene regulatory domain bound specifically by sea urchin embryo nuclear protein. *Mol. Reprod. and Dev.* **25**, 110–122.

THIEBAUD, P., GOODSTEIN, M., CALZONE, F. J., THÉZÉ, N., BRITTEN, R. J. AND DAVIDSON, E. H. (1990). Intersecting batteries of differentially expressed genes in the sea urchin embryo. *Genes Dev.* **4**, 1999–2010.

TOMLINSON, C. R., KOZLOWSKI, M. T. AND KLEIN, W. H. (1990). Ectoderm nuclei from sea urchin embryos contain a Spec-DNA binding protein similar to the vertebrate transcription factor USF. *Development* **110**, 259–272.

VINSON, C. R., LAMARCO, K. L., JOHNSON, P. F., LANDSCHULTZ, W. H. AND MCKNIGHT, S. L. (1988). *In situ* detection of sequence-specific DNA binding activity specified by a recombinant bacteriophage. *Genes Dev.* **2**, 802–806.

(Accepted 6 February 1991)

Appendix C

Territorial expression of three different *trans*-genes in early sea urchin embryos detected by a whole-mount fluorescence procedure

Robert W. Zeller, R. Andrew Cameron, Roberta R. Franks, Roy J. Britten and Eric H. Davidson

DEVELOPMENTAL BIOLOGY 151, 382-390 (1992)

Territorial Expression of Three Different *trans*-Genes in Early Sea Urchin Embryos Detected by a Whole-Mount Fluorescence Procedure

ROBERT W. ZELLER, R. ANDREW CAMERON, ROBERTA R. FRANKS,[1] ROY J. BRITTEN, AND ERIC H. DAVIDSON

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

Accepted February 24, 1992

We have developed a new procedure for detection of the protein product of chloramphenicol acetyltransferase (CAT) reporter genes in whole mounted sea urchin embryos. The position of a commercially available anti-CAT antibody is visualized by video or confocal microscopy, and thus the spatial domains of exogenous reporter gene expression can be determined with regard to the intact three-dimensional structures of the embryo. We show that in pluteus stage embryos CAT protein expression patterns for *SM50 · CAT* or *CyIIIa · CAT* reporter genes are similar to those previously obtained by *in situ* hybridizations with radioactive probes. Taking advantage of the superior resolution of cellular CAT expression patterns using the antibody visualization method, we found for the first time that, in addition to the expression in aboral ectoderm, some cells in the ciliated band of the pluteus express *CyIIIa · CAT*. The expression of a new fusion construct, *CyIIa · CAT*, was also examined. As expected from the localization of endogenous *CyIIa* mRNA, CAT protein was expressed under control of the *CyIIa* promoter in gut and skeletogenic mesenchyme cells.

INTRODUCTION

The initial processes by which spatial patterns of differential gene expression are instituted in the sea urchin embryo occur during cleavage. Five different territories, defined by their unique embryonic fates and each the product of an invariant set of early lineage elements, can be recognized in the early embryo (Davidson, 1989; Cameron and Davidson, 1991). By late cleavage, marker genes have been activated within specific territories (Davidson, 1986, 1989). In order to unravel the molecular basis of territorial specification, we and others have made extensive use of gene fusions in which reporter sequences are expressed under the control of the regulatory regions of such regionally expressed marker genes. By this means regulatory sequence elements that are necessary and sufficient to generate appropriate territorial patterns of fusion gene expression in aboral ectoderm (Gan *et al.*, 1990; Hough-Evans *et al.*, 1987, 1988, 1990) and in skeletogenic mesenchyme (Sucov *et al.*, 1988) have been identified. We have found that the widely used bacterial drug resistance gene encoding chloramphenicol acetyltransferase (CAT) serves as a reliable reporter sequence in sea urchin embryos. The enzyme is easily extracted in embryo homogenates and it can be assayed quantitatively, which is extremely useful for a variety of purposes, for example, for temporal analysis of expression (see, e.g., Franks *et al.*, 1990; Livant *et al.*, 1988, 1991). There is no similar endogenous sea urchin enzymatic activity. However, a drawback for studies of spatial expression of exogenous CAT fusion genes is that it has been necessary to section the embryos in order to detect CAT mRNA by *in situ* hybridization. No generally useful whole-mount *in situ* hybridization procedure has yet been found, nor has a cytological whole-mount detection method for CAT enzyme been available. In this work we demonstrate a new whole-mount procedure for CAT reporter gene expression. This method relies on cytological detection of CAT enzyme by fluorescence of a commercial anti-CAT antibody in permeabilized embryos expressing CAT fusion genes (Harlan *et al.*, 1991; Zuber *et al.*, 1989). Though the method in our hands is not yet as sensitive as *in situ* hybridization, the use of whole mounts affords three-dimensional resolution at the cellular level significantly higher than that of sections. We have taken advantage of this resolution to make the unexpected observation that the *CyIIIa · CAT* fusion is expressed in some ciliated band cells, as well as in all regions of the aboral ectoderm, as has been observed earlier (Hough-Evans *et al.*, 1987, 1988, 1990). Some regions of the ciliated band are in fact partially composed of cells deriving from certain aboral ectoderm lineages (R. Cameron and E. Davidson, unpublished observations).

In this paper we present the whole-mount antibody staining procedure and then describe its application to three different CAT fusion constructs. Two of these,

[1] Present address: National Institute of Dental Research, Bethesda, MD 20892.

0012-1606/92 $5.00

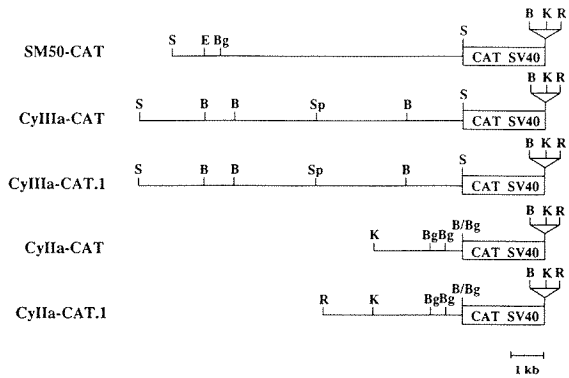


FIG. 1. Restriction maps of fusion gene constructs. Solid lines represent the upstream regulatory region and the hatched box represents the CAT and SV40 regions. Plasmid vector sequences are not indicated. B, *Bam*HI; Bg, *Bgl*II; B/Bg, *Bam*HI/*Bgl*II fusion site; E, *Eco*RV; K, *Kpn*I; R, *Eco*RI; S, *Sal*I, Sp, *Sph*I.

SM50 · CAT and *CyIIIa · CAT*, have been studied earlier, while the third, *CyIIa · CAT*, is an entirely new construct.

MATERIALS AND METHODS

Eggs and Embryos

Strongylocentrotus purpuratus gametes were collected and embryos injected as previously described (McMahon *et al.*, 1985). Briefly, the eggs were dejellied in pH 5 filtered sea water (FSW) and electrostatically attached to 60-mm plastic Petri dishes previously coated with 1% protamine sulfate. An inverted Leitz microscope outfitted with a 16°C cold stage was used for microinjections. Just before microinjection, attached eggs were fertilized with a freshly prepared sperm suspension. Each egg was then microinjected one to three times with 1000–2000 copies of the appropriate fusion gene construct along with carrier DNA (fivefold mass excess of *Hin*dIII-digested sea urchin DNA). In a typical experiment, each of the CAT reporter genes was injected into 400–500 zygotes. Embryos were reared at 16°C and after 24 hr of development, the hatched embryos (about 70–80% of the singly injected zygotes and 50% or less for multiply injected zygotes) were pooled into flasks containing 50 ml sea water and allowed to grow 3–6 days until well-developed plutei were formed. At this stage, the oral and aboral territories of the embryo were well distinguished, so that cell identities could be assigned with no ambiguity.

CAT Fusion Genes

An *SM50 · CAT* fusion, two *CyIIIa · CAT* fusions, and two *CyIIa · CAT* fusion genes were used. Schematic diagrams of these are shown in Fig. 1. The 14-kb *CyIIIa · CAT* construct was linearized at the *Sph*I site, which is located about 5 kb upstream from the start codon (Flytzanis *et al.*, 1987). The *pCyIIIa · CAT.1* construct contains the same regulatory domain as *CyIIIa · CAT*, but fused to a different CAT vector assemblage (Franks *et al.*, 1990). For injection it was linearized at the *Kpn*I site downstream from the CAT and SV40 sequences. In the *SM50 · CAT* construct (Sucov *et al.*, 1988) the sea urchin sequence is fused to the CAT vector at the second exon. It contains 2.2 kb of upstream *SM50* sequence and about 7 kb consisting of the first intron and the first two exons. For injection it was linearized at the *Bam*HI site downstream of the CAT and SV40 sequences. The two *CyIIa · CAT* constructs utilize upstream sequences originally isolated from a λ genomic library (Scheller *et al.*, 1981). To construct *CyIIa · CAT*, a *Kpn*I-*Bam*HI fragment that extends from a *Kpn*I site located 2.7 kb upstream of the ATG start codon to 37 bp downstream of the start codon was subcloned into pUC18. This 2.7-kb region includes a 600- to 700-bp leader intron. The *Kpn*I-*Bam*HI fragment was then isolated and fused at a *Bgl*II site to the 1.6-kb CAT fragment from pUC^{PL}CAT (Bond-Matthews and Davidson, 1988) to create *CyIIa · CAT*. For injection it was linearized at the *Kpn*I site (Fig. 1). *CyIIa · CAT.1* was derived from an *Eco*RI fragment of the λ clone. It contains a regulatory region longer than that of *CyIIa · CAT*, including an additional 1.6 kb of upstream sequence. For injection it was linearized at a downstream *Bam*HI site that follows the CAT reporter assemblage, as shown in Fig. 1.

CAT Antibody Staining

Embryos were placed in well slides and fixed in cold 90% methanol, 10 m*M* EDTA for 6 min as described by Harris (1986). Antibody staining methods were a modification of those described by Crowther *et al.* (1990). Briefly, fixative was removed with several washes of sterile (autoclaved) Millipore-filtered sea water (SMFSW) and the embryos were transferred to microtiter well plates (Falcon). The remaining steps were carried out under a dissecting microscope. Embryos were digested with 2 μg/ml proteinase K in SMFSW for 1 hr at room temperature. The proteinase K was removed and the embryos were rinsed for 15 min in 25 m*M* glycine in SMFSW and then rinsed for 1 hr at room temperature with several changes of buffer A (0.1% BSA, 0.05% Triton X-100 in SMFSW). The embryos were washed for at least 8 hr at 4°C with buffer B (buffer A containing 10% normal goat serum and 1% Triton X-100). Buffer B was removed and 50 μl of a 1/100 dilution of a commercial rabbit polyclonal antibody to chloramphenicol acetyltransferase (5′ → 3′, Inc.) in

buffer B was added and incubated for at least 8 hr at 4°C. The primary antibody was removed with several changes of buffer A and 100 μl of a 1/125 dilution of a fluorescein- isothiocyanate (FITC)-labeled goat- anti-rabbit secondary antibody (EY Labs) in buffer B was added and incubated for at least 8 hr at 4°C. After several washes with buffer A, embryos were wet mounted under a coverslip and observed. Visual images were recorded with a light-intensifying video camera (SIT) attached to an image-processing system (Imaging Technologies), or on a Bio-Rad confocal microscope. Composites of fluorescent and bright field images were assembled after background subtraction and dynamic range adjustment. Embryos that had not been microinjected were used for controls.

RESULTS AND DISCUSSION

Fluorescence Detection of SM50 · CAT Expression

In an initial series of experiments we studied the spatial distribution of CAT enzyme in embryos bearing the *SM50 · CAT* fusion. The spatial pattern of expression of this construct had been observed by *in situ* hybridization in previous work (Sucov *et al.*, 1988). The *SM50* gene encodes a spicule matrix protein and is expressed, beginning in late cleavage, exclusively in the lineages of the skeletogenic mesenchyme territory, both before and after ingression (Benson *et al.*, 1987; Killian and Wilt, 1989; Sucov *et al.*, 1987; Katoh-Fukui *et al.*, 1991). The *SM50* regulatory sequences included in the *SM50 · CAT* fusion were shown to direct the same pattern of CAT mRNA accumulation as that observed for endogenous *SM50* mRNA (Sucov *et al.*, 1988). At the pluteus stage the skeletogenic mesenchyme cells are easily distinguished in whole mounts by their linear arrangement along the skeletal rods, and since they are located internally, we considered localization of the *SM50 · CAT* reporter expression particularly useful for development of effective permeabilization, immunocytological, and fluorescence detection protocols.

The *SM50 · CAT* construct was introduced into zygotes by single injections of 1000–2000 copies of the linearized plasmid (see Materials and Methods). Morphologically normal embryos were collected after 96 hr and processed for CAT detection. About 50% of the processed embryos expressed detectable CAT protein. Expression was observed exclusively in skeletogenic mesenchyme cells (we examined about 100 embryos in this series of experiments). As expected from the mosaic incorporation resulting from single injections of exogenous DNA (McMahon *et al.*, 1985; Hough-Evans *et al.*, 1988; Franks *et al.*, 1988; Livant *et al.*, 1991), in most of the embryos only a portion of the 64 mesenchyme cells present in the pluteus stage embryo displayed CAT expression. Examples of video image recordings of *SM50 · CAT* expression patterns are shown in Figs. 2A–2C. In the embryo shown in Figs. 2A and 2B the cells displaying detectable CAT enzyme constitute about half of the total number of skeletogenic mesenchyme cells (i.e., 27/64 cells; note that not all the labeled cells can be seen in any one plane of focus). Thus in this case stable nuclear incorporation most likely occurred in one of the first two blastomeres (in 75% of singly injected eggs, stable nuclear incorporation occurs after second, third, or fourth cleavage; Hough-Evans *et al.*, 1988; Livant *et al.*, 1991). Therefore, even if there were instead two incorporations after second cleavage, the *SM50 · CAT* construct must have been present in all territories of the embryo, and thus its exclusive expression in skeletogenic mesenchyme (Figs. 2A and 2B) is significant. In the example shown in Fig. 2C, the embryo contained 13 obviously labeled skeletogenic mesenchyme cells and again no cells other than skeletogenic cells express the CAT reporter. In these embryos there could have occurred a single incorporation after either second, third, or fourth cleavage. We think it likely that in a few skeletogenic cells in each of these embryos the signal was not intense enough to have been detected and amplified, thus accounting for the difference between the 13 labeled cells and the 16 cells expected in a clone containing one-fourth of the skeletogenic mesenchyme cells. Other explanations are not excluded, e.g., loss of exogenous DNA or failure to visualize all the skeletogenic mesenchyme cells because of optical occlusion.

We conclude from these *SM50 · CAT* experiments that the immunolocalization and permeabilization procedures suffice for the detection of CAT reporter expression in cells located internally within the embryo. The appropriate, specific pattern of fusion gene expression was exclusively observed. Therefore, the procedure described under Materials and Methods eliminates all significant background fluorescence. These experiments showed that cell type specificity of CAT fusion gene expression can be assessed immunocytologically in whole embryos. The optical imaging procedures we applied in this and the following procedures (see Materials and Methods) provide superb cellular and structural resolution, relative to *in situ* hybridization (particularly ^{35}S hybridization) in sections. However, the *SM50* promoter appears to be particularly powerful in late embryos (Killian and Wilt, 1989; Sucov *et al.*, 1987, 1988) and sensitivity is thus relatively a less serious issue with the *SM50 · CAT* fusion than with other territory-specific reporter gene fusions, as shown below.

CyIIa · CAT Expression

Cox *et al.* (1986) showed that the *CyIIa* cytoskeletal actin gene is expressed in various portions of both the

definitive vegetal plate and the skeletogenic territories, the founder cells for which are adjacent in the midcleavage embryo (Cameron *et al.*, 1987; Cameron and Davidson, 1991). *In situ* hybridization with a probe specific for *CyIIa* mRNA (Lee *et al.*, 1986; Cox *et al.*, 1986) displays this message in ingressed skeletogenic mesenchyme and in the vegetal plate at the mesenchyme blastula stage. In the mid- to late gastrula, *CyIIa* expression is limited to secondary mesenchyme. In pluteus stage embryos, this gene is expressed in the stomach, intestine, and coelomic rudiments. However, it remained unclear whether skeletogenic mesenchyme cells reinitiate transcription of *CyIIa* at the pluteus stage (Cox *et al.*, 1986), or whether it was other forms of mesenchyme that displayed CyIIa mRNA in the later embryos studied by Cox *et al.* (1986). Two different fusions, *CyIIa·CAT* and *CyIIa·CAT.1*, that contain 2.7 and 4.3 kb, respectively, of upstream sequence were constructed as described briefly in Materials and Methods.

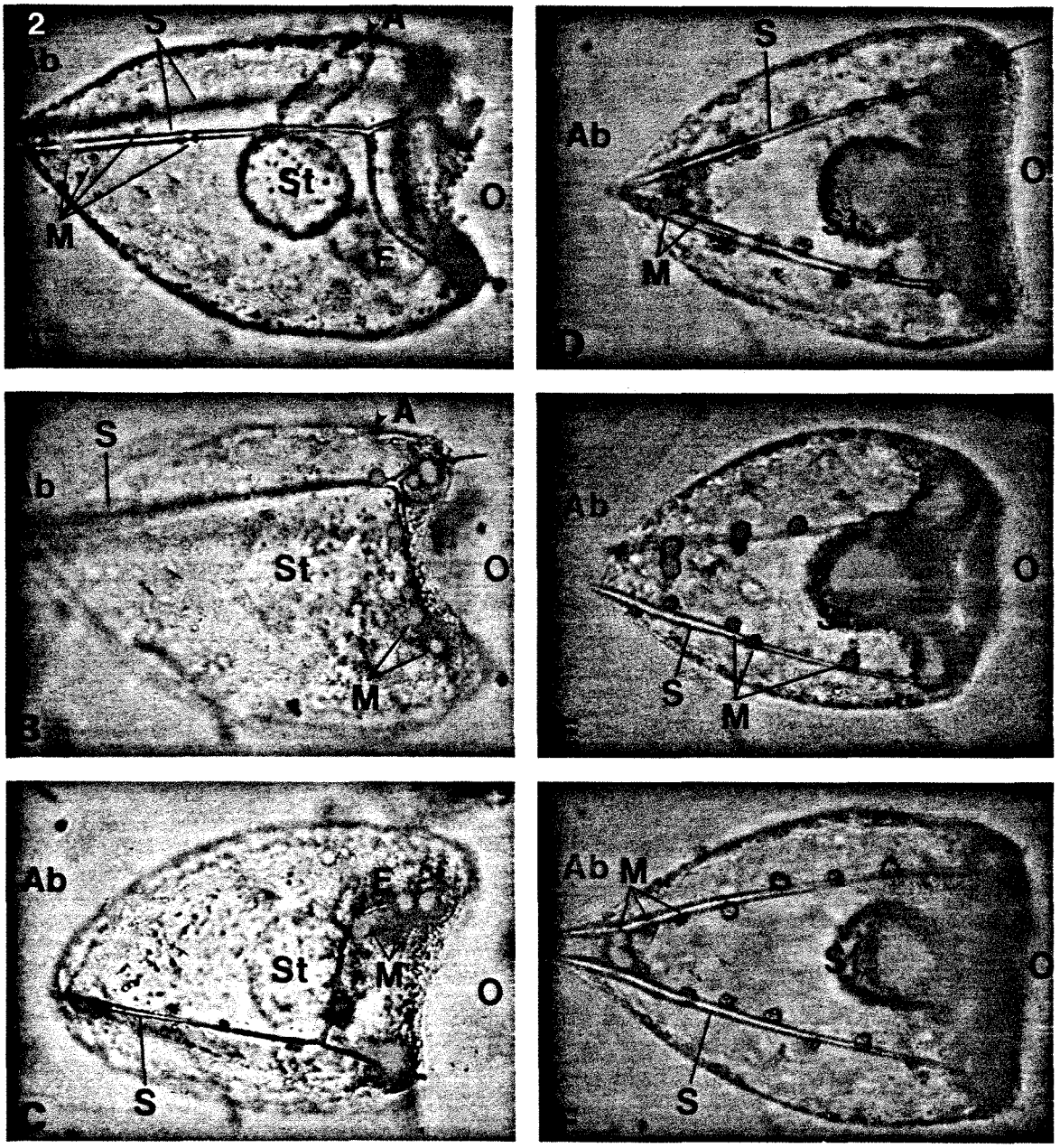
Zygotes were singly microinjected with 1000–2000 copies of either fusion gene, and 50 morphologically normal plutei were collected for antibody processing after 72 hr. Detectable *CyIIa·CAT.1* expression was observed in about 50% of these embryos and the expression pattern recorded. In 75% of the labeled embryos, *CyIIa·CAT.1* expression occurred in both skeletogenic mesenchyme cells and cells forming the stomach and intestine. The remaining embryos expressed *CyIIa·CAT.1* in either skeletogenic mesenchyme or cells of the stomach or intestine. No expression was ever observed in oral or aboral ectoderm. Since stable incorporation of exogenous DNA occurs randomly in all lineages (Hough-Evans *et al.*, 1988; Livant *et al.*, 1991), and many embryos therefore would include the *CyIIa·CAT.1* sequence in ectoderm lineages, the regulatory sequences included in the *CyIIa·CAT.1* construct suffice either to repress expression in ectoderm territories or to promote expression exclusively in vegetal plate and skeletogenic territories (or both). Four plutei displaying CAT expression driven by the *CyIIa* regulatory domain are shown in Figs. 2D–2F. These embryos contained the longer of the two fusion constructs, i.e., *CyIIa·CAT.1.* However, the same results were obtained with the *CyIIa·CAT* fusion, indicating that the regulatory sequences required to generate the same spatial pattern of expression as that observed for *CyIIa* mRNA at pluteus stage are located within the 2.7 kb of sea urchin sequence included in this gene fusion. CAT expression is clearly evident in the skeletogenic mesenchyme cells that are arrayed along both of the skeletal rods, which are clearly visible, and also in the stomach. Although difficult to distinguish in Fig. 2, because the intestine lies in the same optical axis as the stomach in these particular embryos, other examples show that *CyIIa·CAT* expression occurs in the intestine as well as the stomach (as does *CyIIa* mRNA; Cox *et al.*, 1986). It was not determined whether expression occurs in secondary mesenchyme derivatives such as the coelomic pouches and esophageal muscle, and embryos at an earlier stage when secondary mesenchyme is being delaminated were not examined. Since in all the embryos shown in Figs. 2D–2F *both* skeletogenic and vegetal plate derivatives manifestly include the *CyIIa·CAT* construct, the latest time that a single stable incorporation event could have given rise to this pattern (Hough-Evans *et al.*, 1988) would have been a third cleavage macromere, though of course such an event could have occurred earlier, after first or second cleavage. In either case about $\frac{1}{4}$ or $\frac{1}{2}$ of vegetal plate and skeletogenic derivatives would have contained and expressed the construct (see Cameron *et al.*, 1987, 1991). While this is difficult to determine for the gut, CAT expression was detected in 17–28 of the 64 skeletogenic mesenchyme cells in each of the embryos shown in Figs. 2D–2F (see figure legend).

Immunofluorescence Detection of CyIIIa·CAT Fusion Genes

The *CyIIIa* cytoskeletal actin gene serves as an early and specific marker of aboral ectoderm territorial specification, as shown originally by *in situ* hybridization using a probe specific for this mRNA (Cox *et al.*, 1986). *CyIIIa* is transcriptionally activated in late cleavage (Hickey *et al.*, 1987) within a division or two of completion of the segregation of aboral ectoderm lineage founder cells from lineages destined for other territories (Cameron *et al.*, 1987, 1990). Expression of the *CyIIIa·CAT* and p*CyIIIa·CAT.1* fusions has been extensively studied. The 2.3-kb regulatory domain included in both of these constructs suffices to direct spatial, temporal, and quantitative expression of CAT mRNA that, within the limitations of the measurements, appears identical to that of the endogenous *CyIIIa* mRNA (Flytzanis *et al.*, 1987; Livant *et al.*, 1988, 1991; Hough-Evans *et al.*, 1987, 1988, 1990; Franks *et al.*, 1990).

We examined CAT expression by the immunocytological method in 3- to 6-day-old plutei that developed from zygotes injected two or three times with a total of

FIG. 2. *SM50·CAT* and *CyIIa·CAT* expression in 3- and 4-day-old plutei. Composite false color video images of bright field (blue) and fluorescent (red) exposures are shown superimposed. (A, B) Two different focal planes of the same embryo expressing *SM50·CAT* in which the oral face is to the right, with the anus pointing up. The CAT-expressing mesenchyme cells are located along both of the spicules, which are seen



respectively in the focal planes reproduced in (A) and (B). The total number of CAT-expressing cells in this embryo is 27, not all of which can be distinguished in these illustrations. (C) An *SM50 · CAT*-expressing embryo oriented with the oral face to the right with anus pointing down. A group of closely spaced mesenchyme cells at the tip of the postoral arm is responsible for the extensive labeling in that region. Label was recorded in 13 skeletogenic mesenchyme cells in this embryo. (D–F) *CyIIa · CAT*-expressing embryos oriented with the oral faces to the right and viewed from the abanal surfaces. The intense signal in the gut makes it difficult to distinguish individually labeled cells. By varying the focal plane and signal intensity, it can be seen that in these embryos portions of the intestine also express CAT enzyme. Embryos shown in (D–F) have 28, 17, and 22 labeled skeletogenic mesenchyme cells, respectively. A, anus; Ab, aboral ectoderm; E, esophogus; O, oral ectoderm; M, skeletogenic mesenchyme cells; S, spicule; St, stomach.

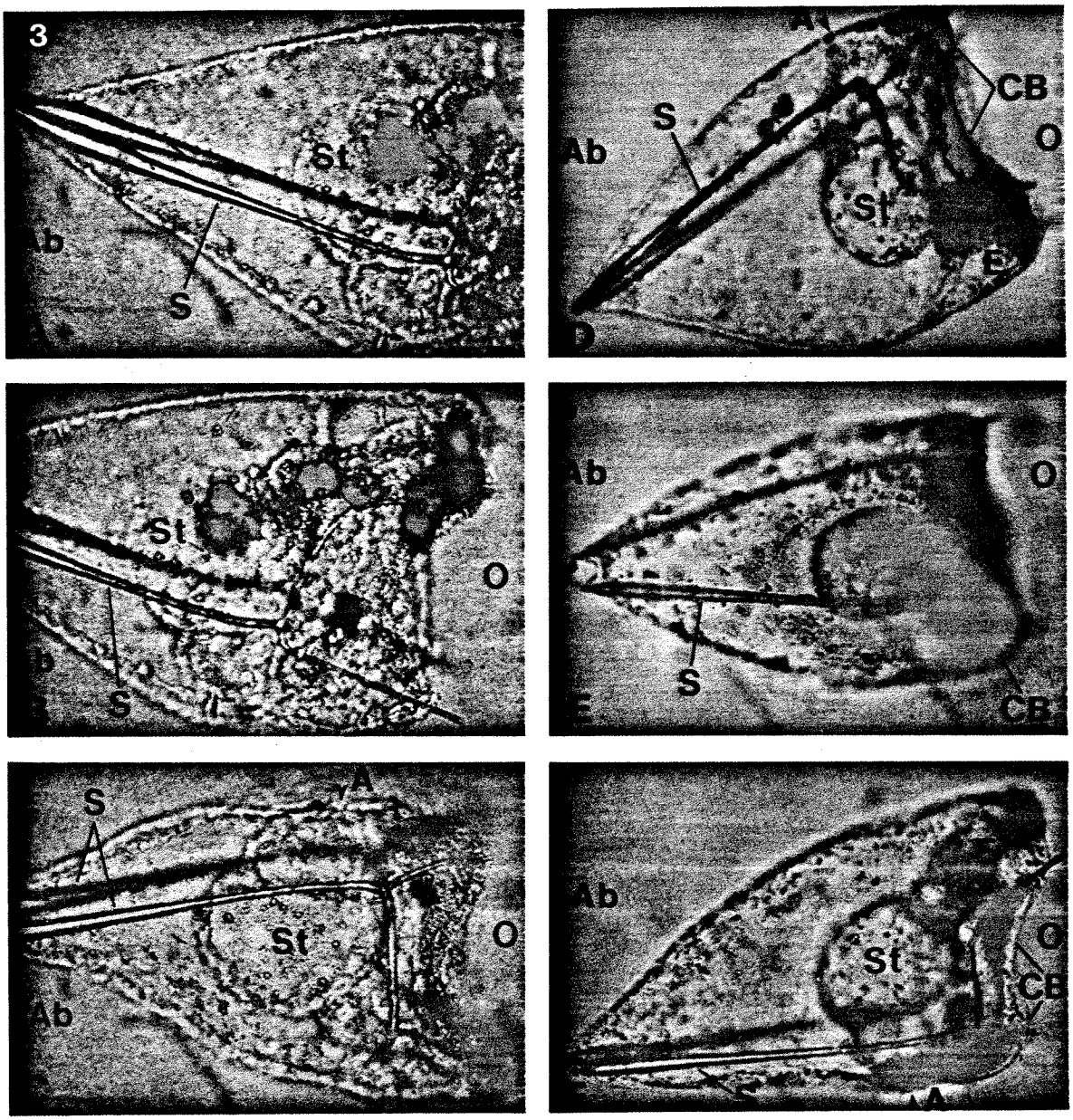


FIG. 3. Composite video images of bright field and fluorescent exposures of 6-day-old plutei expressing p*CyIIIa·CAT.1* (A–C) or 3- to 6-day-old plutei expressing *CyIIIa·CAT* (D–F). (A–B) Two views of the same embryo oriented with the oral face to the right with the anus facing down. The red cells expressing CAT are present in the aboral ectoderm at the tip of the embryo apex, above the gut and in the arm tip. (C) The oral face of this embryo is to the right with the anus facing up. CAT-expressing cells are detected in the aboral ectoderm overlying part of the esophogus and stomach. (D) Confocal image of a 3-day-old pluteus with *CyIIIa·CAT* expression in the ciliated band. Confocal microscopy greatly improves the signal-to-noise ratio in labeled embryos and the three-dimensional analysis of labeled cells is thus enhanced. To obtain these images a single midsagittal bright field image of the embryo was combined with an 11-frame confocal series of fluorescent signal. In each optical section, noise from noncoplanal fluorescence is greatly reduced, and the final image has a much higher resolution than could be obtained using conventional fluorescence microscopy. Cells labeled within the ciliated band are clearly visible. Note that signal extends along the

1000–2000 copies of *pCyIIIa·CAT.1*, as described by Livant *et al.* (1991). Multiple injection greatly increases the fraction of cells in which the exogenous DNA is stably incorporated and subsequently expressed in the appropriate lineages (Livant *et al.*, 1991). In many embryos from some batches of eggs the exogenous DNA is retained in all or almost all cells of the embryo. However, the mortality rate may also be increased by multiple injection, and in these experiments about 50% of the zygotes failed to develop. Some embryos that survived beyond 24 hr were morphologically abnormal and contained loose, disorganized cells within the blastocoel that often expressed CAT enzyme. When morphologically abnormal embryos were eliminated from the sample, no ectopic CAT expression was observed; i.e., in normally formed pluteus stage embryos CAT expression was never detected in any internal cells, including gut or mesenchyme cells, nor was CAT expression ever seen in the facial oral ectoderm that is the product of the *No* and the *Vo* founder cell clones (Cameron *et al.*, 1987, 1991). However, the observed CAT expression in these embryos was limited to relatively confined patches of aboral ectoderm cells and was usually not observed in the expected large regions of aboral ectoderm. Examples are shown in Fig. 3. Different views of the same embryo are reproduced in Figs. 3A and 3B to show all of the antibody labeling. CAT-expressing cells are visible in two groups, one located at the tip of the aboral ectoderm and the second present in the aboral ectoderm extending toward the ciliated band. From their position, we can identify the labeled cells in the posterior apex as descendants of the third cleavage *VA* blastomere, which (at sixth cleavage) gives rise to 2 of the 11 clones constituting the aboral ectoderm (Cameron *et al.*, 1987, 1990). The CAT-expressing cells on the right side of the aboral ectoderm in this embryo derive from the *Na* and the *NL* blastomeres. The aboral ectoderm cells expressing CAT in the embryo shown in Fig. 3C derive from *NL* and *Na* blastomeres. In the ~100 morphologically normal embryos that we examined, we rarely saw extensive CAT expression that would correspond to the *whole* of the regions of the aboral ectoderm contributed by the second to sixth cleavage founder cells as established by lineage tracer experiments (Cameron *et al.*, 1987, 1990). In contrast, even in singly injected embryos, we have observed large regions of cells that displayed CAT mRNA by *in situ* hybridization (see, e.g., Fig. 3 in Hough-Evans *et al.*, 1990), and this is more commonly seen in multiply injected embryos (unpublished data of B. Hough-Evans and D. Livant; cf. Livant *et al.*, 1991). Furthermore, as illustrated in Fig. 2, expression of *SM50·CAT* and *CyIIa·CAT* constructs can indeed be observed in the expected, relatively large clonal elements of the embryo. It is possible that CAT mRNA translation or CAT protein stability are deficient in aboral ectoderm, relative to that in mesenchyme and gut cells. However, the most likely interpretation may be that CAT protein expression is somewhat lower in aboral ectoderm cells because the *CyIIIa* promoter is less active in late embryos (Lee *et al.*, 1986; Lee, 1986). The flat geometry of the squamous aboral epithelium may also contribute to a lower signal intensity. In any case the implication is that the sensitivity of the immunofluorescence detection method in practice suffices to reveal a signal, after computational background suppression, in only a fraction of the cells that actually express the CAT reporter at levels that are detectable by *in situ* hybridization to CAT mRNA.

In Figs. 3D–3F, some embryos expressing *CyIIa·CAT* had a relatively intense signal localized in regions of the ciliated band, a three- to six-cell-deep structure composed of columnar cells, that lies at the interface between the oral and aboral ectoderm in the postgastrular embryo (Pearse and Cameron, 1991). The long axis of these cells protrudes from the squamous aboral epithelium, and this may account for the bright fluorescence signal that we observed in the whole-mount preparation. An example in which we utilized confocal microscopy to improve spatial resolution is shown in Fig. 3D. The ciliated band cells are difficult to distinguish clearly in most *in situ* hybridization sections, and CAT mRNA has not been noted in previous studies of *CyIIIa·CAT* expression by that method (Hough-Evans *et al.*, 1987, 1988, 1990). In our opinion, published *in situ* hybridizations to *CyIIIa* mRNA do not clearly show endogenous *CyIIIa* actin expression in any cells of the ciliated band.

CONCLUSIONS

The Whole-Mount Immunofluorescence Method

The method described provides a means of assessing spatial CAT expression from fusion constructs in sea urchin embryo whole mounts. CAT is now a widely used reporter in these embryos. Many CAT constructs have

ciliated band and is also observed in additional cells in the aboral ectoderm. (A–B) and (E–F) are images made with conventional fluorescence microscopy (i.e., with image intensification and processing as above). (E) An anal view of a 4-day-old pluteus showing a region of CAT-expressing cells extending from aboral ectoderm into the ciliated band. The oral face is to the right. (F) A 4-day-old embryo showing extensive labeling along the ciliated band with several labeled cells also located toward the posterior tip of the embryo. The oral face is to the right with the anus pointing down. CB, ciliated band; other labels are as described in the legend to Fig. 2.

already been described, and in current work these are being used to introduce regulatory mutations and other variants. CAT is also an advantageous reporter because CAT enzyme activity can be assayed quantitatively with great sensitivity in sea urchin embryo extracts (McMahon *et al.*, 1984). In comparison with *in situ* hybridization to CAT mRNA in sections, the whole-mount immunofluorescence method that we have developed has both advantages and disadvantages. It is easier, since the demanding skills required for good-quality sectioned preparations are not needed, and the cytological resolution afforded by whole-mount immunocytology is far superior. The method also lends itself to analysis by image processing and confocal microscopy. With the confocal microscope a series of optical sections of the embryos is observed, and thereby more useful three-dimensional reconstructions of the data can be generated. Images of these optical sections can then be enhanced through digital processing and combined with a bright field image. The resulting composite picture is of superior quality to conventional fluorescent images. However, the immunocytological method may in practice be somewhat less sensitive than *in situ* hybridization, as we discuss above. Furthermore, while indirect immunofluorescence from CAT-expressing cells can be detected with less-sophisticated equipment, complete analysis in fact depends on instrumentation designed for low-light-level fluorescence microscopy. This is because the fluorescent signals generated by the fluorochrome-conjugated antibody range from low to high intensity, and often several different images at different gain levels need to be recorded, in order to see all of the labeled cells expressing CAT. When quantities of expressed protein and therefore fluorescent signal are low, a low-light-level camera is necessary to see the cells that express CAT.

Expression of Three Different Fusion Genes

We show here that under the control of three different *cis*-regulatory domains the CAT reporter molecule is expressed in three different territorial patterns in the sea urchin embryo. In each case the pattern of expression coincides with the loci of expression of the endogenous genes from which the regulatory domains were derived. The unexpected observation of *CyIIIa·CAT* expression in certain regions of the ciliated band is consistent with the lineage history of these cells, whose parents probably belonged to certain of the aboral ectoderm lineages (R. Cameron and E. Davidson, unpublished observations).

The main import of the observations that we report is confirmation of the fundamental paradigm that spatial patterns of gene expression in the embryo depend only on *trans*-interactions within the regulatory domains of territory-specific genes (Davidson, 1989). This holds for a gene expressed in an autonomously specified territory, i.e., the *SM50* gene of the skeletogenic lineages, as for genes expressed in conditionally specified territories, as are *CyIIIa* and *CyIIa*. Active transcriptional regulatory factors within each territory interact with the promoter region of the reporter gene, the expression of which is in each case activated or repressed depending on the presence or absence of the appropriate factors. Thus the patterns of CAT protein expression that we report here again indicate that in molecular terms, a territory may be thought of as a region in which a specific set of functional *trans*-factors is present. The spatial distribution of functional *trans*-factors within the territorial founder cells thus determines which batteries of genes shall be expressed within each region of the embryo.

We thank Dr. Scott E. Fraser for the use of the confocal microscope. This research was supported by NIH Grant HD-05753 to E.H.D. and NSF Grant DCB-8911829 to R.A.C. R.W.Z. is supported by NIH Training Grant GM-07616.

REFERENCES

Benson, S., Sucov, H., Stephens, L., Davidson, E., and Wilt, F. (1987). A lineage-specific gene encoding a major matrix protein of the sea urchin embryo spicule. I. Authentification of the cloned gene and its developmental expression. *Dev. Biol.* **120,** 499–506.

Bond-Mathews, B., and Davidson, N. (1988). Transcription from each of the *Drosophila* actin 5c leader exons is driven by a separate functional promoter. *Gene* **62,** 289–300.

Cameron, R. A., and Davidson, E. H. (1991). Cell type specification during sea urchin development. *Trends Genet.* **7,** 212–218.

Cameron, R. A., Fraser, S. E., Britten, R. J., and Davidson, E. H. (1990). Segregation of oral from aboral ectoderm precursors is completed at fifth cleavage in the embryogenesis of *Strongylocentrotus purpuratus*. *Dev. Biol.* **137,** 77–85.

Cameron, R. A., Fraser, S. E., Britten, R. J., and Davidson, E. H. (1991). Macromere cell fates during sea urchin development. *Development* **113,** 1085–1092.

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

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.

Crowther, R. J., Meedel, T. H., and Whittaker, J. R. (1990). Differentiation of tropomyosin-containing myofibrils in cleavage-arrested ascidian zygotes expressing acetylcholinesterase. *Development* **109,** 953–959.

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

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

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

Franks, R. R., Anderson, R., Moore, J. G., Hough-Evans, B. R., Britten, R. J., and Davidson, E. H. (1990). Competitive titration in living sea

urchin embryos of regulatory factors required for expression of the CyIIIa actin gene. *Development* **110,** 31–40.

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

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

Harlan, R. E., Mondal, D., Coleman, R., and Prakash, O. (1991). Immunocytochemical localization of chloramphenicol acetyltransferase as a single-cell marker of transfected and transgenic cells. *Biotechniques* **10,** 304–306.

Harris, P. M. (1986). Cytology and immunocytochemistry. *Methods Cell Biol.* **27,** 243–262.

Hickey, R. J., Boshar, M. F., and Crain, W. R., Jr. (1987). Transcription of three actin genes and a repeated sequence in isolated nuclei of sea urchin embryos. *Dev. Biol.* **124,** 215–227.

Hough-Evans, B. R., Britten, R. J., and Davidson, E. H. (1988). Mosaic incorporation and regulated expression of an exogenous gene in the sea urchin embryo. *Dev. Biol.* **129,** 198–208.

Hough-Evans, B. R., Franks, R. R., Cameron, R. A., Britten, R. J., and Davidson, E. H. (1987). Correct cell-type specific expression of a fusion gene injected into sea urchin eggs. *Dev. Biol.* **121,** 576–579.

Hough-Evans, B. R., Franks, R. R., Zeller, R. W., Britten, R. J., and Davidson, E. H. (1990). Negative spatial regulation of the lineage specific CyIIIa actin gene in the sea urchin embryo. *Development* **110,** 41–50.

Katoh-Fukui, Y., Noce, T., Ueda, T., Fujiwara, Y., Hashimoto, N., Higashinakagawa, T., Killian, C. E., Livingston, B. T., Wilt, F. H., Benson, S. C., Sucov, H. M., and Davidson, E. H. (1991). The correct structure of the SM50 spicule matrix protein of *Strongylocentrotus purpuratus*. *Dev. Biol.* **45,** 201–202.

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

Lee, J. J. (1986). "Activation of Sea Urchin Actin Genes During Embryogenesis: Nuclear Synthesis and Decay Rate Measurements of Transcripts from Five Different Genes. PhD Thesis, California Institute of Technology.

Lee, J. J., Calzone, F. 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.

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

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

McMahon, A. P., Flytzanis, C. N., Hough-Evans, B. R., Katula, K. S., Britten, R. J., and Davidson, E. H. (1985). Introduction of cloned DNA into sea urchin egg cytoplasm: Replication and persistence during embryogenesis. *Dev. Biol.* **108,** 420–430.

McMahon, A. P., Novak, T. J., Britten, R. J., and Davidson, E. H. (1984). Inducible expression of a cloned heat shock fusion gene in sea urchin embryos. *Proc. Natl. Acad. Sci. USA* **81,** 7490–7494.

Pearse, J. S., and Cameron, R. A. (1991). Echinodermata Echinoidea. *In* "Reproduction of Marine Invertebrates" (A. C. Giese and J. S. Pearse, Eds.), Vol. 6, pp. 514–663. Boxwood Press, Pacific Grove, CA.

Scheller, R. H., McAllister, L. B., Crain, W. R., 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. *Mol. Cell. Biol.* **1,** 609–628.

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

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

Zuber, M. X., Strittmatter, S. M., and Fishman, M. C. (1989). A membrane-targeting signal in the amino terminus of the neuronal protein GAP-43. *Nature* **341,** 345–348.