Characterization of the sea urchin homologue of the replication factor A 70 kD subunit and the novel interspersed repeat family to which it binds

Thesis by

Roger Anderson

In Partial fulfillment of the Requirements

for the Degree of

Doctor of Philosophy

California Institute of Technology

Pasadena, California

1996

(Submitted June 1995)

This work is dedicated to He by whom all things are, that are.

"Science is the study of how God made things"
James E. Talmage

Acknowledgments

I want to start by thanking my wonderful and beautiful eternal wife, Christeve; without whom I would be lost. She has been a source of strength and support when all else seemed to fail. I love her with all my heart. I thank my two sons Ryan and Brett for their patience and long suffering while Daddy was always working. I would also like to thank my family and friends for their love and support. I could not have made it this far without them. An extra special thank you to my wonderful mother-in-law Phyllis Helton who helped me edit my thesis and to my Grandfather Felix Salamon for the example and inspiration that he has been to me throughout my life.

I would like to thank Eric for teaching me the fundamentals of development and guiding me through my development as a scientist. His devotion to quantitative biology pushed me to my limits. I would also like to thank the members of my committee; Howard Lipshitz, James Strauss, David Anderson and Scott Frasier, who have given me strength and encouragement over the years. Special kudos for Howard who was even on my candidacy committee lo those many years ago.

A big thank you goes to the members of the Eric Davidson Lab who helped me, guided me, taught me and put up with me, especially; Bob® Zeller, Barbara Hough-Evans, Jane Rigg, Ann Cutting, Frank Calzone, Pat Leahy, Jim Moore, Carmen Kirchhamer, Roberta Franks, Jim Coffman, Donal Manahan, Henry Sucov, Joe Minor and my two assistants Catherine Moore, and Mary Chen.

"Bob" is a registered trademark of Microsoft Corporation

Abstract

The 5' regulatory region of the *Strogylocentrotus purpuratus* cytoskeletal actin gene, *CyIIIa*, contains a 150 bp inverted interspersed repeat. The center of the inverted repeats contains binding sites for a 12 zinc finger DNA binding protein; **SpZ12-1**. This repeat was also found in the flanking region of the genes for two factors; **SpZ2-1** and **SpP3A2** which bind to the *CyIIIa* promoter. The high degree of conservation of the repeats is indicative of potential protein binding sites. Gel-shift assays using sea urchin egg cytoplasmic extracts revealed a second binding activity, on a separate region of the repeat, for a protein which we later identified as the sea urchin homologue of the 70 kD subunit of replication factor A , **SpRFA-70**.

Replication factor A is a well known single-stranded DNA (ssDNA) binding protein that is required for DNA replication (Kim et al., *Mol. Cell. Biol.* **12**, 3050-3059, 1992). We demonstrate that in addition to its non-specific affinity for ssDNA, **SpRFA-70** has a high sequence specific binding affintiy for double-stranded DNA (dsDNA). The equilibrium constant (K_{eq}) for SpRFA-70 binding to ssDNA is 1.7 $x10^{10}$ M^{-1} and the K_{eq} for binding to the specific dsDNA site is $1.8X10^{9}$ M^{-1} as determined by protein excess titration. Quantitative gel retardation assays reveal that the **SpRFA-70** dsDNA binding activity is two orders of magnitude more prevalent per egg than per whole embryo at 24 hours. These data suggest that **SpRFA-70** functions as a regulatory factor predominantly in the egg and early embryo.

A genomic library was screened with short oligo probes to determine the number of sites for the DNA binding factors **SpZ12-1** and **SpRFA-70**. The results of this screen indicate there are ~460 **SpRFA-70** sites and ~318 **SpZ12-1** sites. These sites match to a consensus sequence by >80%. There are 40-80 instances where these sites occur together. This is 30 fold higher than would be expected by random association. The distance between the two binding sites is also conserved. This arrangement is the same as that which occurs in the CyIIIa regulatory region. That this configuration is so conserved implies a potential regulatory interaction for these two factors.

Table of Contents

Chapter 1

Introduction

Identification of small specific sequence sites in cloned DNA libraries

This introduction sets the stage for the chapters that follow. Various means of identifying sequences that are of interest in the study of development are described. Also discussed is a method of short oligonucleotide screening that was modified to identify, in the entire genome, multiple instances of important DNA sequences. The important parameters of this method will be detailed and a range of other possible uses will be outlined.

The second chapter of this thesis deals with a novel use of specific sequence oligonucleotide screening. The quantity and quality of SpRFA-70/SpZ12-1 repeat element copies in the genome of *Strongylocentrotus purpuratus* is determined. The SpRFA-70/SpZ12-1 repeat is so named for the two DNA-binding proteins SpRFA-70 (Ch.3) and SpZ12-1 (Wang *et al.*, 1995) that bind to different sites in this repeat (Fig. 1). The SpRFA-70 region is referred to as the P10 site or P10 repeat in Chapters 2 and 3.

The third chapter deals with the isolation and characterization of the protein that binds to a specific but as yet undetermined sequence in the *SpRFA-70* region of the SpRFA-70/SpZ12-1 repeat. This protein turned out to be the sea urchin homologue of the replication factor A 70 kD subunit. Replication factor A (RFA) is a three subunit protein that functions as a single-stranded DNA binding protein during replication (Kim *et al.*, 1992). The relative affinity and the equilibrium constants were determined for SpRFA-70 binding to double-stranded DNA (dsDNA) and to single-stranded DNA (ssDNA). These data are compared with previously published data regarding the DNA binding capacities of human and yeast RFA-70.

Appendix 1 describes the application of the specific oligonucleotide screening method on a genome blot to determine the range of animal species that carry the two binding sites of the SpRFA-70/SpZ12-1 repeat family. The results are consistent with earlier findings that repeat families are conserved only among closely related species (Moore *et al.*, 1978). A method for determining the function of a set of protein recognition sites from the 5' flanking region of the CyIIIa actin gene is described in Appendix 2.

Why would it be useful to be able to recover multiple occurrences of a particular protein binding site in the genome?

It is generally accepted that gene regulation in development is a function of the coordinated expression of gene networks (Britten and Davidson, 1971). Gene network studies require multiple members of the network in question to demonstrate their common or shared elements. To find the full range of regulatory proteins for these networks it will be essential to identify the sites of protein:DNA interaction. Once these sites have been identified this information can be used to look for additional copies of those sequences in the genome. By comparing the expression patterns of genes sharing various binding sites, those factors that are causal for the common expression patterns seen for the members of a particular battery of genes can be more easily identified. In other words, the appearance of a common binding site in the promoter regions of several genes that are related only on the basis of their temporal or spatial pattern of expression, could indicate that the shared factor is responsible for the similar behavior. Obviously, further testing using in situ hybridization and reporter constructs would be necessary to clarify the situation. It will

also become easier to identify the factors that act to specify the expression patterns of the various gene networks as more sequence data from promoter regions is accumulated. As an example of this, after a number of neuro-specifc gene transcripts were isolated researchers have been able to identify elements that contribute to their common expression patterns by comparisons of their regulatory regions (Schoenherr and Anderson, 1995).

In this introductory chapter I describe a method for obtaining additional occurrences of previously identified short sequences sites, here specific protein binding sites in the genomic DNA, and I show how this method can be used to determine the number of these sites in the genome that are homologous within a designated criterion of homology. Most methods of identifying sites of protein:DNA interaction do not and can not directly identify additional genes or genomic regions that share common elements. This method does not rely on common expression patterns, on mutation studies or on extensive (=expensive) sequencing projects for the identification of additional sites of potential interaction for known DNA binding factors. It can be used to identify additional copies of any given sequence, of sufficient length, in the genome.

Traditional methods used to identify protein binding sites

There are several biochemical methods for identifying protein:DNA sites of interaction. 1) Reporter constructs using segments from the DNA 5' to a transcribed region are often used to define the segment of DNA flanking a transcribed region that is necessary and sufficient for correct expression of the particular gene. Additional mutation or deletion constructs can then be used to narrow down the exact location, identity and function of individual protein binding sites. 2) Studies using purified factors or nuclear

extracts can be used to identify protein binding sites on isolated genomic regions by gel-shift (Thézé *et al.*, 1990) or by footprint analysis (Wang *et al.*, 1995). 3) A set of potential binding sites for a protein can be isolated from a pool of degenerate oligonucleotides using an affinity column of purified protein attached to a resin and a cycle of PCR and gel-shift separations (Aufiero *et al.*, 1994). All of these biochemical methods require the prior identification of the gene and the isolation of the genomic region from which the gene is expressed.

Genetic studies are another valuable method for the identification of potential protein binding sites but they require as a subsequent test one or more of the experiments mentioned above. Genetic evidence can lead to the identification of putative target genes for given transcription factors. Identification of the specific target sites requires extensive comparisons of promoter regions and/or detailed in vitro mutagenesis studies. A difficulty in genetic screening for gene networks is the possibility that redundancies in regulatory mechanisms will mask the effect of the deletion of a particular gene product. Another important factor is that these methods require an organism in which genetic studies can be performed.

A third method of identifying potential binding sites is computer assisted promoter comparison. This method does not require that the genes have similar patterns of expression that they be from the same organism or even from the same genus. In some cases, comparisons of flanking regions of genes from similar organisms or homologous genes can indicate potential binding sites, because of the selective restrictions on DNA sequences that interact with a regulatory protein. In theory and in practice such sites will

mutate more slowly as they require a corresponding change in the binding protein in order to maintain the proper function and will hence appear as conserved sequence islands. However, frequently the subtle base changes allowable at protein binding sites and the short recognition sequences make the task of identification of sites by homology alone nearly impossible. Not until a number of sites have been identified for a particular factor can the allowable substitutions be determined. Clearly there needs to be some direct means of identifying multiple transcripts that have common regulatory elements in their flanking DNA.. This work demonstrates the use of sequence specific oligonucleotide screening to isolate additional genomic regions containing identified protein binding sites.

Short oligonucleotide screening of cloned DNA libraries

For several years short oligonucleotide probes have been used to screen libraries in attempts to isolate clones for specific proteins or gene families (Wood *et al.*, 1985, Honoré *et al.*, 1993). Typically the labeled probe consists of a set of degenerate oligonucleotides based on amino acid sequence from a conserved domain in homologous proteins or from amino acid sequence obtained from purified protein. These probes are hybridized to the target sequence under controlled conditions. The excess probe is removed by repeated washing steps with decreasing amounts of sodium chloride. This causes all but the most complementary probe:target hybrids to dissociate. The dissociation is a function of the melting temperature (T_m) of the hybrid. In sodium chloride the T_m is based on the length of the probe, the GC concentration of the probe and the molarity of the salt in the buffer. There are several difficulties in this method when using short probes. One difficulty is the increased effect of the base composition on the

T_m with short synthetic probes. Due to the limited number of bases the GC percentage can drastically differ between the various probes in a degenerate mix making it very difficult to determine the correct melting temperature (T_m). The GC percentage in longer probes tends to average out and usually ends up being closer to that of the average DNA of the organism (in sea urchin DNA the GC content is about 40% and is slightly higher in expressed sequences).

To render the melting reaction in the wash steps length dependent only on probe length, reagents have been developed that eliminate the effects of base composition on the T_m. Von Hippel's group has extensively investigated the use of two such reagents: tetraalkylammonium (TAA) salts (Melchior and Von Hippel *et al.*, 1973) and betaine (N,N,N-trimethylglycine, Rees *et al.*, 1993). Both of these products enable the user to control the conditions of the hybridization or the washes. By doing so distinct numbers of allowable base pair mismatches or differences can be selected without regard to GC% of the probe or target sequence.

The TAA salts that have been the most carefully studied are tetramethylammonium chloride (TMACL) and tetraethylammonium chloride (TEACL). TMACL does not effect the overall stability of the DNA hybrids as do longer side chain TAA salts, such as TEACL. The long side chain TAA salts destabilize the basepair stacking and so reduce the overall T_m of the duplex by 30°C or more. The TAA salts reduce the effect of the GC% on T_m by binding to the A:T base pairs. Experiments probing the B form of DNA with TMACL and TEACL analogs indicate that their binding to the DNA occurs in the major groove and that they associate with and stabilize A:T base pairs. This stabilization

results in the elimination of base composition effect on melting temperatures (Melchior and Von Hippel 1973). The increased preferential contact with the A:T base pairs is probably due to hydrophobic interactions with the methyl group on the thymidine base. The presence of TAA+ salts binding in the major groove is confirmed by experiments using DNA from a native strain of T4 phage, that has normally glucosylated dC residues in its DNA, and a mutant strain of T4 that does not have glucosylated DNA (Rees *et al.*, 1993). The glucosylated (hydroxymethyl) cytosine residues are known to fill the major groove of DNA. This occlusion inhibits the TAA+ salt from associating with the DNA as evidenced by a lack of stabilization of native T4 DNA.

Betaine is believed to function in a manner similar to the TAA salts in reducing the effect of base composition on the T_m. Betaine, like the long side chain TAA salts, reduces the T_m by ~20°C. The mechanism of destabilization has not been specifically identified. The temperature decrease is probably due to the reduced polarity of the solvent (Rees *et al.*, 1993). The duplex destabilization property of betaine and the long side chain TAA salts does not make them as suitable as TMACL for some purposes. In order to use small oligonucleotide probes the T_m in the final wash solution must be well above room temperature. For longer probes and different situations the temperature reduction capacity of TEACL or betaine can be very useful to bring the working temperature range into more convenient limits.

To illustrate the range of effective use for TMACL with different length oligonucleotides, the number of base pair mismatches allowed, before the duplex melts, has been plotted versus the blot wash temperature in a solution of 3M TMACL buffer

(Fig. 2). The correlation between allowed mismatch and wash temperature is linear for each length Thus the T_m is dependent on length only. The results in Chapter 2 demonstrate that this correlation is very specific and exact. As the probe lengths increase past 100 bp the effect of length on the T_m is less pronounced, as indicated by formula (1) in the legend to Fig. 2. This also reduces the effectiveness of the TMACL in eliminating the base composition effect. The procedure used in Ch 2 requires TMACL only in the wash steps. This is a variation from the methods described by others where TMACL is used in the hybridization and washing buffers (Wood *et al.*, 1985, Honoré *et al.*, 1994). It proved not to be necessary to use TMACL in the hybridization buffer. Hybridization in standard conditions using 0.9 M NaCl buffer at T_m -20°C, was sufficient for proper identification of target sites. The use of TMACL only in the wash steps allows greater control over the amount of probe left duplexed to the filter DNA.

Oligonucleotide Probe Design

Obviously specific sequence probes are relatively simple to design. The only factors to consider are the length and the prevalence of the particular sequence in the genome as discussed above. Degenerate oligonucleotide probes have several factors that make their use more difficult and challenging. To design a probe from homologous regions of DNA those positions that are the most highly conserved should be chosen as the basis for the probe. If the starting materials are proteins, then the additional factors of codon redundancy and codon usage of the organism under study must be considered. Choosing peptide sequences that have the fewest possible codon choices reduces the degeneracy of the probe, and therefore increases the actual amount of specific probe in the

final hybridization. A brief look at Figure 2 shows that increased probe length increases the specificity and the melting temperature. What is not shown however, is the increase in less than full-length probes that are produced in the synthesis of longer oligos. After 40 synthesis cycles less than 50% of the starting strands will be completely elongated even if the rate of successful strand elongation is ~98%. Another factor not included in the chart is the additional cost of longer probes.

One of the additional benefits of using TMACL in this type of screening protocol is that the filters can be autoradiographed at any stage of the process and thus intermediate wash steps can be analyzed to show the various levels of conservation of the target site for the probe in use, whether it is a repeat fragment or a protein binding site. The washes can also be done empirically when the degree of similarity between the probe and target is unknown. The results are analyzed between washings until all of the probe has been removed. A genome blot can also be washed in this manner to determine the optimal conditions for screening a cloned DNA library. This is particularly important in cases where the exact sequence of the target is not known, such as in the case of homologous proteins from different species. This greatly reduces the chances of missing a clone due to excessively stringent washing of the filters.

Additionally inosine can be substituted for positions where there is a high degree of degeneracy. Inosine is able to pair with A, C, and T in mild sodium buffer without problems(Honoré *et al.*, 1993). Inosine also reduces the stability of the duplex as a whole thereby reducing the overall T_m of the hybrid. In these cases the ability to wash in steps is

a necessity. The use of inosine also contributes to an increase in false positives obtained in the screening process.

Specific short oligonucleotide probes

Using specific sequence short oligonucleotide probes, genome blots or cloned DNA libraries can be tested directly for additional instances of any particular repeat family or known target site greater than ~12 bases long. A sample flowchart of our protocol is outlined in Fig. 3. If the repeat family is highly prevalent (>1,000 copies), based on qualitative estimates from a genome blot, then the number of genomic clones that will be needed in the screen to isolate additional copies of the repeat is very low.

Organization of Thesis

This introduction to the technology developed in my work constitutes Chapter 1.

Chapter 2 describes the identification of the SpRFA-70/SpZ12-1 inverted repeat element. This repeat element was initially identified as a conserved sequence that occurs in the *cis*-regulatory domain of the CyIIIa cytoskeletal actin gene and in the flanking sequences of the genes for two transcription factors that bind to the same target sites in the CyIIIa promoter; SpZ2-1 and Sp3A2. The number of copies of two different segments of the repeat is determined. The high frequency in which these two segments of the repeat occur together is a possible indication of a function for this combined site though each segment also occurs independently. The identification of a protein that binds with high affinity to a particular region of the repeat element is described. The quantity of protein in the egg and 24 hour embryo nucleus is compared. These findings are described in: Anderson *et al.*, 1994

Chapter 3 describes the characterization of a factor that binds to the SpRFA-70/SpZ12-1 inverted repeat element. Also described is the cloning of a cDNA that codes for this protein. We identify this protein as the sea urchin homologue of replication factor A 70 kD subunit (SpRFA-70). We demonstrate that it binds with high specificity to one particular section of the repeat. SpRFA-70 is shown to have both double- and single-stranded DNA binding activities. The discussion section of this chapter also summarizes the work that comprises this thesis. This paper is in preparation for publication.

Appendix 1 describes a search for the presence of the two regions of the SpRFA-70/SpZ12-1 repeat in the genomes of other organisms. Genomic DNAs from several sea urchin species as well as 15 other unrelated species were used to screen for these regions.

Appendix 2 details a series of in vivo competition experiments in which various portions of the CyIIIa promoter region were co-injected with a reporter construct consisting of the CyIIIa control region fused to the CAT reporter protein coding sequence. This paper was published as Franks *et al.*, 1990

Future Prospects

Several avenues of exploration have now opened up with regard to SpRFA-70 and the regulation of gene expression. One area which still requires some work is the identification of the actual binding site for SpRFA-70. This will require the cloning and expression in bacteria of all three subunits of the holoenzyme. Researchers using the human RFA have been successful at expressing all three subunits in the same bacterium using a common T7 promoter. With recombinant SpRFA it should be possible to obtain sufficient quantities of the factor to obtain a clear footprint.

It would be interesting to see if, for any of genes for the three subunits, there is an SpRFA-70 site in the upstream flanking regions. If present this would imply a self-regulatory mechanism and demonstrate at least one regulatory role for SpRFA-70. Preliminary work indicated that a clone containing coding region for SpZ12-1 has an SpZ12-1 site but not an SpRFA-70. Since little is known of the function of RFA in gene regulation, if all of the genes for the RFA subunits have RFA-70 sites in their control regions this set of genes might be an ideal network or battery on which to work.

References

1. Anderson, R. F., 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
2. Aufiero, B., Neufeld, E. J. and Orkin, S. H. (1994). Sequence-specific DNA binding of individual cut repeats of the human CCAAT displacement/cut homeodomain protein *Proc. Natl. Acad. Sci. USA* **91**, 7757-7761
3. Britten, R. J. and Davidson, E. H. (1971). Repetitive and non-repetitive DNA sequences and a speculation on the origins of evolutionary novelty. *The Quarterly Review of Biology* **46**, 111-138
4. Franks, R. R., Anderson, R. F., 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
5. Honoré, B., Madsen, P. and Leffers, H. (1993). The tetramethylammonium chloride method for screening cDNA libraries using highly degenerate oligonucleotides obtained by backtranslation of amino-acid sequences *J. of Biochem. and Biophys. Meths.*, **27**, 39-48
6. Kim, C., Snyder, R. O., and Wold, M. S. (1992). Binding properties of replication protein A from human and yeast cells. *Mol. Cell. Biol.* **12**, 3050-3059
7. Melchior Jr., W. B. and Von Hippel, P. H. (1973) Alteration of the relative stability of dA•dT and dG•dC base pairs in DNA *Proc. Natl. Acad. Sci. USA* **70**, 298-302

8. Moore, G. P., Scheller, R. H., Davidson, E. H. and Britten, R. J.(1978). Evolutionary change in the repetitive frequency of sea urchin DNA sequence. *Cell* **15**, 649-660

9. Rees, W. A., Yager, T. D., Korte, J. and Von Hippel, P. H. (1993) Betaine can eliminate the base pair composition dependence of DNA melting *Biochemistry* **32**, 137-144

10. Schoenherr, C. J. and Anderson, D. J. (1995). The neuron-restrictive silencer factor (NRSF): A coordinate repressor of multiple neuron-specific genes *Science* 267, 1360-1363

11. 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

12. 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

13. Wang, D. G.-W., Kirchhamer, .C. V., Britten, R.B., and Davidson, E. H. (1995). *SpZ12-1*, a negative regulator required for spatial control of the territory-specific CyIIIa gene in the sea urchin embryo. *Development* **121**, 1111-1122

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

Fig. 1 *Diagram illustrating the two regions of the SpRFA-70/SpZ12-1 repeat.* The gray area represents the region containing the SpRFA-70 binding site. The gray arrow above that section indicates the position and orientation of the 28 bp oligonucleotide probe used to detect that portion of the whole repeat. The specific binding site for SpRFA-70 is somewhere within the sequence of the probe. The black area represents the section containing the SpZ12-1 site. The black arrow below the SpZ12-1 region indicates the site and position of the 21 bp oligonucleotide probe used to detect this segment of the repeat. The exact location of binding for SpZ12-1 is indicated by the box. The same probes were used for PCR amplifications.

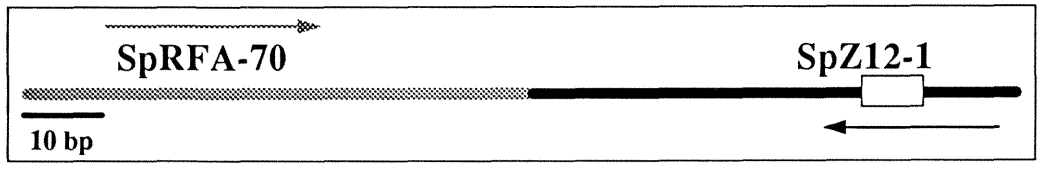

SpRFA-70
SpZ12-1
10 bp

Fig. 2 *Number of allowed mismatched base pairs vs wash temperature.* The lines show the decreasing number of allowed mismatches at the particular temperature for the indicated probe lengths. For specific length probes we use the following relations to calculate the number of allowable mismatches and the appropriate wash temperature:

(1) $$T_m = 93^\circ C - \left[\frac{600}{L}\right]$$

(2) $$\text{mismatch} = \left[\frac{L}{100}\right] \bullet (T_m - \text{washtemp})$$

(3) $$\text{washtemp} = \left[\frac{100 \bullet \text{mismatch}}{L}\right] - T_m$$

T_m is the melting temperature, and L is the length of the oligonucleotide.

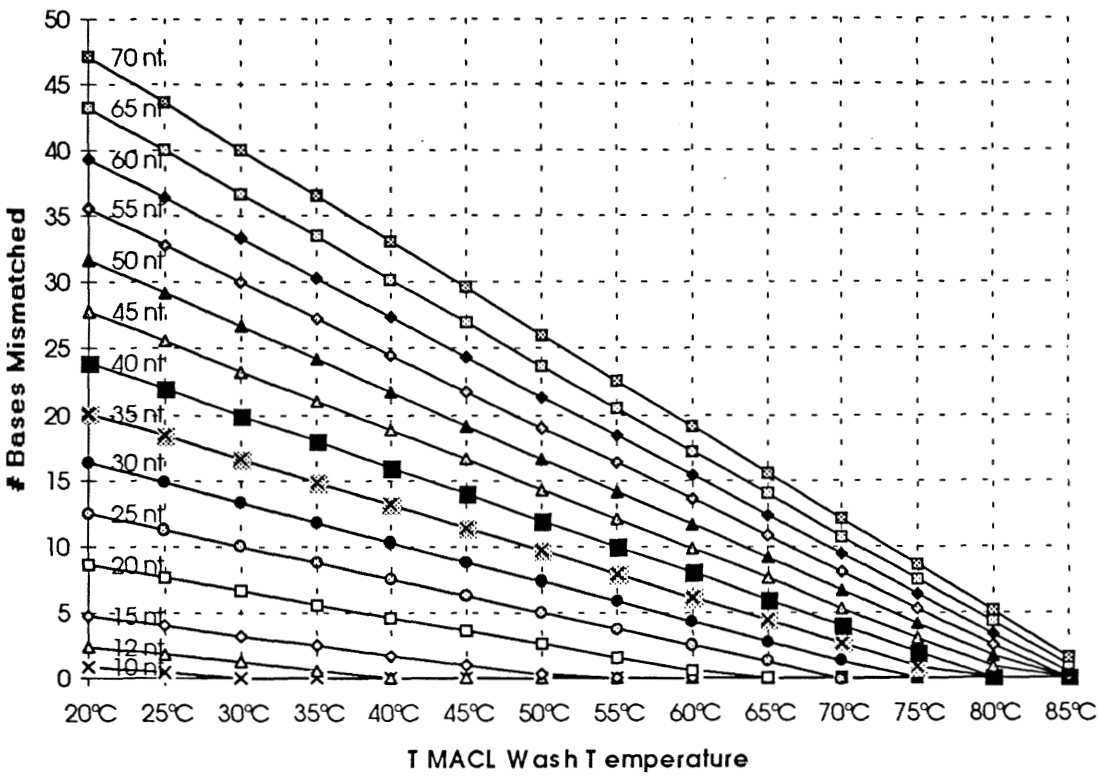
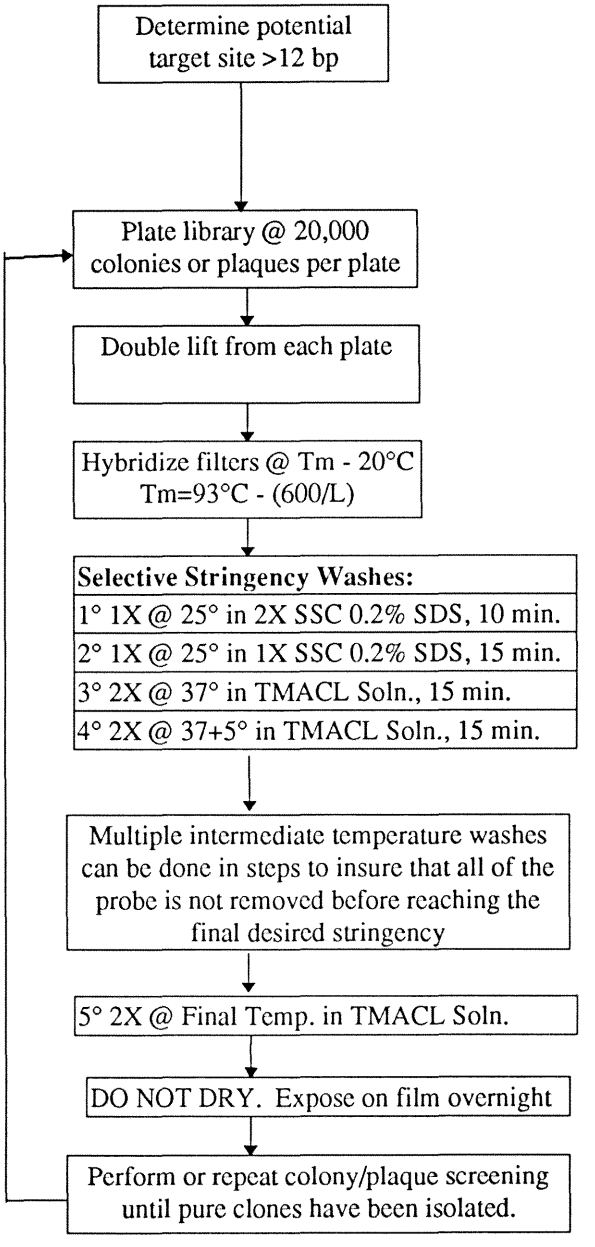

50
45
40
35
30
25
20
15
10
5
0
Bases Mismatched
70 nt
65 nt
60 nt
55 nt
50 nt
45 nt
40 nt
35 nt
30 nt
25 nt
20 nt
15 nt
12 nt
10 nt
20°C 25°C 30°C 35°C 40°C 45°C 50°C 55°C 60°C 65°C 70°C 75°C 80°C 85°C
T MACL Wash T emperature

Fig. 3 *Screening procedure flowchart* During the wash steps, the detection of lower levels of homology is possible by autoradiography of the filters after intermediate temperature washes. It is important that the filters are not allowed to dry during any of these steps. Libraries should be plated at 20,000 colonies or plaques per plate. For genomic libraries 2- 3 genomic equivalents should be used for determining the number of copies per genome.


Determine potential target site >12 bp
Plate library @ 20,000 colonies or plaques per plate
Double lift from each plate
Hybridize filters @ Tm - 20°C Tm=93°C - (600/L)
Selective Stringency Washes:
1° 1X @ 25° in 2X SSC 0.2% SDS, 10 min.
2° 1X @ 25° in 1X SSC 0.2% SDS, 15 min.
3° 2X @ 37° in TMACL Soln., 15 min.
4° 2X @ 37+5° in TMACL Soln., 15 min.
Multiple intermediate temperature washes can be done in steps to insure that all of the probe is not removed before reaching the final desired stringency
5° 2X @ Final Temp. in TMACL Soln.
DO NOT DRY. Expose on film overnight
Perform or repeat colony/plaque screening until pure clones have been isolated.

Chapter 2

Repeated sequence target sites for maternal DNA-binding proteins in genes activated in early sea urchin development

Roger F. Anderson, Roy J. Britten and Eric H. Davidson

DEVELOPMENTAL BIOLOGY 163, 11-18 (1994)

Repeated Sequence Target Sites for Maternal DNA-Binding Proteins in Genes Activated in Early Sea Urchin Development

ROGER ANDERSON, ROY J. BRITTEN, AND ERIC H. DAVIDSON

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

Accepted January 26, 1994

This communication concerns a very highly conserved inverted repeat sequence element that serves as a target site for a sea urchin maternal DNA-binding factor. The maternal factor is present in relatively large amounts in unfertilized eggs, but is about 100× less prevalent per embryo in 24-hr embryo nuclear extract. The inverted repeat target site is found in the regulatory domain of the *CyIIIa* cytoskeletal actin gene and also in two upstream genes encoding transcription factors that bind to a functionally important *cis*-regulatory element of the *CyIIIa* gene. There are about 460 copies of the inverted repeat target site per genome. About 15% of these sites occur in a nested arrangement together with a second inverted repeat that binds another previously characterized maternal transcription factor. This arrangement is the same as that which occurs in the *CyIIIa* gene, and it may be of regulatory significance with respect to activation of certain genes in oogenesis and early embryogenesis.

INTRODUCTION

This communication concerns a highly conserved inverted repeat sequence which occurs several hundred times in the genome of *Strongylocentrotus purpuratus*. This sequence motif is likely to be of functional importance for early development because, as we show below, the repeat contains specific target sites for a maternal DNA-binding protein. In the following, we provisionally refer to this DNA-binding protein(s) as **P10**. Inverted repeat sequence elements are often observed to serve as target sites for transcription factors, which interact with them as facing homotypic or heterotypic pairs [for example, in the target sites for the yeast $\mathbf{mat}_{\alpha}$ and **MCM1** factors (Johnson and Smith, 1992) and sites for regulators of the HLH (e.g., Ellis *et al.*, 1990) and leucine zipper (Cao *et al.*, 1991) classes]. In an experimental study of the transcription factors required to establish the territorial expression of the *CyIIIa* cytoskeletal actin gene in the early *S. purpuratus* embryo, we encountered an inverted repeat target site for a tightly binding protein factor, initially referred to as "**P6**" (Calzone *et al.*, 1988; Thézé *et al.*, 1990). This factor, now termed **SpZ12–1**, contains 12 sequence elements very similar to canonical Zn fingers. As will be reported elsewhere, **SpZ12–1** is present in egg cytoplasm and embryo nuclei, and its interaction with its target sites in the regulatory domain of the *CyIIIa* gene is probably required during embryogenesis for spatially accurate expression (unpublished data of G.-W. Wang). In the *CyIIIa* gene, the two halves of the newly discovered inverted repeat which is our present subject, i.e., the sequence that includes the **P10** target sites, immediately flank on either side the inverted repeat to which two **SpZ12–1** molecules bind. We show that the same unique arrangement of these two different inverted repeat factor-binding sites occurs a number of times in the genome, though each inverted repeat also occurs without the other. Curiously, **P10** target sites are present not only in the *CyIIIa* regulatory domain, but also in the vicinity of two upstream genes encoding transcription factors (Calzone *et al.*, 1991; Höög *et al.*, 1991) that recognize *different regulatory sites in the CyIIIa gene.* Interactions at these other sites are also required for proper *CyIIIa* expression. These two regulatory genes, and the *CyIIIa* gene itself, are all activated in oogenesis and are expressed in embryogenesis. **P10** may thus serve as a maternal regulator of a functionally related network of genes required early in the sea urchin life cycle.

MATERIALS AND METHODS

Extracts

24-hr embryo nuclear extracts were prepared as described previously (Calzone *et al.*, 1988). Egg extracts were prepared from a postmitochondrial fraction of egg cytoplasm, as will be described in detail elsewhere (F. Calzone, unpublished data). Ovary and testis extracts were prepared by a modification of the procedure of Calzone *et al.* (1988): Gonads were removed from adults and diced with a razor blade. Mature gametes were re-

0012-1606/94 $6.00

moved by suspending the gonad fragments in 2× vol filtered seawater and centrifuging them at full speed in a tabletop centrifuge for 5 min. This was repeated until the supernatant was clear of gametes when viewed under a dissecting microscope. The gonads were then twice washed in 5 vol of ice-cold 1 *M* glucose and spun 5 min at full speed in a tabletop centrifuge. The pellet was resuspended in 10 vol of ice-cold buffer consisting of 10 m*M* Tris-HCl, pH 7.4, 1 m*M* EDTA, 1 m*M* EGTA, 1 m*M* spermidine, 0.36 *M* sucrose, 1 m*M* DTT, and frozen in liquid nitrogen. Nuclear extracts were later prepared as previously described (Calzone *et al.*, 1991), with the exception that the ovary extract was not fractionated with 4 *M* $(NH_4)_2SO_4$ as were the other nuclear extracts.

Site Frequency Determinations

Genomic site frequencies were obtained by screening an EMBL3 genomic library, as follows: Probes used were, for the **P10** site, GTAATGTAACCAAATTTAATGTATATGTC, and for the **SpZn12–1** site, TGTTGCTAGGTAGGTCAAGC. Probes were labeled with T4 polynucleotide kinase, and hybridizations were carried out at 42°C for the **P10** probe and 32°C for the **Zn12–1** probe, in 0.9 *M* NaCl, 0.012 *M* EDTA, 0.18 *M* Tris, pH 8.0, 5× Denhardt's solution, 50 m*M* Na-phosphate, pH 7.4, 0.25% SDS, and 100 μg/ml sheared denatured calf thymus DNA. The following washes were then carried out: 15 min at 25°C in 2× SSC, 0.2% SDS; 15 min at 25°C in 1× SSC, 0.2% SDS; and 2× for 30 min in 3 *M* Me_4NCl (tetramethylammonium chloride), 50 m*M* Tris (pH 8.0), 2 m*M* EDTA, 0.1% SDS, at 32°C. For the filters probed with the **Zn12–1** oligonucleotide, two additional rinses in the Me_4NCl mix for 30 min at 37°C were performed, and for the filters probed with the **P10** oligonucleotide two washes in the Me_4NCl mix were carried out at 54°C. These final washes allow for a 4- to 5-bp mismatch with respect to each probe. For each screen two sets of Amersham Hybond N+ nylon filter disks were used to lift plaques off the plates.

RESULTS

*Occurrence of the **P10** Inverted Repeat in Three Known Genes*

The observation that initially focused our interest on the inverted repeat sequence here termed the "**P10** target site" was its occurrence in the vicinity of three genes that are likely to be functionally related. The first of these genes is the *CyIIIa* gene, which encodes a cytoskeletal actin. The expression of this downstream gene is strictly confined to the aboral ectoderm of *S. purpuratus* embryos (Cox *et al.*, 1986). The second of these genes is a regulatory gene, *SpP3A2*, which encodes a transcription factor of novel sequence, at least two target sites for which exist in the regulatory domain of the *CyIIIa* gene (Thézé *et al.*, 1990; Calzone *et al.*, 1991). The third is another regulatory gene, *SpZ2-1*, (formerly *P3A1*), which encodes a second transcription factor that is apparently a 2-Zn finger protein and that recognizes almost exactly the same target sites as does the **P3A2** protein (Calzone *et al.*, 1991; Höög *et al.*, 1991). Interference with factor interactions at these "P3A" target sites, by *in vivo* competition (Hough-Evans *et al.*, 1990; Franks *et al.*, 1990) or by site deletion (unpublished data of C. Kirchhamer), has been shown to result in ectopic expression of *CyIIIa* fusion gene constructs. Either or both of the products of these two regulatory genes are thus probably required for spatial regulation of embryonic *CyIIIa* expression (Calzone *et al.*, 1991; Höög *et al.*, 1991).

The inverted **P10** repeat target site was first noticed in scanning DNA sequences obtained upstream of the *SpZ2-1* gene, and a computer search revealed almost exactly the same inverted repeat within the regulatory domain of the *CyIIIa* gene. According to genome blots (not shown) this same sequence element is also present as a dispersed repeat in the DNAs of two other *Strongylocentrotid* species, but is not detectable in *Lytechinus* DNA. These particular sequences had escaped detection as protein-binding sites in our earlier analyses of the *CyIIIa* gene (Calzone *et al.*, 1988; Thézé *et al.*, 1990), for the reason that they probably function during oogenesis, rather than in embryogenesis, as shown below. In the *CyIIIa* gene the two elements of the **P10** inverted repeat target site occur in a particularly interesting arrangement. Thus they lie immediately external to another inverted repeat which is a target site for the known *CyIIIa* regulatory factor **SpZ12–1** (see above). The location of the **P10** sequences upstream of the *CyIIIa* and of the *SpZ2-1* genes is shown diagrammatically in Fig. 1a. A subsequent search of sequences in the vicinity of the *SpP3A2* gene then revealed contiguous **SpZ12–1** and **P10** sites in the 3′ region of this transcription unit as well (Fig. 1a). Two additional sequence features that the two regulatory genes included in Fig. 1a share with the *CyIIIa* gene are that the *SpZ2-1* gene also contains a canonical target site for its own product (or that of *SpP3A2*) and that the *SpP3A2* gene also contains a target site for the *SpZ12-1* factor.

The unusually high degree of similarity among the **P10** target site sequences discussed here is shown explicitly in the alignment given in Fig. 1b. It can be seen that in the seven copies of the **P10** site shown from the three genes analyzed, 46 of 66 nucleotides (nt) are *identical* in the shared region of the sequence, and at an additional 11 positions the sequence is identical in 6 of 7 occurrences considered. The two halves of each inverted repeat are about as similar as the repeats from different

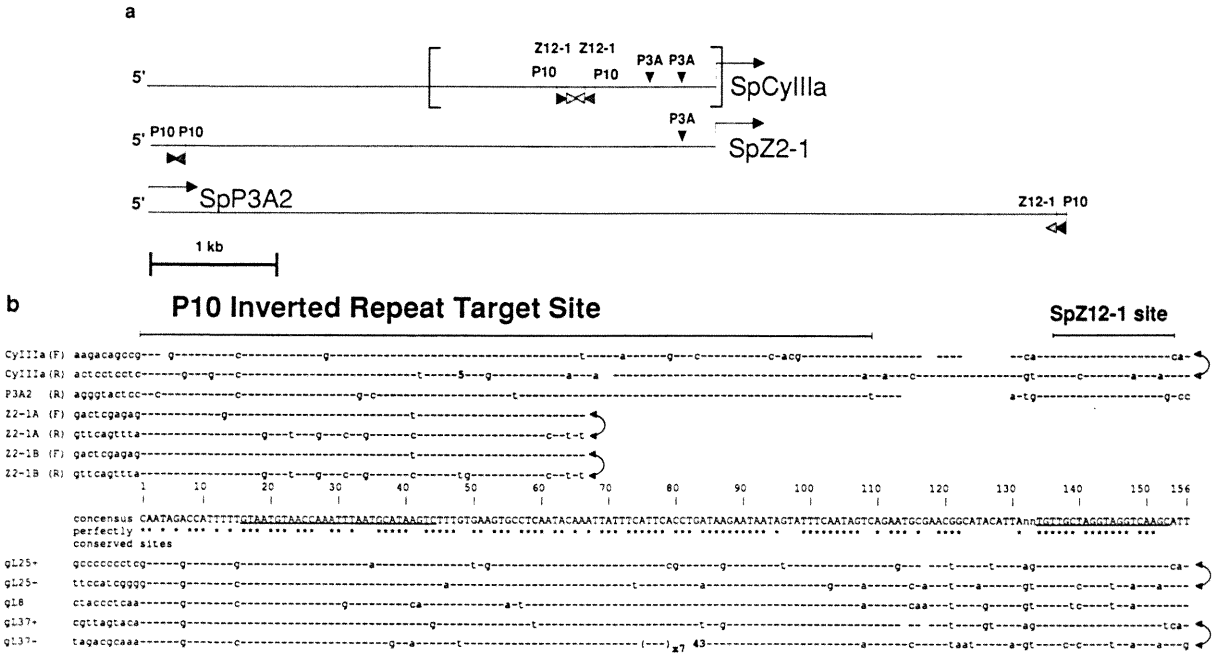


FIG. 1. Location and alignment of **P10** inverted repeat target sites in the *CyIIIa*, *SpP3A2*, and *SpZ2-1* genes. (a) Location of **P10** sites with reference to origin of transcription. The necessary and sufficient regulatory region of the *CyIIIa* gene, indicated by brackets, is known to include at least 50 target sites for DNA-binding factors, serviced by about 10 such factors. These have all been partially characterized and some are now cloned (Thézé *et al.*, 1990; Coffman and Davidson, 1992; Calzone *et al.*, 1988, 1990; unpublished data). Target sites for the **P10** binding activity (see text and b) are shown as horizontal solid triangles, indicating the orientation of each repeat half unit. Transcription start sites are indicated as bent arrows. The intron/exon structures of the *SpZ2-1* and *SpP3A2* genes are not known, nor have the *cis*-regulatory regions of these genes been functionally defined. Target sites for the **SpZ12-1** transcription factor (Calzone *et al.*, 1988; Thézé *et al.*, 1990) are also shown, as open horizontal triangles, and the target sites of the **SpZ2-1** and **SpP3A2** transcription factors (Calzone *et al.*, 1991; Höög *et al.*, 1991), indicated as "P3A sites," are shown as vertical triangles. The P3A sites of the *CyIIIa* gene have been experimentally shown to bind either factor (Höög *et al.*, 1991), while the site shown upstream of the *SpZ2-1* gene is here identified merely by sequence (its presence, however, suggests the possibility of autoregulation). (b) Sequence alignments. The first seven sequences shown are the **P10** and **SpZ12-1** sites indicated in (a), followed by their obvious consensus. Four examples, i.e., two inverted pairs, of the **P10** sites upstream of the gene encoding **SpZ2-1** are shown, representing two different alleles that were sequenced. In the *CyIIIa* sequence shown (F), nucleotide 1 of the concensus corresponds to nucleotide 1153 of the regulatory domain of the *CyIIIa* gene (Thézé *et al.*, 1990). The five sequences shown below the consensus are derived from genomic clones obtained by direct target site screening, as described in the text. The two underlined regions below the consensus indicate the sequences of oligonucleotides used for target site screening. Symbolism: A dash indicates a match to the consensus; a space indicates a gap introduced for alignment purposes; a bold number indicates the presence of inserted nonmatching bases with respect to the consensus; and lower case letters indicate a specific difference from the consensus at that site. An F or R after the clone name indicates that the sequence is Forward, reading 5′ to 3′, or Reversed, 3′ to 5′, relative to the transcribed region (the R sequences shown are the complementary strand to that represented in the F sequences). For the genomic isolates + and − designations are used to distinguish between the inverted copies of the repeat. Note gL8 only has one copy of the repeat.

genes are to one another. Figure 1b also shows that the inverted repeat sequences of the *CyIIIa* and *SpP3A2* genes are 40–60 nt longer than the minimum shared length. These additional regions (nt 67 to 130 of consensus in Fig. 1b) are equally complementary to one another, as the shorter sequences held in common in all seven versions shown. At the right-hand end of Fig. 1b can be seen the inverted repeats that constitute the two *SpZ12-1* sites of *CyIIIa* located within the **P10** repeats and the single aligned *SpZ12-1* site in the *SpP3A2* gene. As can be seen in Fig. 1b, the homology ends abruptly at the same location at the external termini of the **P10** repeat (i.e., the left-hand side of the alignment shown in the figure).

The Shared Inverted Repeat Is Recognized by a Maternal DNA-Binding Factor

Figure 2a displays gel-shift experiments that indicate a specific reaction between proteins present in an egg cytoplasmic extract and a probe consisting of the inverted repeats of the *SpZ2-1* gene (see Fig. 1b). The bind-

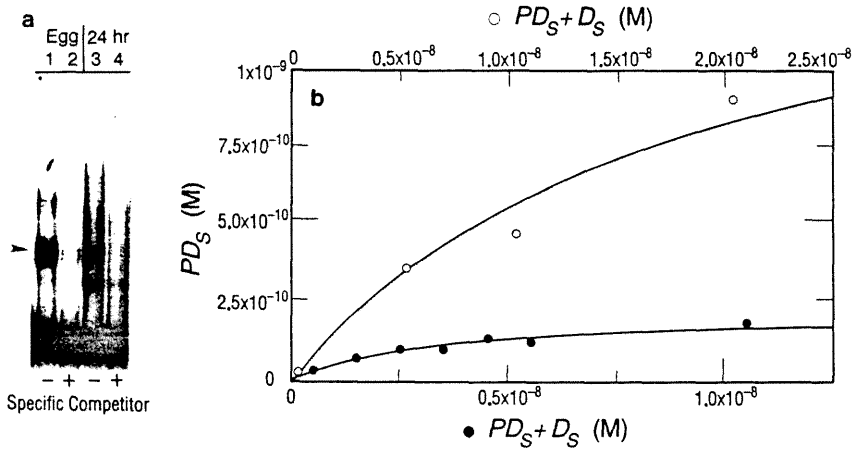


FIG. 2. **P10** DNA-binding activity in egg cytoplasm and blastula stage embryo nuclei. (a) Detection of **P10**-DNA complexes by gel shift. The probe was a DNA fragment consisting of the *SpZ2-1* (A) pair of inverted repeat elements shown in Fig. 1b. The probe was present at 0.1 ng/reaction. For the experiments shown in lanes 2 and 4, 80 ng of the same target site as was used for the probe, but unlabeled, were added as specific competitor. All reactions also contained 3 μg poly(dA:dT) as nonspecific competitor. Gel shifts were carried out as described earlier (Calzone *et al.*, 1988) and the same results were obtained with different extract preparations (not shown). The arrow indicates the specifically competed **P10** complex (the lower band seen with the nuclear extract indicates a nonspecific, general DNA-binding protein(s) that cannot be specifically competed). Lanes 1 and 2 show reaction of a cytosol egg cytoplasm extract and lanes 3 and 4 the reaction of our standard 24-hr embryo nuclear extract (Calzone *et al.*, 1988). No **P10** activity was seen in an egg cytoplasm fraction that consists of proteins recovered from an RNP particle fraction (not shown). (b) Estimation of relative binding affinity and quantity of active **P10** factor by quantitative gel-shift titrations. The procedure and analysis is described by Calzone *et al.* (1988). Molar quantities of specific **P10**-DNA complex ($\mathbf{PD}_s$) are shown as a fraction of amount of specific DNA ($\mathbf{D}_s$) present, i.e., the sum of labeled (probe) and unlabeled (competitor) DNA fragments in the reaction. Data were fit by least squares to the function, $\mathbf{PD}_s = K_r P_0 D_s/(1 + K_r D_s)$, where P_0 is the molar quantity of active **P10** factor in the reaction and K_r is the ratio of equilibrium constants for the reaction of **P10** with probe DNA, and with poly(dA:dT) (Emerson *et al.*, 1985; Calzone *et al.*, 1988).

ing activity was confined to a cytosol fraction of egg cytoplasm and was absent from a ribonucleoprotein fraction (data not shown). The competition experiment shown in lane 2 of Fig. 2a indicates the sequence specificity of the reaction between the maternal protein(s) and this probe (see legend, and Table 1). DNA-protein complexes of the same mobility are formed by protein(s) present in 24-hr embryo nuclear extract, as shown in lanes 3 and 4 of Fig. 2a. This suggests that the maternal **P10** activity and the embryo nuclear activity could be the same protein, although as shown below its properties are different in the embryo. Other experiments (not shown) demonstrate that similar complexes are formed with this probe by protein(s) present in pluteus stage (72-hr) embryo nuclear extract and in nuclear extracts of both ovary and testis.

TABLE 1
PREVALENCE AND RELATIVE BINDING CONSTANT FOR P10 FROM EGG CYTOPLASM AND EMBRYO NUCLEI

Extract	K_r	P'_0 (molec/egg or embryo)
Egg cytoplasm	1.4×10^5	2.3×10^7
Egg cytoplasm	1.1×10^5	1.8×10^7
Egg cytoplasm	2.7×10^5	1.2×10^7
Blastula nuclei	1.9×10^4	1.4×10^5
Blastula nuclei	4.3×10^4	6.9×10^4

Note. Data are from several different experiments, such as those shown in Fig. 2b (see legend for details and definition of K_r). P'_0 is calculated from the molar concentrations of P10 per reaction determined as in Fig. 2b by application of the number of eggs or embryos represented per microliter of extract.

To measure the relative specificity of the interaction between the egg and embryo protein(s) and the inverted repeat probe, we carried out a series of quantitative gel-shift competitions by methods described previously (Calzone *et al.*, 1988), as illustrated in Fig. 2b. These experiments also provided an estimate of the prevalence in egg cytoplasm and embryo nuclei of the specific **P10** DNA binding protein(s). These data are presented as titrations of the factor present in the extracts as the concentration of the DNA target sequence (i.e., labeled probe plus unlabeled competitor) is increased. Results are shown in Table 1. The egg cytoplasm factor displays a fairly high relative specific affinity, indicated by the K_r value of about $1\text{-}3 \times 10^5$, typical of many of the transcription factors we have studied (Calzone *et al.*, 1991). The **P10** factor(s) is also reasonably prevalent in the egg: If one were to imagine the egg cytoplasm divided into about 500 cells, as at the blastula stage, there would

be ~40,000 molecules/average cell. However, Table 1 shows that only about 200 molecules of active factor are recovered per average nucleus in the blastula stage nuclear extract, or ~0.5% of what was in the egg cytoplasm. Of course the factor could have been present in the nuclei but have been preferentially lost from the nuclear extract, although this is not usually observed with other transcription factors that we have studied (Harrington *et al.*, 1992; Calzone *et al.*, 1988; Livant *et al.*, 1988; our unpublished data). In any case, the activity of the factor in the nuclear extract has apparently changed. The K_r value for **P10** in the embryo nuclear extract is less than one-fifth of its value for the maternal form, which is obviously a greater difference than can be accounted for as experimental error (Table 1). However, the point is moot, since unless the factor is sharply localized in the embryo there is probably not enough of it per nucleus to cause significant occupancy of its target site (e.g., see calculations in Calzone *et al.*, 1988). Thus, **P10** is clearly *a maternal factor*, which may operate in oocytes or early embryos.

TABLE 2
RECOVERY OF P10 AND SpZ12-1 TARGET SITES BY DIRECT LIBRARY SCREENING

Initial probe	+/Plaques screened		(%)	+ Sites/G.e.
P10 site	46/5,300		(.87)	460
Z12-1	60/10,000		(.60)	318
Counter-screen	Probe	++ Plaques	++/Ge.	(%)	++/+ Plaques (%)
P10 plaques	Z12-1 site	8	80	(.15)	(17.4)
Z12-1 plaques	P10 site	7	37	(.07)	(11.9)

Note. An EMBL3 genome library was screened under criterion-controlled conditions with the oligonucleotide probes indicated in Fig. 1b, as described under Materials and Methods. The genome size of *S. purpuratus* is 8×10^8 bp, and at an average insert size of 15 kb; 1 genome equivalent (G.e.) is about 5.3×10^4 plaques. Plaques scored as positive (+) were observed on double lifts and then verified on isolation and rescreening. Of initial positive plaques recovered, 89 and 73% of the P10 and Z12-1 site screens, respectively, survived rescreening and purification.

P10 Target Sites in the Genome and Their Coincidence with SpZ12-1 Target Sites

The inverted repeat target sites for the **P10** factor, as well as for the **SpZ12-1** factor shown in Fig. 1b are sufficiently long and invariant as to suggest that their prevalence could be directly measured in the genomic DNA. To this end we constructed oligonucleotide probes representing maximally conserved regions and utilized them at a carefully controlled criterion to screen a genomic library (see Materials and Methods). Results are collated in Table 2. Relevant regions of a number of these isolates were sequenced, and, as shown in the lower portion of Fig. 1b, in each case a *bona fide* example of a **P10** target site was in fact recovered. Independent screens were carried out for **P10** sites, and for **SpZ12-1** sites, and the set of isolates from each screen was then counterscreened for the other site, to determine the frequency with which both occur together, as in the *CyIIIa* and *SpP3A2* genes. Sequences from some examples of these "double-positive" isolates are included in Fig. 1b, where it can be seen that the **Z12-1** target-site sequences recovered are also extremely similar to those found earlier.

Table 2 shows that there are per genome about 460 **P10** sites (i.e., inverted repeat sites or single sites) that fall within the range of similarity to the probe sequences allowed by the screening criterion and about 320 **SpZ12-1** target sites. Two estimates of the fraction of these isolates that contains *both* sites are shown, from which it can be concluded that there are 40–80 occurrences of the **P10:SpZ12-1** double site per haploid genome. This is about 30-fold more than would occur if these sites were randomly assorted: from Table 2 the fraction of genomic λ clones that would be expected to contain both sites would be about 5×10^{-5} rather than the $\sim 1.5 \times 10^{-3}$ observed. Note that these values refer only to the highly conserved sites allowed by the screening conditions. Larger numbers of double-positive isolates might have been obtained at more relaxed criteria. Indeed, the ancestral inverted repeat could have contained both **P10** and **SpZ12-1** target sequence elements. As shown in Fig. 3, the organization of the double sites is anything but random. This figure displays a PCR estimate of the sequence length separating the **P10** site probe sequence and the **SpZ12-1** site probe sequence on 11 of the double site isolates. In every case the lengths are within a few nucleotides of the lengths observed in the *CyIIIa* and *SpP3A2* genes, as shown in the right lanes of Fig. 3. Sequences for several of the same genomic **P10:SpZ12-1** sites as used for the PCR experiment are shown in Fig. 1b. We also screened for **P10** target sites in a late gastrula cDNA library. Relevant regions of 12 of these isolates were examined (data not shown) and found to contain the expected **P10** and **SpZ12-1** target site sequence elements, in either 5′ or 3′ noncoding regions of the transcripts. The significance of this observation is that it adds evidence that the **P10** target site marks transcription units expressed in the embryo. We note in this connection that >90% of embryo transcripts of gastrula stage are also represented in maternal poly(A) RNA (Galau *et al.*, 1976). It is interesting and perhaps significant that 4 of the 12 of these cDNA

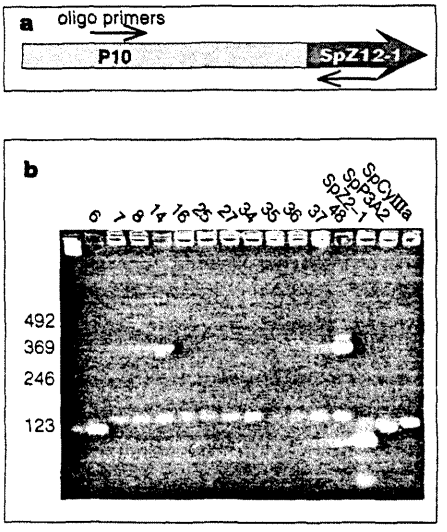


FIG. 3. Detection of similar **P10:SpZ12-1** target site sequences in genomic isolates by PCR. (a) Location of PCR primers in the **P10:SpZ12-1** target site sequence. The specific primers used were as indicated in Fig. 1b (the same as used for the library screening described in text). Vent polymerase from NEB was used according to the following PCR protocol, viz, denaturation at 96°C for 5 min, 35 amplification cycles of 96°C, 40 sec, 65°C 35 sec, 1 min ramp to 72°C, and 72°C for 1 min. A 10-min 72°C incubation followed after the 35 cycles to ensure that all extensions were complete. The samples were then maintained at 4°C until they were loaded onto a 4% agarose gel for size determination. The bands in (b) are visualized by ethidium bromide staining. (b) Amplification products shown are from 12 genomic isolates (6–48) and from the known *SpZ2-1*, *SpP3A2*, and *CyIIIa* sequences. The exact sequences of the amplified regions from the three latter genes are shown in Fig. 1b. Note that the *SpZ2-1* gene sequence, which lacks the internal sites for the **SpZ12-1** factors, produces a smaller band, of the size expected for amplification of the inverted **P10** repeat sequence alone. It can be seen that the bands produced by the unknown genomic isolates are very close to the length separating the primers in the **P10:SpZ12-1** target site sequences of *CyIIIa* and *SpP3A2*. Primary sequences verifying this interpretation are shown for genomic clones 8, 25, 27, and 37 in Fig. 1b.

clones displayed *both* **SpZ12–1** sites and **P10** sites, in the usual arrangement, and 4 others displayed somewhat degenerate **SpZ12–1** sites, also in the appropriate sequence locations. This may be compared to ~15% incidence of high-fidelity double sites in the genomic isolates. Thus, double sites may occur more frequently in the immediate vicinity of embryonically expressed genes (such as *CyIIIa*).

DISCUSSION

Inferences with Respect to Maternal Function of ***P10***

The distribution of the tightly conserved inverted repeat **P10** target sites is particularly interesting when considered in light of the observation that these sites interact with a *specific* maternal DNA-binding factor(s). The prevalence of the **P10** factor(s) in egg cytoplasm lies within the range observed for other transcription factors that interact with *CyIIIa* regulatory sites (unpublished data of F. Calzone). However, **P10** is unusual in that only 0.5% of the activity measured in egg cytoplasm is later recoverable from blastula stage nuclear extracts. This, of course, could be an artifact, due to sensitivity of **P10** to extraction conditions, or its exclusion from these extracts, a point that will be resolvable only when this factor(s) is cloned and antibody probes are available. However, we also show that the **P10** present in nuclear extracts at blastula stage has been partially inactivated, so that the stability of its interaction with target sites appears at least 4× lower than with the maternal form. Whatever its state in later embryos, it seems reasonable to assume that **P10** is indeed a maternally active factor or factors. This would imply that it is present in active form at appreciable concentrations in oocytes (and early embryo) nuclei. On these assumptions its role may be illuminated by the developmental pattern of expression of the three known genes in which we observe **P10** target sequences.

With respect to function, these genes share two notable characteristics: *CyIIIa*, *SpP3A2*, and *SpZ2-1* probably belong to the same regulatory network, as discussed earlier, and all three genes are in fact activated during oogenesis. Earlier studies show that there are 1000–2000 molecules of maternal mRNA per egg for all three genes at fertilization (Lee *et al.*, 1986; Cutting *et al.*, 1990). Both **SpP3A2** and **SpZ2–1** proteins are present in the egg as well (unpublished data of R. Zeller). After fertilization, however, the patterns of expression of these three genes differ sharply. **SpP3A2** and **SpZ2–1** mRNA levels remain low, within a factor of two to three of their maternal levels on a per embryo basis (Cutting *et al.*, 1990). The *CyIIIa* gene is transcriptionally quiescent after fertilization, until about 10 hr into cleavage. It is then abruptly reactivated and thereafter is transcribed at a fairly high rate but only in the aboral ectoderm territory of the embryo. In aboral ectoderm *CyIIIa* message accumulates to a steady state of about 200 copies per cell (Lee *et al.*, 1986). **P10** could function in some way as an element in the mechanism by which these three genes are mobilized for the initial onset of their expression during oogenesis and/or early embryogenesis. The unique and easily recognizable **P10** target sites might thus mark members of a particular battery of genes (Britten and Davidson, 1969) that are activated during oogenesis and utilized after fertilization.

The ***P10*** *Gene Battery in Oogenesis*

The measurements of Table 2 provide an estimate of the number of members that such a battery might in-

clude, or at least of those members that utilize very well conserved **P10** target sites. This estimate might be too small because our criterion screening standards might have excluded more divergent **P10** sites that are nonetheless functional, or it might be too large if it includes conserved **P10** sites that are unconnected with functional genes. Table 2 indicates about 460 conserved **P10** target sites per genome. However, this number represents only about 4% of the number of different maternal mRNA species estimated from complexity measurements to be present on early embryo polysomal mRNA (reviewed by Davidson, 1986, pp. 71–74). This argument prompts the obvious queries as to what sort of subset of maternally expressed genes a **P10** regulatory system might service, and what maternal role might such a system play.

It is tempting to consider the conserved **P10** sites as markers of a particular regulatory network of genes of which the *SpP3A2* and *SpZ2-1* transcription factors and the *CyIIIa* cytoskeletal gene are all members. The function of interaction at these sites might be, for example, to alter chromatin conformation in adjacent regulatory domains and thereby to potentiate current or later interactions of other specific transcription factors that will in the end differentially regulate these genes. For example, **P10** binding might be required for the *CyIIIa* regulatory system to be accessible to its early embryonic factors. Furthermore, a result of such "preactivation" might be that all genes so affected might be transcribed to some extent in oogenesis. This could account for an old mystery, viz, that almost all genes expressed in particular ways during embryogenesis are also represented by low levels of maternal mRNA (Galau *et al.*, 1976; reviewed in Davidson, 1986, pp. 165–170).

*Relation of **P10** Sites with **SpZ12-1** Sites*

The **SpZ12-1** factor is evidently utilized later in embryogenesis in various ways. For example, we know that its interaction with *CyIIIa* target sites is specifically required to prevent expression in mesenchyme cells, and late in embryogenesis the gene encoding this factor is intensely expressed, but only in certain embryonic territories (G-W. Wang, unpublished data). However, **SpZ12-1** is also a maternal factor, and like **P10** (this work) it is found in testis nuclei (F. Calzone, unpublished data) as well as in eggs. Around 15% of occurrences of the inverted **SpZ12-1** target sites in the genome are immediately flanked by **P10** target sites (Table 2; Fig. 3). Thus, though **SpZ12-1** clearly has an embryonic role while **P10** may not function after the earliest stages of embryogenesis, these factors may function together in a subset of genes during oogenesis. The organization of these sites (Fig. 2) suggests alternative forms of interaction: Thus, in *CyIIIa* and many other genes (Fig. 3) the arrangement would be **P10**$_{\rightarrow}$:**Z12-1**$_{\rightarrow}$:$_{\leftarrow}$**Z12-1**:$_{\leftarrow}$**P10**, although in the later embryo only the **Z12-1**$_{\rightarrow}$:$_{\leftarrow}$**Z12-1** interaction is functional in the *CyIIIa* gene. In other genes, such as *SpZ2-1*, the interaction would be confined to **P10**$_{\rightarrow}$:$_{\leftarrow}$**P10**. Thus, there must be three classes of protein:protein interaction that could occur at different times and in different regions of the organism among these factors, i.e., **P10:P10**,**Z12-1:Z12-1**, and **Z12-1:P10**. The precise organization of these DNA target sites thereby provides for several alternative potential regulatory states given the same two factors.

We thank Dr. Frank J. Calzone for providing egg and embryo extracts and Ms. Mary Chen for technical assistance. We are most appreciative of the efforts of Professor Carl Parker of this Institute who reviewed a draft of the manuscript. Research was supported by NIH Grant HD-05753. R.A. was supported by NIH Training Grant GM-07616.

REFERENCES

Britten, R. J., and Davidson, E. H. (1969). Gene regulation for higher cells: A theory. *Science* **165**, 349–358.

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., 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.

Cao, Z., Umek, R. M., and McKnight, S. L. (1991). Regulated expression of three C/EBP isoforms during adipose conversion of 3T3-L1 cells. *Genes Dev.* **5**, 1538–1552.

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

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., 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. Natl. Acad. Sci. USA* **87**, 7953–7957.

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

Ellis, H. M., Spann, D. R., and Posakony, J. W. (1990). *extramachrochaetae*, a negative regulator of sensory organ development in *Drosophila*, defines a new class of helix-loop-helix proteins. *Cell* **61**, 27–38.

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. *Cell* **41**, 21–30.

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.

Galau, G. A., Klein, W. H., Davis, M. M., Wold, B. J., Britten, R. J., and Davidson, E. H. (1976). Structural gene sets active in embryos and adult tissues of the sea urchin. *Cell* **7**, 487-505.

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., 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., 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.

Johnson, A. D., and Smith, D. L. (1992). A molecular mechanism for combinatorial control in yeast: MCM1 protein sets the spacing and orientation of the homeodomains of an $\alpha 2$ dimer. *Cell* **68**, 133-142.

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

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.

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.

Chapter 3

Identification, Isolation and Cloning of the P10 site binding factor : Replication Factor A

Roger F. Anderson, James G. Moore, Michael G. Harrington, James A. Coffman, Roy J. Britten, and Eric H. Davidson

Abstract

Previous experiments identified a protein binding site in a repeat element that is present in the *Strogylocentrotus purpuratus* genome > 400 times, provisionally labeled the P10 site. We purified the P10 site binding factor using a specific sequence double-stranded DNA affinity column. Cloning of a cDNA, by use of amino acid sequence from the protein responsible for the gel-retardation of the P10 site, enabled us to identify this factor as the sea urchin homologue of the replication factor A 70 kD subunit (SpRFA-70).

Replication factor A has been shown to have a number of activities and interactions as a component of DNA replication. The most prominent and well-characterized function is the single-stranded DNA (ssDNA) binding capacity of the 70 kD subunit. The non-specific affinity of RFA-70 for random single-stranded DNA has been documented to be significantly higher than its non-specific affinity for random double-stranded DNA (dsDNA) sequence. We show that SpRFA-70 has a relative affinity for the specific P10 dsDNA site that is much higher than its affinity for random dsDNA. The equilibrium constant for SpRFA-70 binding to ssDNA is 1.7×10^{10} and the K_{eq} for binding to the P10 site dsDNA is $1.8X10^{9}$ as determined by protein excess titration.

Introduction

This study was undertaken to clone and describe a potential transcription factor identified in our earlier work. The factor was first characterized by gel-shift activity as recognizing a specific sequence site in double-stranded DNA (dsDNA). This specific site was found flanking several genes. Two features of this site made it very interesting, first its location; part of an inverted repeat and second, the 400+ occurrences of this site in the *Strogylocentrotus purpuratus* genome. We describe the cloning of the message encoding the binding factor, and the characterization of the binding activities of this protein. The translated sequence of the isolated clone revealed the protein product to be a sea urchin homologue of the 70 kD subunit of replication factor A (RFA 70). RFA is a three subunit complex that has been previously described as a single-stranded DNA (ssDNA) binding protein with a relatively low affinity for dsDNA, which is essential for DNA replication (Kim *et al.*, 1992 and refs. therein). The activities of the whole protein are cell cycle regulated (Carmichael *et al.*, 1993) and in *Xenopus* oocytes antibody staining of RFA marks the sites where DNA replication will begin after fertilization (Adachi and Laemmli 1992). However, RFA has been shown to have other roles in yeast (Luche *et al.*, 1993) and in this work, based on our findings, we demonstrate that this may be true in sea urchins as well.

We have not yet determined the function of RFA with respect to the gene regulatory regions to which it binds. We have demonstrated the presence of a high degree of conservation of the specific binding site. We have shown that there is a

high prevalence of "active" factor in the egg. We also show that RFA has a double stranded DNA site specific binding activity and we measured the equilibrium constants for binding to both double- and single-stranded DNA. We confirm the capacity of RFA to bind dsDNA in a site specific manner as described for the RFA 70kD subunit of yeast (Luche *et al.,* 1992). We note that RFA has been shown to interact with VP16, p53 and GAL4 (Li and Botchan, 1993, He, *et al.,* 1993) which indicates that it also has a capacity to function with other DNA binding factors. Thus SpRFA may operate as both a replication factor and gene expression regulatory factor.

Materials and Methods

Affinity Column Purification of Proteins: Egg extract and 24 hour embryo nuclear extract were prepared as described in Calzone *et al.,* 1988. The extracts were each applied to double-stranded DNA affinity columns to purify the unknown factor from the respective extracts. The procedure detailed in Thézé *et al.,* 1990 was followed with the following exceptions. (1)DNA was bound to Pierce 3M EMPHAZE™ beads using their protocol. This resin also has a higher capacity for attachment and the efficiency is much greater than afforded by the cyanogen bromide method and sepharose resin. The binding reaction was quenched with MOPS buffer pH 7.9. (2)The elution of the proteins from the affinity column was accomplished with washes up to 2 M KCl Buffer. (3) A single-stranded poly dC column was used for a second level of purification in later steps after the protein had been identified. The column was constructed by attaching single-stranded poly dC (Pharmacia) homopolymers of an avenge length of 450 bases to Pierce 3M EMPHAZE™ beads using their protocol. Additional steps in protein purification and isolation were performed as described in Coffman *et al.,*. 1992. The sequence of the double-stranded binding site (upper strand) used is as follows:

5' GTAATGTAACCAAATTTAATGTATATGTCGTAAT 3'.

This is the same sequence used to determine the number of repeat copies in the sea urchin genome.

Protein DNA UV Crosslinking: The protein:DNA complex was crosslinked according to the protocol outlined by Gray, *et al.,*. 1990. In our case a 20 minute exposure to long wave UV light was the most efficient.

Degenerate Probe Sequence: Amino acid sequence was experimentally determined using automated Edman degradation of peptide fragments of gel purified protein (Coffman *et al.,* 1992). Based on the sequence with the highest degree of confidence: EAEDFDGSGN, we had the following oligonucleotide synthesized :

GAA GCI GA(A/G) GA(C/T) TT(C/T) GA(C/T) GGI TCI GGI AA.

This probe was used to screen a 48 hour cDNA plasmid library. Hybridization was accomplished at 42°C. The filters were then washed in a tetramethylammonium chloride (TMACL) solution to a final wash temperature of 68°C as described previously (Anderson *et al.,*. 1994).

Subcloning and Expression of a Non-Toxic Peptide Fragment : Antibodies against RFA subunits from other species are not cross reactive (Kim *et al.*, 1992), therefore we generated antibodies against SpRFA. Expression of the entire SpRFA-70kD subunit is toxic to bacteria but we were able to express a 50kD subfragment of the protein by deleting putative binding domains on both ends of the factor. The subcloned region consisted of a 780 basepair Xba I - Bgl II fragment that was cloned in frame with sequence encoding a histidine tag in the expression vector pRSETA (Invitrogen, San Diego, CA). The resulting clones were transfected into pLys-S cells and grown overnight on supplemented agar plates. Single colonies were tested for proper expression by growing small volume cultures of 3 ml for 3 hours. These cultures were then induced to express the recombinant peptide by the addition of IPTG. Three aliquots of 500 µL each were taken every hour and lysed by osmotic shock. The resulting lysates were separated on an SDS-PAGE gel and compared to uninduced cell lysates. The presence of an

increased abundance of ~50kD protein was used to indicate proper expression. This peptide fragment was purified and antibodies raised against it in rabbits.

Antibody Detection: The 50 kD peptide fragment was purified by nickel-NTA chromatography as previously described Zeller *et al.,* (1995). This was used as an antigen to induce antibody production in rabbits by Cocalico Biologicals (Reamstown, PA). Immunoblotting and detection were performed as outlined in Zeller *et al.,* 1995.

Gel shift Assays for Relative Affinity and Equilibrium Measurements: Standard gel shift assays were performed as described previously (Calzone *et al.,* 1988) to determine relative affinity (specific to non-specific) and protein concentration. Exceptions to the prescribed protocol were 1) reduction of non-specific competitor for gel shift assays involving column purified proteins; 2) Use of a mixture of poly dA:dT and poly dI:dC in a 2:5 ratio; and 3) an increase in final KCl concentration in the reaction to 0.125 M. The target sequences used were:

DS	**5'**-GTAATGTAACCAAATTTAATGTATATGTCGTAAT-**3'**
	3'-CATTACATTGGTTTAAATTACATATACAGCATTA-**5'**
A	**5'**-GTAATGTAACCAAATTTAATGTATATGTC-**3'**
B	**5'**-ATTACGACATATACATTAAATTTGGTTAC-**3'**
C	**5'**-GATCGGATCCTTACAGCTCTTGGCATGTAAGAAG-**3'**
D	**5'**-GTGAATATGTGCACCCCTCCCCCCCCCCTCCCCCAACCATCAAAT-**3'**
E	**5'**-TTATCATTTGATGGTTGGGGGAGGGGGGGGGGAGGGGTGCACATA-**3'**

A,B,C,D, & E are all ssDNA oligonucleotides, DS is a pair of annealed oligos.

For the equilibrium measurements, saturation binding of the probe in the absence of non-specific competitor was used. These data were fitted by non-linear least squares estimations of the parameters.

Results

Purification of the protein

To purify sufficient quantities of the factor responsible for the observed site specific DNA binding activity we constructed a dsDNA affinity column by attaching concatemers of the binding site sequence to a resin. The affinity column is used to retain factors that bind specifically to the target site during passage of egg or nuclear extract over the matrix. The DNA sequence chosen for the column is the same 34 bp fragment as was previously used for the library screening and site quantification (Anderson *et al.*, 1994). The average concatemer length of binding site fragments in the column was ~1 kb. Thus there were approximately 30 head-to-tail copies of the fragment per molecule. Thirty-five ml of high speed fractionated egg extract were passed over this dsDNA column on three separate occasions in our effort to purify the factor(s). We estimate, based on BSA quantitation standards, that approximately 1 µg of total factor was isolated from each column run.

With the notable exception of SpZ12-1, which required guanidinium-HCl or high concentrations of urea to release it from binding to the appropriate double-stranded DNA affinity column (Wang *et al.*, 1995), most factors in our experience elute under column buffer wash conditions of 0.3 M to 0.9 M KCl. Our initial attempts to isolate the factor under these conditions were unsuccessful. As measured by the capacity to retard the target probe there was little or no activity in a gel mobility assay in the resulting column fractions. From tests to resolve binding conditions with egg extract we concluded that the appropriate KCl concentration optimal for gel-shift binding was 0.125 M KCl. This is

somewhat higher than is normally observed for other factors (0.08 - 0.1 M KCl). This led to our investigation of higher elution salt concentrations. We found that >1 M KCl buffer is needed to more effectively elute this protein from the affinity column.

To determine which fractions had the least amount of contaminating proteins, aliquots from each column fraction were displayed by SDS-PAGE. The protein appears largely pure in the high salt fractions, as can be seen in the representative gel shown in Figure 1a. Gel-shift activity was detectable to some extent in all fractions, but the peak of greatest activity occurred in the 1.3 to 1.7 M KCl fractions. The gel-shift activity profile paralleled the 75 kD band, the most prominent band in the 1.3 M to 1.7 M fractions. From the gel pattern, it is apparent that the great majority of non-binding proteins have been removed in the higher salt fractions.

We were unable to renature the protein after gel electrophoresis for use in gel shifts. Nor could the protein be renatured so that it would react with target site DNA when immobilized on filters. We thus crosslinked the protein to the DNA probe using UV irradiation in order to determine the size of the factor(s) binding to the probe. This repeatable assay demonstrated that the only factor from the column fraction extracts that bound to the probe was the 75 kD protein. The probe, a 34 bp double-stranded oligo nucleotide, contributes 20 kD to the complex, and a monomer of the protein accounts for the remaining 75 kD of the 95 kD complex (Fig. 1b).

Amino Acid and DNA Sequencing

Partial amino acid sequences were used to design probes to isolate the cDNA. The amino acid sequence was determined from lys-C digested, gel purified and isolated 75 kD

protein, that yielded numerous peptide fragments. These fragments were separated and isolated by HPLC, according to our established protocols (Coffman *et al.*, 1992). The amino acid sequence was determined by automated Edman degradation of the peptide fragments.

Degenerate oligonucleotide probes were derived from the amino acid sequences, and we isolated a clone that contains the complete coding region for the SpRFA-70 kD subunit from a 48 hour whole embryo cDNA plasmid library . Selective hybridization washes were used to identify only those clones with the highest homology to the probe sequence. Two rounds of rescreening resulted in the isolation of a single positive colony out of ten potential isolates from the primary screen. The 2.2 kb insert of this clone was sequenced and found to contain a continuous, 625 amino acid open reading frame. No other significant open reading frames were found. The molecular weight of the translation product of this cDNA is approximately 70 kD, close to our size estimations based on electrophoretic measurements.

The entire translated protein sequence was then used to search the protein databases for homologies. A strong similarity was found with both the human and *Xenopus* RFA 70 kD subunits. The amino acid sequences for the human, *Xenopus*, yeast and sea urchin proteins were multiply aligned using the ClustalV program (Fig. 2), with the percent similarity determined for each pairwise alignment (Table 1). The sequence of the two peptide fragments used for cloning are denoted by double underlines around positions 329 and 369. Note that the peptide sequences which were determined by Edman degradation are exact matches to the sequence obtained from the translation of the cDNA.

The region of highest homology with the proteins of other species is the putative "zinc finger" indicated in the Figure. Another well conserved region is the basic domain on the fourth line of the aligned sequences, amino acids 186 to 224. Other regions of potential interest are indicated above the amino acid sequence. It is not yet known what functions, if any, should be ascribed to these regions. The regions indicated are based solely on homology to similar domains in other proteins. It is clear, however, from the degree of sequence similarity and the conservation of particular important positions, that this clone is a copy of the message for the sea urchin RFA 70kD subunit.

Characterization of the Relative Binding Affinity and Equilibrium Binding Constants

We measured the relative affinity (Kr) of SpRFA-70 for this specific double-stranded DNA site versus random double-stranded DNA to show that the binding of SpRFA-70 is specific for this site. The Kr is measured by quantitative probe excess titration gel-shift assays as described previously (Anderson *et al.*, 1994). The Kr of the protein in whole egg extracts for the specific dsDNA site is approximately $2x10^5$. This indicates an affinity that is 100,000 times greater for the specific site than for random sequence dsDNA. The measured relative affinity constant for purified SpRFA-70 is approximately $8.6x10^4$. The SpRFA-70 used for this measurement is the 1.25 M KCl buffer eluted fraction of extract that was purified by double-stranded DNA affinity chromatography and subsequently by single-stranded DNA affinity chromatography (Fig. 4a). A simple fit of the data from one of several trials can be seen in Fig. 3a. The conservation of specific site recognition and binding in a protein purified by its affinity for

a specific site in double-stranded DNA and by its capacity to bind single-stranded DNA is further proof that we have purified the same protein that was initially identified by specific double-stranded DNA binding in egg extract.

To further characterize the SpRFA-70 kD subunit we measured the equilibrium constants (K_{eq}) for the binding of the protein to short double- or single-stranded DNA probes. The K_{eq} was determined by protein excess titration; representative data from the protein saturation experiments can be seen in Figures 3b(dsDNA) and 3c(ssDNA). For this measurement we used the protein concentration of the column purified factor, determined by probe excess titration (Fig. 3a). The fit shown in Fig. 3b is for one determination of the specific sequence dsDNA K_{eq} . The fit shown in Fig. 3c is for one determination of several random sequence ssDNA K_{eq} measurements. The results from 20 separate experiments, using either site specific dsDNA or random sequence ssDNA, to determine the K_{eq} can be seen in Table 2. The measured K_{eq} constants are 1.8×10^9 for the specific site dsDNA probe and 1.7×10^{10} for random sequence ssDNA probes 15 - 45 bases long.

The probe saturation curves were tested for conformity with expectation based on the possibilities that one, two or three molecules of protein bound per DNA molecule. Previously published information indicated that RFA does not bind well to random dsDNA sequence (Kim *et al.*, 1992 and refs. therein). According to our results, the best fit of the data for the specific site double-stranded DNA probe is that of a simple second order reaction with a protein:DNA ratio of 1:1. Tests with the random sequence single-stranded probes A, B and C show that with SpRFA-70 binds to these ssDNAs with a

protein:DNA ratio >1; Table 2 shows that the X^2 value is lowest for the assumption of two molecules of SpRFA-70 binding per probe molecule. Probes D & E are more homopolymeric: D is a C-rich and E is a G-rich oligonucleotide (see Materials and Methods), and the binding to these probes occurs in a bimolecular manner with one SpRFA-70 protein molecule binding per ssDNA molecule. It therefore appears that the homopolymers or near homopolymers behave differently than do random sequence single-stranded oligonucleotides (if we let ***n*** vary, the best fit of the data for these two probes is actually ***n***=1.5). We conclude that the SpRFA-70 subunit is <u>both</u> a site specific double-stranded DNA binding protein and a non-specific single-stranded DNA binding protein. However, the K_{eq} is 10 - 50 X higher for random single-stranded DNA over the specific site double-stranded DNA (Table 2).

With the K_{eq} values we have determined, we can derive the equilibrium constant K_n for the binding of random sequence dsDNA by SpRFA-70. Using the K_r above, K_n is calculated to be 2 X 10^4 ($K_n = K_{eq}/K_r$). This is interesting because this K_n is ~36 times higher than the K_n (5.6 X 10^2, c.f. Calzone *et al.*, 1988) measured for the average site-specific dsDNA binding protein. This indicates that SpRFA-70's affinity for random dsDNA is higher than most sequence specific DNA binding proteins' affinity for random dsDNA. However, the K_n value of 2 X 10^4 also indicates that SpRFA 70 has an affinity for random dsDNA that is ~10^6 times lower than its affinity for random sequence ssDNA (K_{eq}=2X10^{10}).

Proof that the factor cloned is the same factor characterized

To show that the same factor responsible for the binding to the random sequence ssDNA is also the protein responsible for site specific dsDNA binding, we constructed a ssDNA affinity column. We combined several 0.9 M KCl fractions eluted from the dsDNA affinity column and ran them over the single-stranded poly dC DNA affinity column to further purify the SpRFA-70 kD subunit. It is highly unlikely that any factor other than the RFA 70 kD subunit could bind specifically to both the dsDNA column and to this ssDNA column. A steeper elution gradient than previously used was applied and the activity from each fraction was determined as above. The gel-shift activity is found in all fractions with the peak of activity occurring in the first 1.0 M fraction. The purity of the resulting fractions was determined by SDS-PAGE separation of the proteins in each fraction (Fig. 4a). After a second purification column only one protein is visible in the 1.0 and 1.25 M KCl fractions. In this manner we were able to obtain highly purified protein and show more conclusively that the factor responsible for the gel-shift activity is that found in the 75 kD band. This experiment verifies that the protein binding to the site specific dsDNA probe is also a protein that is purified by affinity to a single-stranded DNA column.

We analyzed samples of both recombinant and column purified protein on an SDS-PAGE gel by antibody blotting, using an antibody raised against the cloned protein fragment. The proteins were transferred to nitrocellulose and, after appropriate blocking steps, incubated in a solution with a rabbit antibody raised against the SpRFA protein. This antibody reacted with both the protein fragment expressed in bacteria and with the

protein purified by dsDNA and ssDNA column chromatography (Fig. 4b). Though there is a large amount of protein present in the bacterial lysate lane, and there are some background hybridization bands, the control preimmunization serum did not show any reaction with an identically blotted and treated filter(data not shown). When crude egg extract was subjected to SDS-PAGE electrophoresis and reacted with the antibody, one band only was recognized by the antibody (data not shown). These data demonstrate that the protein purified, isolated, cloned and characterized is the sea urchin homologue of the RFA 70kD subunit.

Discussion

Properties of RFA 70

Replication Factor A is a three subunit (14, 35 and 70 kD) enzyme that has been characterized in some detail. Among the more well documented activities attributed to RFA are single-stranded DNA binding and interaction with SV40 large T antigen (Wold and Kelly, 1988). RFA has been shown to interact with several additional replication factors including helicases, DNA polymerase α, DNA primases, Topoisomerases I and II, DNA strand repair and factors involved in homologous recombination (Blackwell and Borowiec 1994 and refs. therein). Some of these interactions are attributable to the 35 and 14 kD subunits and are modulated by various kinases. In *Xenopus* RFA has been shown to mark the sites of replication in the oocyte prior to fertilization (Adachi and Laemmli 1992). The ssDNA binding activity is a property of the 70 kD subunit and is cell cycle regulated (Carmichael *et al.,* 1993), is reported not to be sequence specific (Kim *et al.,* 1992) and does not require the other two subunits. RFA holds the DNA in a single-stranded form to facilitate replication. It is essential for replication and viability. RFA from different species can substitute as single-stranded binding proteins but they have not been shown to functionally interact with the other replication machinery components in cross species in vitro replication assays (Wold and Kelly 1988). The interaction of the RFA 70 kD subunit and DNA are the most pertinent activities for us in the context of this research.

SpRFA-70 binds a specific dsDNA target site

In this paper we demonstrate that the protein which binds to the P10 sites of CyIIIa, SpZ2-1 and P3A2 is the sea urchin RFA 70 kD subunit. This contention is based on the sequence of the protein purified from the 75 kD band and the subsequent cloning of the message for that protein. The determination that the binding factor was found in the 75 kD band of dsDNA affinity column extracts is based on the purification profiles of the 75 kD band and the gel-retardation activity for the specific site probe(Fig. 1a). We observed the specific site binding activity in all elution fractions containing the 75 kD band from both dsDNA and ssDNA affinity columns(Figs. 1a and 4a). This was confirmed by crosslinking the probe to the protein in the gel-shift complex and determining the size of the resulting product to be 95 kD which is the sum of 75 kD plus the 20 kD probe. (Fig. 1b).

The amino acid sequences determined by Edman degradation for the peptide fragments from the 75 kD band matched perfectly with the translation product from the cDNA obtained for SpRFA-70. Alignment of the amino acid translation from the cloned cDNA to the amino acid sequences of known RFA 70kD subunits from three other species clearly shows the homology between these proteins (Fig. 2). The conservation of the 4 cysteine metal binding, proline rich and basic domains are further indications of the matching identity. SpRFA-70 retains a 45% similarity to the *Xenopus* and human proteins (Table 1).

Further evidence that the binding factor is SpRFA-70 comes from our findings on the single-stranded DNA binding activity of the purified protein. Previous publications

have repeatedly demonstrated the non-specific affinity of RFA 70 for ssDNA (Henricksen *et al.*, 1992 and refs. therein). The binding of SpRFA-70 to ssDNA is sequence independent and occurs with a K_{eq} of approximately $2x10^{10}$ which matches the K_{eq} determined for human RFA(hRFA) measured by Kim *et al.*, (1992). (Note: recombinant hRFA generously provided by Dr. Marc S. Wold showed a similar gel-shift pattern for the P10 specific site and we were able to compete away the binding with cold P10 site probe DS, in the same manner as with the SpRFA-70.) In some cases a high degree of cooperativity has been indicated for the RFA binding to ssDNA in yeast and we also saw some indication of this in our K_{eq} determinations (Table 2).

Our most concrete evidence that the cloned factor and the purified factor are the same protein comes from concurrent immunodetection of the native column-purified protein and a bacterially expressed peptide fragment from SpRFA-70 (Fig. 4b) with an antibody raised against the bacterially expressed fragment. The same anti-SpRFA-70 fragment antibody also reacted with a 75 kD protein in 24 hour embryo nuclear extract. This shows conclusively that the purified factor is the same protein as that encoded by the message we cloned and that protein is the sea urchin replication factor A 70 kD subunit (SpRFA-70).

SpRFA-70 also has a dsDNA binding activity

We have shown that SpRFA-70 not only binds ssDNA but it also has a site specific dsDNA binding activity for at least one identified set of sites in the sea urchin genome. The fact that the P10 specific site binding activity was purified by both dsDNA and ssDNA affinity chromatography thus demonstrates it has the capacity for both dsDNA and

ssDNA binding. The presence of both of these activities in the unfractionated extracts of egg and 24 hour embryos, and our demonstration that the purified SpRFA-70 from column fractions performs the same functions on the specific site dsDNA and on the random sequence ssDNA, is additional evidence that we have identified and purified the appropriate protein. In yeast it has been established that RFA binds specifically to the *URS1* site of the *CAR1*(arginase) gene and that it acts as a repressor at that site(Luche *et al.*, 1990 and Luche *et al.*, 1992, Viljoen *et al.*, 1992 and refs. therein).

What is SpRFA-70 doing in the early stages of development

We earlier identified over 400 sites in the sea urchin genome to which SpRFA-70 probably binds, though we have no evidence for functions at any of these sites. We have demonstrated that there is a correlation between the presence of SpRFA-70 binding sites and presence of SpZ12-1 sites (Anderson *et al.*, 1994). In light of the fact that RFA is known to interact with the gene regulatory factors VP16, p53, and GAL4 (Li and Botchan, 1993, He, *et al.*, 1993), SpRFA-70 may interact with SpZ12-1 to modulate the regulatory properties of SpZ12-1 (positively or negatively). We have not however, been able to demonstrate a function for SpRFA-70 binding site in the *CyIIIa* promoter region during embryogenesis as we have for SpZ12-1 (Wang *et al.*, 1995).

In our previous publication we also showed that there is a much higher abundance of active SpRFA-70 in the unfertilized sea urchin egg than is present at 24 hours in nuclear extract. One reason for this difference could be the activation state of the protein at the different times. Another, and more probable reason is that the requirements for DNA replication in the fertilized egg are much higher in the first stages of cleavage than at any

other time in development. It has been well documented that eggs have large amounts of DNA polymerase, primases and other proteins required for replication (Davidson, 1986).

The greater abundance of RFA in the egg, on a per nucleus basis, indicates that it probably has a greater effect on gene regulation in the early stages. There are several mechanisms one can envision about how RFA could function in these early stages. If RFA were to act as a repressor in the mature oocyte this could keep a large network of genes silent awaiting derepression at fertilization. If a large percentage of RFA protein is not activated, so as to be able to bind specifically to dsDNA , prior to fertilization then it could act, if activated at fertilization, as an enhancer to stimulate a large network of genes as part of the large increase in gene expression seen in response to fertilization. We do not as yet have any evidence regarding the modulation of site specific dsDNA binding for SpRFA-70 so we can only speculate about these possibilities at this time. Most likely there will be some combination of these two events. The dual role of RFA as a replication factor and as a gene regulatory factor makes this protein even more interesting as a subject for study.

In conclusion we have identified the factor which binds to the P10 site of the CyIIIa regulatory region, the SpZ2-1 flanking region and downstream of the P3A2 coding region to be SpRFA-70. We determined the equilibrium constant for SpRFA binding to ssDNA to be $1.7 \times 10^{10}\ M^{-1}$ and $1.8 \times 10^{9}\ M^{-1}$ for dsDNA. We propose that this protein interacts with SpZ12-1 based on the frequency of occurrence together of both of these sites, the large number of sites for each in the genome, and the large quantities of both factors in the unfertilized egg (Wang *et al.*, 1995 and Anderson *et al.*, 1994).

References

1. Adachi, Y. and Laemmli, U. K. (1992). Identification of nuclear pre-replication centers poised for DNA synthesis in Xenopus egg extracts: Immunolocalization study of replication protein A. *J. Cell Bio.* **119**, 1-15

2. Anderson, R. F., 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

3. Blackwell, L. J., and Borowiec, J. A. (1994). Human replication protein A binds single-stranded DNA in two distinct complexes. Mol. Cell. Biol.**14**, 3993-4001

4. Calzone, F.J., Thézé, N., Thiebaud, P., Hill, R. 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 and Development* **2**, 1074-1088

5. Calzone, F.J., Höög, C., Teplow, D. B., Cutting, A. E., Zeller, B. 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

6. Carmichael, E. P., Roome, J. M., and Wahl, A. F. (1993). Binding of a sequence-specific single-stranded DNA-binding factor to the Simmian Virus 40 core origin inverted repeat domain is cell cycle regulated. *Mol. Cell. Biol.* **13**, 408-420

7. 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

8. 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. Natl. Acad. Sci. USA* **87**, 7953-7957

9. Davidson, E. H. (1986). Gene activity in early development. *3rd ed. Academic Press, Orlando FL.*

10. Gray, T. A., Gumucio, D. L. and Collins, F. S. (1990). A two-dimensional electrophoresis system for DNA-binding proteins. *J. Meth. Cell Mol. Biol.* **2**, 147-154

11. He, Z., Brinton, B. T., Greenblett, J., Hassell, J. A. and Ingles, C. J. (1993). The transactivator proteins VP16 and Gal4 bind replication factor A. *Cell* **73**, 1223-1232

12. Hendricksen, L. A., Umbricht, C. B., and Wold, M. S. (1994). Recombinant replication protein A: expression, complex formation, and functional characterization. J. Biol. Chem. **269**, 11121-11132

13. Höög, C., Calzone, F.J., Cutting, A. E., Zeller, B. W., 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

14. Kim, C., Snyder, R. O., and Wold, M. S. (1992). Binding properties of replication protein A from human and yeast cells. *Mol. Cell. Biol.* **12**, 3050-3059

15. Li, R., and Botchan, M. R. (1993). The acidic transcriptional activation domains of VP16 and p53 bind the cellular replication protein A and stimulate in vitro BPV-1 DNA replication. *Cell* **73**, 1207-1221

16. Luche, R. M., Sumrada R., and Cooper, T. G. (1990). A cis-acting element present in multiple genes serves as a repressor protein binding site for the yeast CAR1 gene. *Mol. Cell. Biol.* **10**, 3884-3895

17. Luche, R. M., Smart, W. C., and Cooper, T. G. (1992). Purification of the heteromeric protein binding to the URS1 transcription repression site in *Saccharomyces cerevisiae. Proc. Natl. Acad. Sci. USA* **89**, 7412-7416

18. Luche, R. M., Smart, W. C., Marion, T., Tillman, M., Sumrada, R. A. and Cooper, T. G. (1993). *Saccharomyces cerevisiae* BUF protein binds to sequences participating in DNA replication in addition to those mediating transcriptional repression (URS1) and activation. *Mol. Cell. Biol.* **13**, 5749-5761

19. 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

20. Viljoen, M., Kovari, L. Z., Kovari, J. A., Park, H.-D., van Vuuren, H. J. J. and Cooper, T. G. (1992). Tripartite structure of the *Saccharomyces cerevisiae* arginase (CAR1) gene inducer-responsive upstream activation sequence. J. of Bacteriology **174**, 6831-6839

21. Wang, D. G.-W., Kirchhamer, .C. V., Britten, R.B., and Davidson, E. H. (1995). SpZ12-1, A negative regulator required for spatial control of the territory-specific CyIIIa gene in the sea urchin embryo. *Development* **121**, 1111-1122

22. Wold, M. S. and Kelly, T. (1988). Purification and characterization of replication protein A, a cellular protein required for in vitro replication of simmian virus 40 DNA. *Proc. Natl. Acad. Sci. USA* **85**, 2523-2527

23. Zeller, R. W., Britten, R. J. and Davidson, E. H. (1995). Developmental utilization of P3A1 and P3A2: Two proteins which recognize the same DNA target site in several sea urchin gene regulatory regions. *Submitted*

Table 1:

Percent Similarity for RFA 70 kD Subunits from four different organisms: Sequences and alignments are shown in figure 2.

	Sp	Human	Xenopus	Yeast
Sp	-	45.3%	43.0%	28.2%
Human	45.3%	-	71.8%	29.6%
Xenopus	43.0%	71.8%	-	29.0%
Yeast	28.2%	29.6%	29.0%	-

Sp = Strogylocentrotus purpuratus.
Percent Similarity is based on the number of perfect matches over the total length for the best matching sequence.

Table 2: Equilibrium Measurements for both dsDNA and ssDNA Target Sites of SpRFA

Probe	type	Length	K_{eq} (p) M^{-1}	X^2 values p = 1	p = 2	p = 3	t
DS	ds	34	1.8E+09 (1)	0.4	1.3	2.5	3
A	ss	29	1.3E+10 (2)	1.4	0.1	0.2	5
B	ss	29	1.7E+10 (2)	1.5	0.1	0.2	5
C	ss	34	2.2E+10 (2)	1.0	0.1	0.5	2
D	ss	45	6.0E+10 (1)	0.3	0.6	1.7	3
E	ss	45	4.9E+10 (1)	0.4	1.2	2.1	2

K_{eq} is given in moles/L^{-1}, based on the chi square values for the fit (X^2) when p = to 1, 2 or 3 molecules of protein per probe. When p = 2 the protein is complexed with the DNA in a 2 to 1, protein to DNA ratio. t = number of times the experiment was repeated with the particular probe.

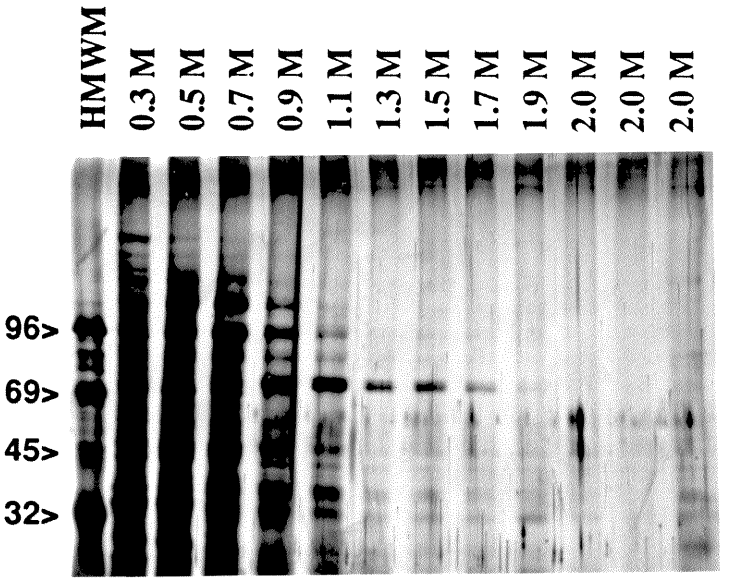
The function from which these values were calculated is: $$\frac{[PD]}{[PD]+[D]} = \frac{K_{eq}^{n} \cdot [P]^{n}}{1 + K_{eq}^{n} \cdot [P]^{n}}$$

Where $[P]$ is the total concentration of binding protein, $[D]$ is the concentration of free probe, and $[PD]$ is the concentration of probe complexed with protein.

Fig. 1. ***Purification and size determination of the protein that binds the P10 Site***

(a) *Salt elution fractions from a dsDNA site specific affinity column separated on a 6.0% SDS-PAGE gel:* Lane 1 (HMWM) shows a high molecular weight protein ladder. The remaining lanes show the eluted fractions at the indicated molarities of KCl Buffer. Column fractions were separated on the SDS-PAGE gel by electrophoresis for 1.5 hours at 130VDC. The specific binding activity is present in all of the fractions from this and several other similar column/gel runs. (b) *UV crosslinked protein DNA complex.* High salt fractions (1.1 and 1.3M KCl buffer) from a column purification as in fig 1a were used to determine which of the many bands on the gel represents the protein(s) responsible for the observed DNA binding activity. The scale on the left represents the unlabelled molecular weight markers visualized by silver staining of the gels prior to film exposure.

(a)


HMWM
0.3 M
0.5 M
0.7 M
0.9 M
1.1 M
1.3 M
1.5 M
1.7 M
1.9 M
2.0 M
2.0 M
2.0 M
96>
69>
45>
32>

(b)

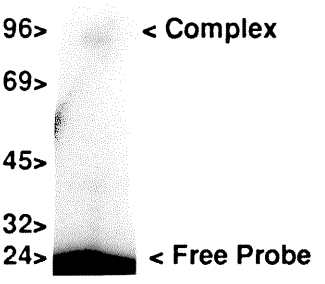

96>
69>
45>
32>
24>
< Complex
< Free Probe

Fig. 2. ***Alignment of SpRFA-70 Amino Acid Sequence to Xenopus, Human and Yeast Sequences From Prosite/GenProt.***

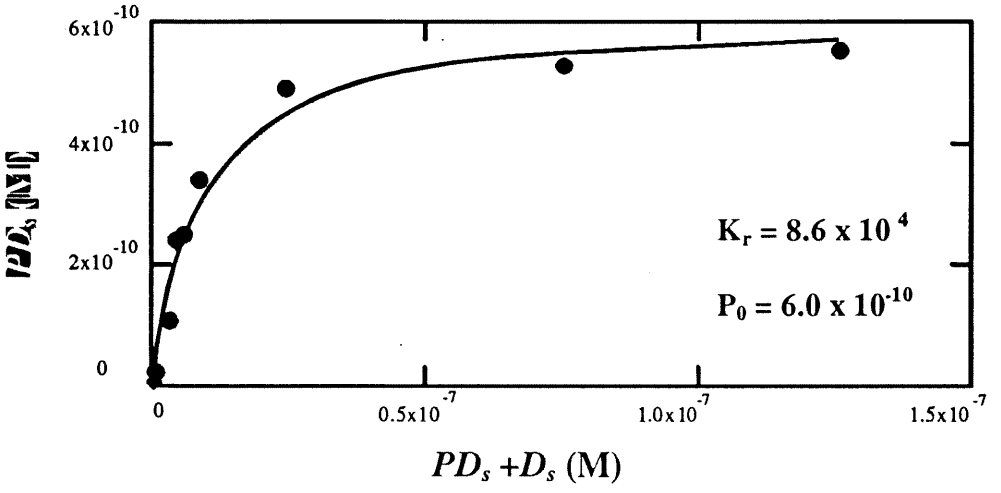
The sequence of a clone containing the full coding region for SpRFA-70 is compared to and aligned with sequences from Prosite/Genbank files for Human RFA (P27694), *Xenopus laevis* RFA(Q01588) and Yeast *Saccharomyces cerevisiae* RFA(P22336). These sequences were aligned using the ClustalV program, with minor adjustments made by hand afterwards. The uppercase letters and a hat (^) mark positions of conservation between the SpRFA-70 kD subunit and any one of the other three proteins. Asterisks (*) below the sequences indicate positions where all four match exactly. Putative functional domains have been indicated according to homologies with identified domains in other proteins. The # symbol above the sequence on the second line indicates hydrophobic amino acids, which are repeated every 7 positions. This could give rise to a coiled coil helical domain. The sea urchin, human, and *Xenopus* proteins all share a proline rich domain (9-12 residues), but there is no conservation of the positions of these residues. The basic domain is well conserved between all four homologues, as is the single 4C zinc finger that can be found near the carboxy terminus. The amino acid sequences determined from the peptide fragments are found starting at amino acid positions 329 and 369. The experimentally determined peptide sequences corresponded exactly with the amino acid sequence derived from the DNA sequence. The percent similarities of the RFA 70 sequences of various species are shown in Table 1.

```
SpRFA-70   Mn-nlLSRGAIAAIFrgenvSC-PvLQllackkmnaaasGkavdRYRLMlSDGehT-ctA  57
rfa1_XENLA MalpqLSeGAIsamlg-gdsSCkPtLQvi---nirpintGngppRYRLlmSDGlnTlssf
rfa1_HUMAN M-vgqLSeGAIAAImqkgdtnikPiLQvi---nirpittGnsppRYRLlmSDGlnTlssf
rfa1_YEAST MssvqLSRGdfhsIFtnkqrydnPtggvyqvyntrksdgansnrknliMiSDGiyhmk-A
           *    ** *^^^^^^     ^^ * ^^           ^    ^^^^^^^**  ^   ^
            #       #      #      #      #      #    -  #      #
SpRFA-70   MLATQLNemVstgELdvkaamklknyscNpiandRRVIvVlD-LDiVKkgseiGVsIGdP 116
rfa1_XENLA MLATQLNslVdnnlLatncicqvsrfivNnlkdgRRVIiVme-LDvlKsadlvmgkIGnP
rfa1_HUMAN MLATQLNplVeeeqLssncvcqihrfivNtlkdgRRVvilme-LevlKsaeavGVkIGnP
rfa1_YEAST lLrnQaaskfqsmELqrgdiirviiaepaivrerkkyvllvDdfelVqsradmvnqtstf
           ^*^^*^^  ^   ^*           ^     ^^^^ ^ ^ ^^ ^^     ^^ ^^ ^
           [--------------Proline rich domain---------------]
SpRFA-70   tP-mrapGQGggpAPAqqq-srPdstvshdPpqtarptsyGTgapasAipalqgtfyGqs 174
rfa1_XENLA qP--yndGQpqpaAPApasapaPapsklqnnsapppsmnrGTsklfg---------gGsl
rfa1_HUMAN vP--yneGlGqpqvappapaasPaassrpqPqngssgmgstvskaygAsktfgkaagpsl
rfa1_YEAST ldnyfsehpnetlkdeditdsgnvanqtnasnagvpdmlhsnsnlnanerkfane-----
            ^     ^^^   ^^^      ^      ^          ^^     ^          ^
                     [------------Basic domain-------------]
SpRFA-70   nsamGGSttspkKVqPIsSLTPYQNrWTIKARVTNKtaIRTWSNaRGEGKLFSmdLLDqS 234
rfa1_XENLA lntpGGSqs---KVvPIaSLnPYQskWTvrARVTNKgqIRTWSNsRGEGKLFSiemvDeS
rfa1_HUMAN shtsGGtqs---KVvPIaSLTPYQskWTIcARVTNKsqIRTWSNsRaEGKLFSleLvDeS
rfa1_YEAST ---npnSqktr-pifaIeqLsPYQNvWTIKARVsyKgeIkTWhNqRgdGKLFnvnfLDtS
               ^^^     ^^ ^^ ^*^***^ **^*****^*  *^***^* *^^*****^  ^^* ^

SpRFA-70   GEIRcTAFkDmVDKyyEmIEiGKVYfVSrGTLKPANrQYTSinNDYELTFNNDTmVEPCv 294
rfa1_XENLA GEIRaTAFneqaDKffsiIEvnKVYyfSkGTLKiANkQYTSvkNDYEmTFNseTsViPC-
rfa1_HUMAN GEIRaTAFneqVDKffplIEvnKVYyfSkgtLKiANkQfTavkNDYEmTFNNeTsVmPC-
rfa1_YEAST GEIRaTAFnDfatKfnEilqeGKVYyVSkakLqPAkpQfTnlthpYELnldrDTviEeCf
           ****^*** ^ ^^*  ^ ^^ ^*** ^* ^^*^^^ *^*^  ^^*^^^^^^^ ^^^*

SpRFA-70   EEDvsiPaVQFDFkSIshLEdtpeDSmiDVIGVCKStsDlTaVTIKSsNREVnKRSlqLV 354
rfa1_XENLA ddsadvPmVQFeFvSIgeLEsknkDtvlDiIGVCKnveevTkVTIKSnNREVsKRSihLm
rfa1_HUMAN EdDhhlPtVQFDFtgIddLEnkskDSlvDiIGiCKSyeDaTkiTvrSnNREVaKRniyLm
rfa1_YEAST dEs-nvPkthFnFikldaiqnqevnSnvDVlGIiqtinphfelTsragkk-fdrRditiV
           ^^^   * ^^*^* ^^  ^^    ^^  *^^*^^^^  ^ ^ ^*^^^ ^^^^ ^*^  ^^

SpRFA-70   DDSqKeVSlTLWGkeAeDFDGSgnPViAvKGARLSgFGGRSLSVLqnSifqvNPDIPkAh 414
rfa1_XENLA DsSgKvVStTLWGedAdkFDGSrqPVvAiKGARLSdFGGRSLSVLssStvmiNPDIPeAf
rfa1_HUMAN DtSgKvVtaTLWGedAdkFDGSrqPVlAiKGARvSdFGGRSLSVLssStiiaNPDIPeAy
rfa1_YEAST DDSgfsiSvgLWnqqAlDFnlpegsVaAiKGvRvtdFGGkSLSmgfsStlipNPeIPeAy
           *^* ^ ^^ ^***  * ^*^^^  ^* * **^*^^ ***^***^^  *    **^** *

SpRFA-70   hLKGWFDSEGhsqDsqSIStrqGsGGgGGAN-----TNWmTfhdVhAqNLGQGEKPDYFT 469
rfa1_XENLA kLraWFDSEGqvvegtSISesrG-GGtGGgN-----TNWkslleVkneNLGhGEKaDYFT
rfa1_HUMAN kLrGWFDaEGqalDgvSISdlks-GGvGGsN-----TNWkTlyeVkseNLGQGdKPDYFs
rfa1_YEAST aLKGWyDSkGrnanfitlkqepGmGGqsaAsltkfiaqriTiaraqAeNLGrseKgDfFs
           *^^*^*^^*   ^  ^^^   ^ ** ^^^^     ^^^ ^   ^ ^ ***^^^*^*^*^
                            [C    C Zinc finger C  C]
SpRFA-70   VKgTIlFvRKENCMYmACPSaeCNKKVsEngdGsYRCEKCsKdyeNFKYRLlLSANvADs 529
rfa1_XENLA svaTIvylRKENClYqACPSqdCNKKVidqqnGlfRCEKCnKefpNFKYRLiLSANiADf
rfa1_HUMAN svaTvvylRKENCMYqACPtqdCNKKVidqqnGlYRCEKCdtefpNFKYRmiLSvNiADf
rfa1_YEAST VKaaIsFlkvdNfaYpACsnenCNKKVlEqpdGtwRCEKCdtnnarpnwRyiLtisiiDe
           ^^ ^^ ^ ^^^*^^* ****  ***** ^   * ^***** ^   ^^^^^^ *^^^ ^*

SpRFA-70   TdNQWaTCFQEtAEQLLlksAqeLGSLKDqgEatekEFnqVFQdAcFidYmFRmRiKmET 589
rfa1_XENLA geNQWiTCFQEsAEsiLgqnAtyLGeLKeknE---qaydeVFQnAnFrsYtFRaRvKlET
rfa1_HUMAN qeNQWvTCFQEsAEaiLgqnAayLGeLKDknE---qaFeeVFQnAnFrsfiFRvRvKvET
rfa1_YEAST TnqlWlTlFddqAkQLLgvdAntLmsLKeedp---nEFtkitQsiqmneYdFRiRaredT
           ^ ^^* *^*^^ *^^^*   *  *^^***  ^    ^^  ^^* ^ ^  ^ ** * ^ ^*

SpRFA-70   YNeEaRlKcTcVsaqPiNvRdYtnkLIkdIRlMA--sA 625
rfa1_XENLA YNdEsRiKaTaVdvkPvdhkeYsrrLImnIRkMAtqgv
rfa1_HUMAN YNdEsRiKaTvmdvkPvdyReYgrrLvmsIRrsAl--m
rfa1_YEAST YNdqsRiryTvanlhslNyRaeadyLadelskall--A
           ** ^ * ^ * ^   ^ ^ ^ ^   *^  ^^ ^^   ^
```

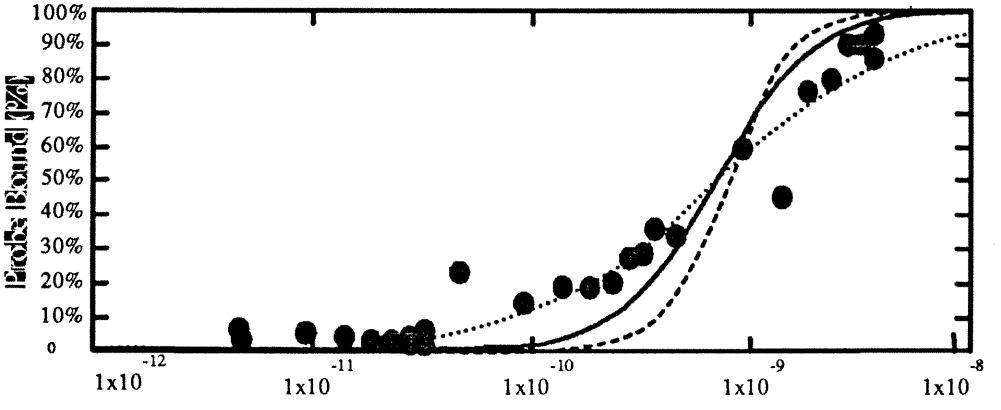
Fig. 3. ***Probe Excess Titration Determination of SpRFA Relative Affinity and DNA Binding Equilibrium Measurements for dsDNA and ssDNA***

(a)Relative binding affinity of SpRFA for specific site double-stranded DNA Estimation of relative binding affinity (Kr) and protein quantity by quantitative gel shift titrations of the SpRFA-70 protein using a specific dsDNA target. The Kr and P_0 values determined in this measurement are shown in the lower right corner of the graph. The Kr and P_0 were determined according to the method described in Calzone *et al.*, (1988). The probe in this case was the 125 bp fragment from the *SpZ2-1* upstream region that contains a pair of inverted SpRFA-70 binding sites. The extract was the first 1.25 M KCl column fraction shown in Fig. 4a. These results are quantitatively similar to those obtained previously with whole embryo nuclear extract (Anderson *et al.*, 1994). (b and c) *Equilibrium constants determined by probe saturation measurements with a specific dsDNA probe or multiple ssDNA probes.* (b)Examples of data are shown for the specific site dsDNA binding using the oligonucleotide probe DS (see materials and methods) and (c) for ssDNA binding using probe C (materials and methods) The extracts used for these measurements were combined salt elution fractions, in which the most prevalent protein was SpRFA-70. The amount of input protein was determined by probe excess titration measurements with the short 34 bp dsDNA probe DS and found to be approximately $2.2 \cdot 10^9$ molecules/μL. The lines represent the theoretical curves for the protein binding to the probe in a 1:1 (- - - -), 2:1 (------) or a 3:1 (-- -- --) protein to DNA complex.

(a)


6x10^-10
4x10^-10
2x10^-10
0
0
0.5x10^-7
1.0x10^-7
1.5x10^-7
PDs [M]
PDs +Ds (M)
Kr = 8.6 x 10^4
P0 = 6.0 x 10^-10

(b)


100%
90%
80%
70%
60%
50%
40%
30%
20%
10%
0
Probe Bound [%]
1x10^-12
1x10^-11
1x10^-10
1x10^-9
1x10^-8

(c)

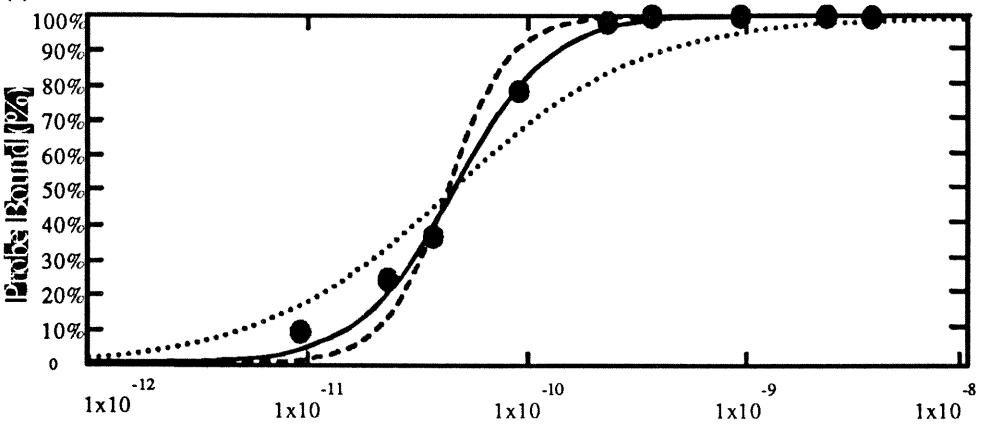
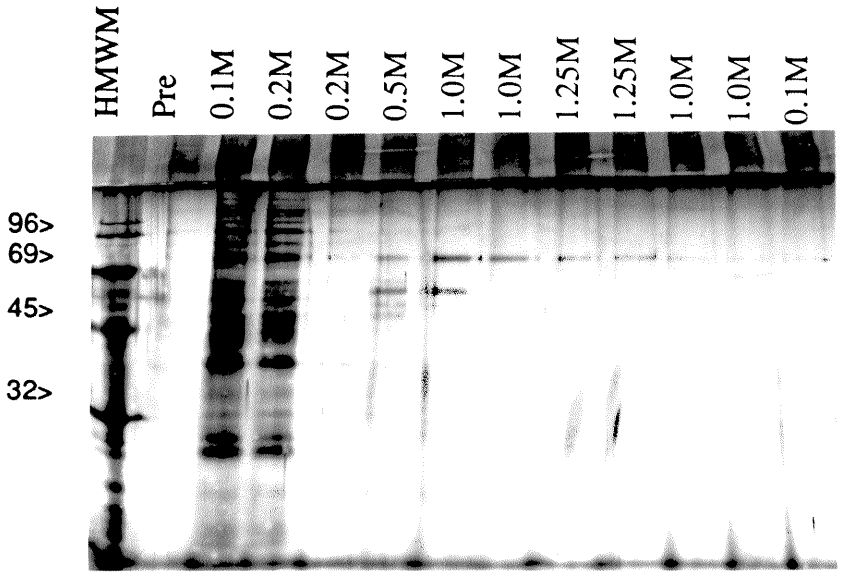

100%
90%
80%
70%
60%
50%
40%
30%
20%
10%
0
Probe Bound [%]
1x10^-12
1x10^-11
1x10^-10
1x10^-9
1x10^-8

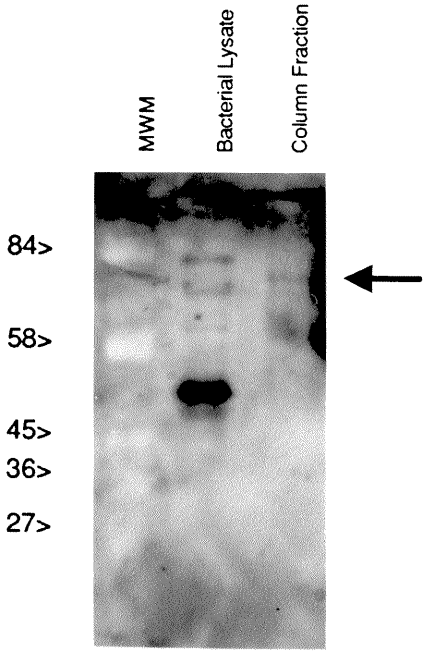
Fig. 4. ***Purification and Immunodetection of SpRFA***

(a)SDS-PAGE separation of fractions from ssDNA poly dC column The extract passed over the column is a combination of several 0.7 M fractions from the dsDNA affinity column (Figure 1a). Lane P1 shows an aliquot of extract taken before application to the column. The lane labeled 0.1 M shows an aliquot of the extract immediately after passage throughout the column. The subsequent lanes show fractions eluted at the indicated KCl buffer molarities. The MW lane shows a high molecular weight protein ladder. The SpRFA-70 binding activity is present in all fractions from this and similar column runs. (b) *Immunodetection of SpRFA in Crude Bacterial Lysate and Affinity Purified Fractions from Egg Extract:* Antibody raised against the expressed peptide fragment was used in the detection. Visualization was effected by use of enzyme labeled secondary antibody (goat anti rabbit-HRP conjugate). The bacterial lysate lane shows the reaction with proteins from an aliquot of bacterial lysate, from cells expressing a peptide fragment of SpRFA-70. The additional bands are due to the high quantity of lysate loaded on the gel. The column fraction lane shows the reaction with native SpRFA-70 protein, purified by affinity chromatography and eluted with a second wash of 1.0M KCl buffer from the single-stranded DNA affinity column. The arrow indicates the position of the native protein and the location of the antibody staining in the "Column Fraction" lane.

(a)


HMWM
Pre
0.1M
0.2M
0.2M
0.5M
1.0M
1.0M
1.25M
1.25M
1.0M
1.0M
0.1M
96>
69>
45>
32>

(b)


MWM
Bacterial Lysate
Column Fraction
84>
58>
45>
36>
27>

Appendix 1

Detection of the SpRFA-70/SpZ12-1 repeat in other species

Roger F. Anderson, Catherine A. Moore and Eric H. Davidson

Abstract

In the *S. purpuratus* genome we identified a low copy (~400 copies per genome), highly conserved repeat element, that we call the SpRFA-70/SpZ12-1 repeat, based on the two factors, SpZ12-1 (Wang *et al.*, 1995) and SpRFA-70 (Ch. 3) which bind to different well conserved regions of this repeat. We looked for the presence of the two regions that contain these binding sites in the genomes of other organisms. Genomic DNAs from five sea urchin species as well as 15 other unrelated species were used to screen for these regions. Except for other species of *Strongylocentrotus*, the repetitive element could not be detected in other genomes.

Introduction

SpRFA-70 is the sea urchin homologue of the replication factor A 70 kD subunit (RFA-70). RFA is a ubiquitous protein found in all eukaryotes that is required for DNA replication (Kim *et al.*, 1992). We demonstrated that SpRFA-70 binds to a specific site in the SpRFA-70/SpZ12-1 repeat (Ch 2). Since RFA is found so widely it follows that its binding site would be present in all animal genomes. We wanted to determine if the SpRFA-70/SpZ12-1 repeat in which the specific binding site is embedded is also found in other genomes. It has been demonstrated previously that interspersed repeat families tend to be restricted to closely related species. The SpRFA-70/SpZ12-1 repeat is a good candidate to see if a repeat with one or more functional protein binding sites would follow this rule. We show by hybridization to blotted genomic DNA from multiple species that there are copies of the *SpRFA-70* and the *SpZ12-1* repeat regions present in the genomes of three species of *Strongylocentrotus*. No significant evidence was found to suggest that either of these regions is conserved outside of the class Echinoidia.

Results

An array of genomic DNA samples was digested, run out in a gel and blotted onto nylon membrane to use in screening for copies of these regions. The source organisms were chosen for their relation to *S. purpuratus*, or their wide use in laboratory research. Specific probes (Fig. 1) were hybridized to the DNA on the filters and then washed under controlled criteria to determine the lower limit of homology allowed. The filters were washed under increasing stringencies and exposed to film between washing steps. The

results of the highest stringency washes after probe hybridization are shown in Figures 2a and 2b. These experiments reveal sites with significant homology to both repeat regions in the three species of *Strongylocentrotus* and a weak homology to the *SpZ12-1* region in the sea urchin *Echinometra viridis*. No significant numbers of homologous copies of these sections of the repeat are detectable under these conditions in any of the other genomes tested. The smear that is barely detectable in the lane showing the calf DNA is probably due to an excess of DNA in that lane.

To determine if the arrangement of binding sites and the repeat length are conserved in the other sea urchin species, we used PCR to amplify the repeat region between the *SpRFA-70* and *SpZ12-1* probe sites in the various sea urchin genomic DNA's. The primers used in the amplification are the same binding site oligonucleotides used to probe the blots (Fig. 1). No amplification products from this repeat family can be observed from any of the genomes other than *S. purpuratus*. Three possible differences in the repeat structure could account for this result: the bases at the ends of the sites are not well conserved, the sites are not in close proximity to one another or the sites are not in the proper orientation for amplification. These results indicate that *S. purpuratus* is probably the only sea urchin species in which the repeat structure we identified is maintained. Yet as we showed earlier, exactly this structure occurs several hundred times and is remarkably well conserved, in the S. purpuratus genome (Anderson *et al.*, 1994).

Discussion

The results shown in Fig. 2 distinguish the phylogenetic distribution of the repetitive sequence elements that our probes identify from the distribution of the binding

sites for SpZ12-1 and SpRFA-70 factors *per se*. Though the phylogenetic distribution of the SpZ12-1 transcription factor has not been investigated specifically, it is certainly not to be expected that this factor would be confined to the genus *Strongylocentrotus*. Transcription factors and their target sites are generally of widespread distribution, as the requirement for coevolution (i.e., of the protein's recognition sequence, and its DNA target site) will retard evolutionary change; and as Wang *et al.* (1995) showed, the Zn finger binding domain of SpZ12-1 in fact closely resembles that of a *Drosophila* Zn finger factor *Krüppel*. As noted above SpRFA-70 is a ubiquitus factor among eukaryotes. The probe that we used to detect the *SpRFA-70* repeat element is 28 bp in length. The length of the actual *SpRFA-70* target site is unknown, though it is a highly specific site that is bound very tightly by a single SpRFA-70 molecule per probe fragment (Chapter 3). The probe for the *SpZ12-1* site is 21 bp in length and includes a single site for the SpZ12-1 protein, which is 10 bp long (Wang *et al.*, 1995). Since the criterion conditions applied in the genome blot of Fig. 2 of this Appendix allowed no more than 3 bp mismatches per probe molecule, we are clearly assaying the presence or absence of the repeat elements in which these two target sites are embedded, not of the sites themselves. These repeat elements are, as shown in Fig. 2, confined to the genus *Strongylocentrotus*.

Interspersed repetitive sequences in sea urchins, as in other taxa, tend to occur in narrow phylogenetic distributions, and in fact it is often observed that the frequency of a given family varies over 10-fold even among congeners (Moore *et al.*, 1978). In this respect the elements assayed here are typical: they are undetectable as such even in other sea urchin genera.

Since the factor target sites are undoubtedly widespread, while the interspersed repeats in which we found them in *S. purpuratus* DNA are tightly confined to this genus, we conclude that these target sites must have been "captures" by a mobile repetitive sequence in the common ancestor of the Strongylocentroids. The repeat then amplified and dispersed around the genome, carrying the target sites with it. This is an interesting and perhaps fundamentally important mechanism of evolution (Britten and Davidson, 1971; Klein et al., 1978; Moore et al., 1978; Britten, in press *Phylogenetics & Evolution*). Most of the *SpRFA-70* and *SpZ12-1* target sites embedded in this compound repeat sequence are doubtless of no physiological significance, since they must have been inserted in their present locations rather recently, i.e., since the divergence of the genus *Strongylocentrotus* from other echinoid genera (≤30my ago; Smith, 1988). However, we found this repeat, containing these target sites, in the vicinity of three known genes (Anderson *et al.*, 1994). In one of these occurancees, viz, in the CyIIIa gene, the SpZ12-1 target site plays an essential regulatory role (Wang *et al.*, 1995), though we could find no role for the SpRFA-70 site, which can be deleted with no effect on expression (unpublished data). Perhaps in this study we are observing an early stage of the basic evolutionary process by which regulatory modules are transported by mobile elements into the cis-control regions of genes, thereby endowing them with new regulatory responses.

References

1. Anderson, R. F., 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
2. Britten, R. J. and Davidson, E. H. (1971). Repetitive and non-repetitive DNA sequences and a speculation on the origins of evolutionary novelty. The Quarterly Review of Biology **46**, 111-138
3. Davidson, E. H. (1986). Gene activity in early development. *3rd ed. Academic Press, Orlando FL.* ,
4. Kim, C., Snyder, R. O., and Wold, M. S. (1992). Binding properties of replication protein A from human and yeast cells. *Mol. Cell. Biol.* **12**, 3050-3059
5. Klein, W. H., Thomas, T. T., Lai, C., Scheller, R. H., Britten, R. J. and Davidson, E. H. (1978). Characteristics of individual repetitive sequence families in the sea urchin genome studied with cloned repeats. Cell **14**, 889-900
6. Moore, G. P., Scheller, R. H., Davidson, E. H. and Britten, R. J.(1978). Evolutionary change in the repetitive frequency of sea urchin DNA sequence. Cell **15**, 649-660
7. Smith, A. B. (1988). Phylogenetic relationship, divergence times, and rates of molecular evolution for camarodont sea urchins. *Mol. Biol. Evol.* **5**, 345-365
8. Wang, D. G.-W., Kirchhamer, .C. V., Britten, R.B., and Davidson, E. H. (1995). *SpZ12-1*, a negative regulator required for spatial control of the territory-specific CyIIIa gene in the sea urchin embryo. *Development* **121**, 1111-1122

Fig. 1 *Diagram illustrating the two regions of the SpRFA-70/SpZ12-1 repeat.* The gray area represents the region containing the SpRFA-70 binding site. The gray arrow above that section indicates the position and orientation of the 28 bp oligonucleotide probe used to detect that portion of the whole repeat. We know that the specific binding site for SpRFA-70 lies somewhere within the sequence of the probe. The black area represents the section containing the SpZ12-1 site. The black arrow below the SpZ12-1 region indicates the site and position of the 21 bp oligonucleotide probe used to detect this segment of the repeat. The exact location of binding for SpZ12-1 is indicated by the box. The same probes were used for PCR amplifications.

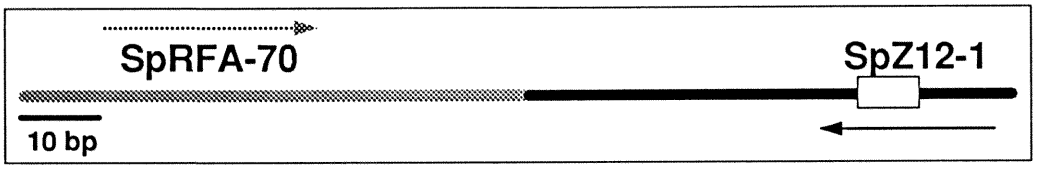
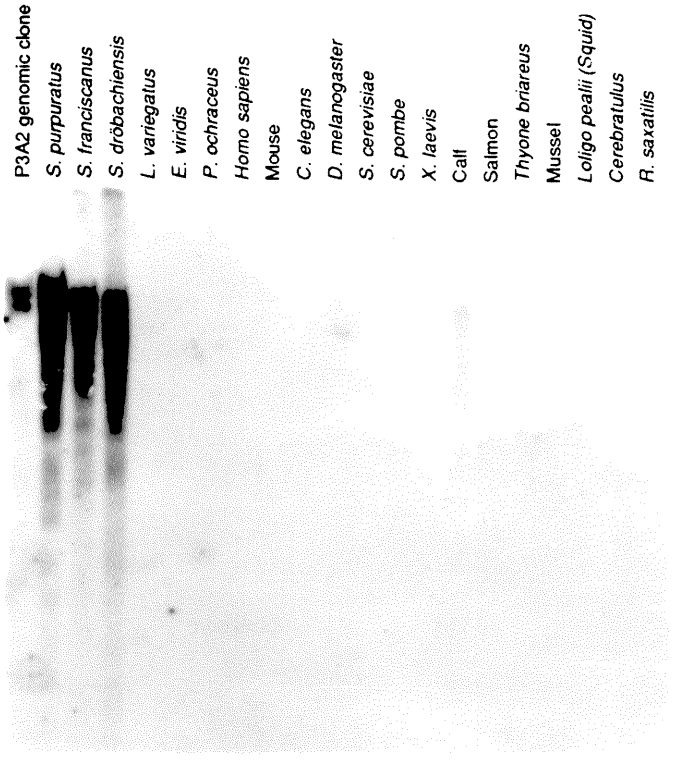
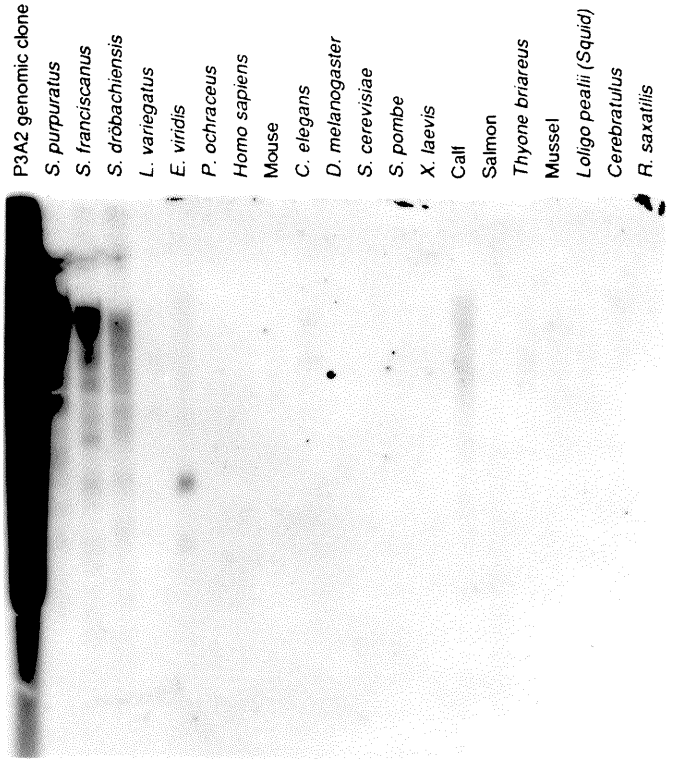

SpRFA-70
SpZ12-1
10 bp

Fig. 2. *Detection of SpRFA-70 and SpZ12-1 repeat copies in genomic DNA from multiple organisms- "Zoo Blot"*: 5 µg of BamH I-digested genomic DNA from the organisms indicated was electrophoretically separated on a 0.7% agarose gel and transferred overnight to a nylon membrane by acidic capillary blotting. The membrane was dried for 15 minutes at 80°C. 10-15 seconds of exposure to a long wave UV light source at a distance of 10 cm was used to cross-link the DNA to the membrane. In separate experiments probes were hybridized to the blot overnight, at 42°C for the *SpRFA-70* site oligo and at 37°C for the *SpZ12-1* probe. (a) Autoradiograph showing the blot after hybridization with the *SpRFA-70* probe and selective stringency washes to 61°C in tetramethylammonium chloride as described previously (Anderson *et al.*, 1994). These conditions allow for a mismatch of up to 3 base pairs within the 28 bp probe (~11%) (b) Autoradiograph showing the blot after hybridization with the *SpZ12-1* probe and selective stringency washes to 52°C in tetramethylammonium chloride. This allows for up to 3 base pair differences within the 21 bp probe (~14%).

(a)


P3A2 genomic clone
S. purpuratus
S. franciscanus
S. dröbachiensis
L. variegatus
E. viridis
P. ochraceus
Homo sapiens
Mouse
C. elegans
D. melanogaster
S. cerevisiae
S. pombe
X. laevis
Calf
Salmon
Thyone briareus
Mussel
Loligo pealii (Squid)
Cerebratulus
R. saxatilis

(b)


P3A2 genomic clone
S. purpuratus
S. franciscanus
S. dröbachiensis
L. variegatus
E. viridis
P. ochraceus
Homo sapiens
Mouse
C. elegans
D. melanogaster
S. cerevisiae
S. pombe
X. laevis
Calf
Salmon
Thyone briareus
Mussel
Loligo pealii (Squid)
Cerebratulus
R. saxatilis

Appendix 2

Competitive titration in living sea urchin embryos of regulatory factors required for expression of the CyIIIa actin gene

Roberta R. Franks, Roger F. Anderson, James G. Moore, Barbara R. Hough-Evans, Roy J. Britten and Eric H. Davidson

Development 110, 31–40 (1990)
Printed in Great Britain © The Company of Biologists Limited 1990

Competitive titration in living sea urchin embryos of regulatory factors required for expression of the CyIIIa actin gene

ROBERTA R. FRANKS, ROGER ANDERSON, JAMES G. MOORE, BARBARA R. HOUGH-EVANS, ROY J. BRITTEN and ERIC H. DAVIDSON

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

Summary

Previous studies have located some twenty distinct sites within the 2.3 kb 5′ regulatory domain of the sea urchin CyIIIa cytoskeletal actin gene, where there occur *in vitro* high-specificity interactions with nuclear DNA-binding proteins of the embryo. This gene is activated in late cleavage, exclusively in cells of the aboral ectoderm cell lineages. In this study, we investigate the functional importance *in vivo* of these sites of DNA–protein interaction. Sea urchin eggs were coinjected with a fusion gene construct in which the bacterial chloramphenicol acetyltransferase (CAT) reporter gene is under the control of the entire CyIIIa regulatory domain, together with molar excesses of one of ten nonoverlapping competitor *subfragments* of this domain, each of which contains one or a few specific site(s) of interaction. The exogenous excess binding sites competitively titrate the available regulatory factors away from the respective sites associated with the CyIIIa · CAT reporter gene. This provides a method for detecting *in vivo* sites within the regulatory domain that are required for normal levels of expression, without disturbing the structure of the regulatory domain. We thus identify five nonoverlapping regions of the regulatory DNA that apparently function as binding sites for positively acting transcriptional regulatory factors. Competition with a subfragment bearing an octamer site results in embryonic lethality. We find that three other sites display no quantitative competitive interference with CyIIIa · CAT expression, though as shown in the accompanying paper, two of these sites are required for control of spatial expression. We conclude that the complex CyIIIa regulatory domain must assess the state of many distinct and individually necessary interactions in order to properly regulate CyIIIa transcriptional activity in development.

Key words: sea urchin embryo, regulatory factors, *in vivo*, CyIIIa actin gene.

Introduction

The CyIIIa cytoskeletal actin gene of the sea urchin *Strongylocentrotus purpuratus* is transcriptionally activated late in cleavage in the presumptive aboral ectoderm territory of the embryo and, throughout embryonic development, the gene continues to be transcribed exclusively in the aboral ectoderm (Shott *et al.* 1984; Cox *et al.* 1986; Lee, 1986; Hickey *et al.* 1987). A necessary and sufficient *cis*-regulatory domain extending approximately 2.3 kb upstream of the CyIIIa transcription start site, which promotes correct temporal and spatial activation of a covalently associated CAT reporter gene, has been identified by gene transfer experiments (Flytzanis *et al.* 1987; Hough-Evans *et al.* 1987, 1988; reviewed by Davidson, 1989). The CyIIIa *cis*-regulatory domain includes about 20 sites where there occur *in vitro* high-specificity DNA–protein interactions. These have been located in gel shift mapping experiments with specific oligonucleotide competitors, using crude embryonic nuclear extracts (Calzone *et al.* 1988; Thézé *et al.* 1990). This set of sites includes 13 wholly non-homologous, distinct recognition sequences, some of which occur several times in the CyIIIa regulatory domain. There are unlikely to be additional binding sites in any of the fragments used in the competition experiments beyond those already characterized, that is, for factors that display a meaningful level of specific binding *in vitro*. Our objective in the present work was to obtain evidence *in vivo* of the functional significance of individual CyIIIa protein-binding sites, or of small subsets of locally contiguous sites. The approach chosen was based on the prior *in vivo* competition experiments of Livant *et al.* (1988). There it was shown that coinjection with the CyIIIa · CAT fusion of low (2–20×) molar excesses of the whole regulatory domain *per se* (i.e. lacking the CAT reporter gene) resulted in a near stoichiometric, quantitative decrease in reporter gene activity as a function of regulatory sequence:CAT reporter fusion gene ratio (molar excess, as used here, refers to the number of copies of competitor sites compared to

CyIIIa · CAT genes.) In the studies described in this and the accompanying paper, we coinjected restriction fragments containing *subregions* of the regulatory domain, together with a near saturating number of complete CyIIIa · CAT genes. The significance of individual target sites (or small sets of target sites) on these subfragments was determined by assessing the activity of the CAT reporter gene later in development. From this approach two forms of information can be derived. When the coinjected restriction fragments contain sites of positive DNA–factor interaction that *in vivo* are required for normal gene function, competition for these specific factors depresses the level of CAT activity measured in the transgenic embryos. In the accompanying paper, we show that ectopic spatial expression results from competition with other specific regulatory target sequences, implying that the interactions occurring in these sequences function *in vivo* to control spatial expression.

Materials and methods

Preparation and injection of S. purpuratus *eggs*

Unfertilized eggs were microinjected as previously described (McMahon *et al.* 1985). Briefly, the eggs were dejellied by exposure to pH 5.0 filtered sea water (FSW) and fixed electrostatically to protamine sulfate-coated dishes. Approximately 2 pl of a DNA solution was injected into the cytoplasm of each egg. After injection the eggs were fertilized and cultured at 15 °C, in the same dishes, in FSW containing penicillin (20 units ml^{-1}) and streptomycin (50 $\mu g\ ml^{-1}$).

CAT fusion genes and whole regulatory domain competitors

Two CyIIIa · CAT fusion genes were used in these experiments. The 14 kb CyIIIa · CAT construct (Flytzanis *et al.* 1987) contains about 7 kb of genomic DNA extending 5′ from the transcription initiation site of the *Strongylocentrotus purpuratus* CyIIIa cytoskeletal actin gene, plus the 5′ leader sequence of the CyIIIa primary transcript that is interrupted by a 2.2 kb intron (Akhurst *et al.* 1987). The fusion point is a *Sal*I site located 11 codons following the start codon of the CyIIIa · CAT message. An SV40 poly(A) addition site and plasmid sequences are also included. The CyIIIa · CAT construct was linearized for microinjection at a unique *Sph*I site located about 2.5 kb upstream of the transcription start site (see Fig. 1). A second fusion gene was created by in-frame subcloning the 9.5 kb *Sal*I fragment from CyIIIa · CAT (Fig. 1) into the CAT vector pUC.PL · CAT gene (Bond-Matthews and Davidson, 1988). The vector was first modified to create a *Sal*I site (R. Franks *et al.* unpublished; Sucov *et al.* 1988). This fusion is called pCyIIIa · CAT · 1. The 6.3 kb *Sph*I–*Kpn*I fragment from CyIIIa · CAT · 1 was gel purified and used for microinjection (Fig. 1). This fragment includes all necessary CyIIIa regulatory sequences.

Subfragments of the CyIIIa regulatory domain used as competitors

Subfragments that were used for the *in vivo* competition experiments were derived either directly from the CyIIIa · CAT construct, or from pUC 18/19 subclones of the CyIIIa regulatory region (Calzone *et al.* 1988; F. Calzone, R. Hill, E. Davidson, unpublished), or equivalent Bluescript subclones (A. Cutting and E. Davidson, unpublished) by digestion with the appropriate restriction enzymes (see legend to Fig. 3). The fragments were gel purified. Subfragments less than approximately 1.0 kb in length were ligated *in vitro* to sperm DNA (prepared as previously described; Lee *et al.* 1984). The sperm DNA was digested with an appropriate restriction enzyme to produce molecules of an average length of about 5.0 kb, the termini of which would be homologous with the competitor DNA termini, or blunt-ended if the competitor DNA molecule bore a blunt end. Ligation reactions were performed at 15 °C overnight in 20–50 μl that contained up to 10 μg of total DNA with an average molar ratio of competitor DNA fragments/carrier DNA fragments of about 5.0. The ligated DNA was digested with *Bgl*II, an enzyme for which none of the competitor DNA subfragments included a site, and molecules ranging in size from approximately 1 kb to 15 kb were obtained for microinjection by fractionation in a 1% low-melt agarose gel, followed by purification by the Geneclean (Bio 101 Inc.) procedure. Alternatively, the *Bgl*II-digested DNA was size-fractionated by gel filtration through Bio-Gel A-5 m (BioRad). The actual concentration of the specific competitor DNA sequence in the final sample used for injection was measured by slot-blot hybridization, using a single-stranded RNA probe. Known numbers of linearized CyIIIa · CAT DNA molecules were included on each filter as standards. Typically, the competitor DNA sequence was present at an average density of about 3.4 copies per 5 kb of carrier DNA in the sample (all samples used fell within the range 2–6 copies per 5 kb carrier DNA except the P7I ligation, which was injected at 1 copy per 5 kb). A few deviations are noted in the legend to Fig. 3.

Microinjection solutions

Each DNA solution was made up in 40% glycerol and contained 550–2000 molecules of a CAT fusion gene per pl, together with competitor DNA and/or carrier DNA (see text). The DNA mass in each injection solution was adjusted by the addition of carrier to yield at least a 5-fold mass excess of carrier DNA plus competitor DNA, over CAT DNA, and to ensure that each injection sample in a given experiment contained the same total DNA concentration.

Measurement of CAT enzyme content per embryo

Embryos were analyzed for CAT enzyme content at 20–24 h postfertilization (hatched blastula) as previously described (McMahon *et al.* 1984), using lysates derived from 40–60 injected embryos and a series of bacterial CAT enzyme standards. Under our conditions, one unit of CAT enzyme activity equals $\sim 2.6 \times 10^{11}$ molecules of enzyme, as established earlier by reference to a purified recombinant CAT enzyme preparation (McMahon *et al.* 1984).

Measurement of CAT DNA and competitor DNA content per embryo

DNA measurements were made at 24 h postfertilization by a modification of the slot–blot procedure described earlier (Flytzanis *et al.* 1987; Franks *et al.* 1988). Centrifugal pellets containing 60–130 injected embryos or uninjected control embryos were mixed in 50 μl of 250 mM Tris–HCl, pH 7.8 with ~100 gastrula stage *Lytechinus variegatus* embryos used as carrier. An equal volume of 0.1 M EDTA, pH 8.0, 0.2 M NaCl, 1% SDS was added, and the samples were incubated with 20 μg of proteinase K for 3 h at 55 °C. The nucleic acids were extracted and treated with 0.4 M NaOH at 65 °C to denature the DNA and hydrolyze the RNA, and were filtered onto

nitrocellulose using a slot-blot apparatus as previously described. For hybridization standards known numbers of CyIIIa · CAT DNA molecules, linearized with *Sph*I, were mixed with *L. variegatus* carrier nucleic acids, treated with 0.4 M NaOH as described above, and applied to each filter. These standards of course include all of the different competitive sequences, since the whole regulatory domain is present in CyIIIa · CAT (see Fig. 1) as well as the CAT gene sequence. The slots were cut exactly in half, and the halves were separately hybridized with single-stranded RNA probes containing sequences representing either the CAT gene (Flytzanis *et al.* 1987) or the CyIIIa 5′-sequence used for *in vivo* competition. Probe specific activities were about 1.3×10^9 cts min^{-1} μg^{-1}. Hybridizations with the CAT probe were carried out at 40°C in 50% formamide as described (Franks *et al.* 1988). Hybridizations with the CyIIIa 5′-competitor sequence probes were performed at temperatures calculated to maximize hybridization efficiency based on G-C content (Casey and Davidson, 1977). The filters were washed twice with 2×SSC (1×SSC is 0.15 M NaCl, 0.015 M sodium citrate), 0.2% SDS at room temperature, twice with 2×SSC, 0.2% SDS at 68°C, and twice with 0.75×SSC, 0.2% SDS at 68°C. Following autoradiography, each half-slot was cut out and counted; quantitation of CAT DNA sequences or CyIIIa 5′-competitor DNA sequences in each half-slot was made by comparison to the signal obtained for hybridization between each probe and the CyIIIa · CAT DNA standards on each filter. The competitor DNA sequence probes also hybridized with the incorporated CyIIIa · CAT sequences present in each half-slot of DNA extracted from the injected embryos. Therefore, the number of competitor DNA molecules in excess of CyIIIa · CAT genes in a given half-slot could be determined by subtracting the number of CyIIIa · CAT DNA molecules that was calculated to be present (using the corresponding half-slot hybridization signal obtained with the CAT probe) from the total number of sequences that hybridized with the competitor DNA probe. Thus the molar ratio of incorporated competitor DNA:CyIIIa · CAT DNA in the transgenic embryos was obtained. The number of CAT DNA molecules and competitor DNA molecules present per half-slot was also computed in terms of average numbers of molecules per embryo.

Competitive efficiency coefficient α

The competitive *in vivo* titrations carried out with the whole regulatory domain by Livant *et al.* (1988) indicated that the efficiency of competition *in vivo* was about 44% of ideal expectation. This was established by deriving a competition efficiency coefficient, α, given the form $y=(1+\alpha x)^{-1}$, where y is the CAT activity relative to the control, in which CyIIIa · CAT but no competitor DNA is introduced into the embryos; and x is the molar ratio of competitor to CyIIIa · CAT DNA actually incorporated in the experimental embryos. Were $\alpha=1$, the factor binding sites included in the incorporated CyIIIa · CAT genes, and those in the incorporated competitor regulatory domains, would have competed with equal efficiency for the limiting DNA binding factors. The microinjection gene transfer system utilized in both the experiments of Livant *et al.* (1988) and those described in this paper, produces a mosaic, though random, pattern of exogenous DNA incorporation (McMahon *et al.* 1985; Flytzanis *et al.* 1985; Hough-Evans *et al.* 1987; Franks *et al.* 1988; see Livant *et al.* 1988). Thus, a probable explanation for the ~2-fold discrepancy from ideal behavior in the experiments of Livant *et al.* (1988) is that the cells incorporating and expressing CyIIIa · CAT in their genomes are not invariably the same cells that incorporate all the competitor sequences.

Results

Competitive titration using the entire CyIIIa regulatory domain

Previously published gene transfer studies demonstrating that correct spatial and temporal expression is mediated by upstream CyIIIa sequence were carried out with the linearized CyIIIa · CAT fusion (Flytzanis *et al.* 1987; Hough-Evans *et al.* 1987, 1988), and this construct was also used in some of the experiments we report here. A second construct, pCyIIIa · CAT · 1, utilized in other experiments, is shown diagrammatically in Fig. 1A. Expression of the 6.3 kb *Sph*I–*Kpn*I fragment of pCyIIIa · CAT · 1 in transgenic embryos is indistinguishable from that of CyIIIa · CAT in quantitative, temporal, and spatial respects. On microinjection into sea urchin egg cytoplasm, linear DNA molecules of these lengths are rapidly ligated, irrespective of the DNA termini, into long end-to-end concatenates that are stably incorporated into one or more blastomere nuclei during cleavage, and are replicated together with host cell DNA throughout embryogenesis (McMahon *et al.* 1985; Flytzanis *et al.* 1985). No differences were detected in the degree of amplification of any of the fusion genes or competition fragments within experiments.

In an initial series of experiments, fragments containing the complete set of detectable protein-binding sequence elements were used as *in vivo* competitors in order to provide base line stoichiometry for comparison with the competitive effects of individual subfragments. In Fig. 2 we present new data that substantiate the observation (Livant *et al.* 1988) that quantitative competitive titrations can be obtained *in vivo* in sea urchin embryos. This Fig. shows the pooled 'whole regulatory domain' controls for the subfragment competition experiments presented in the following. For each point in Fig. 2, at least 1200 molecules of the CyIIIa · CAT reporter, plus the indicated numbers of competitive DNA molecules (9.5 kb *Sal*I or 4.3 kb *Nco*I–*Eco*RI fragments, see Fig. 1) were injected. At least a 5-fold mass excess of carrier and/or competitor DNA was coinjected with the CyIIIa · CAT genes. The carrier used was sea urchin sperm DNA, cut to a mean length of 5 kb. Thus the individual CyIIIa · CAT genes in the incorporated concatenates were distributed around a mean distance apart of 30–70 kb (depending on whether the 14 kb CyIIIa · CAT or the 6.3 kb *Sph*I–*Kpn*I fragment was used), interspersed among randomly selected sequences; or, in the competition samples, among various ratios of CyIIIa regulatory domain and 5 kb randomly selected sequence. Preliminary experiments had shown that the CyIIIa · CAT gene functions 2–3× more efficiently when separated by ≥3.5-fold excess of carrier (possibly due to transcriptional interference in the absence of carrier). Within each experimental series of Fig. 2, the various competition samples and the reference sample, in which CyIIIa · CAT and carrier but no competitor had been introduced, all received exactly the same total mass of DNA per egg; i.e. the carrier DNA content was adjusted to take into account the

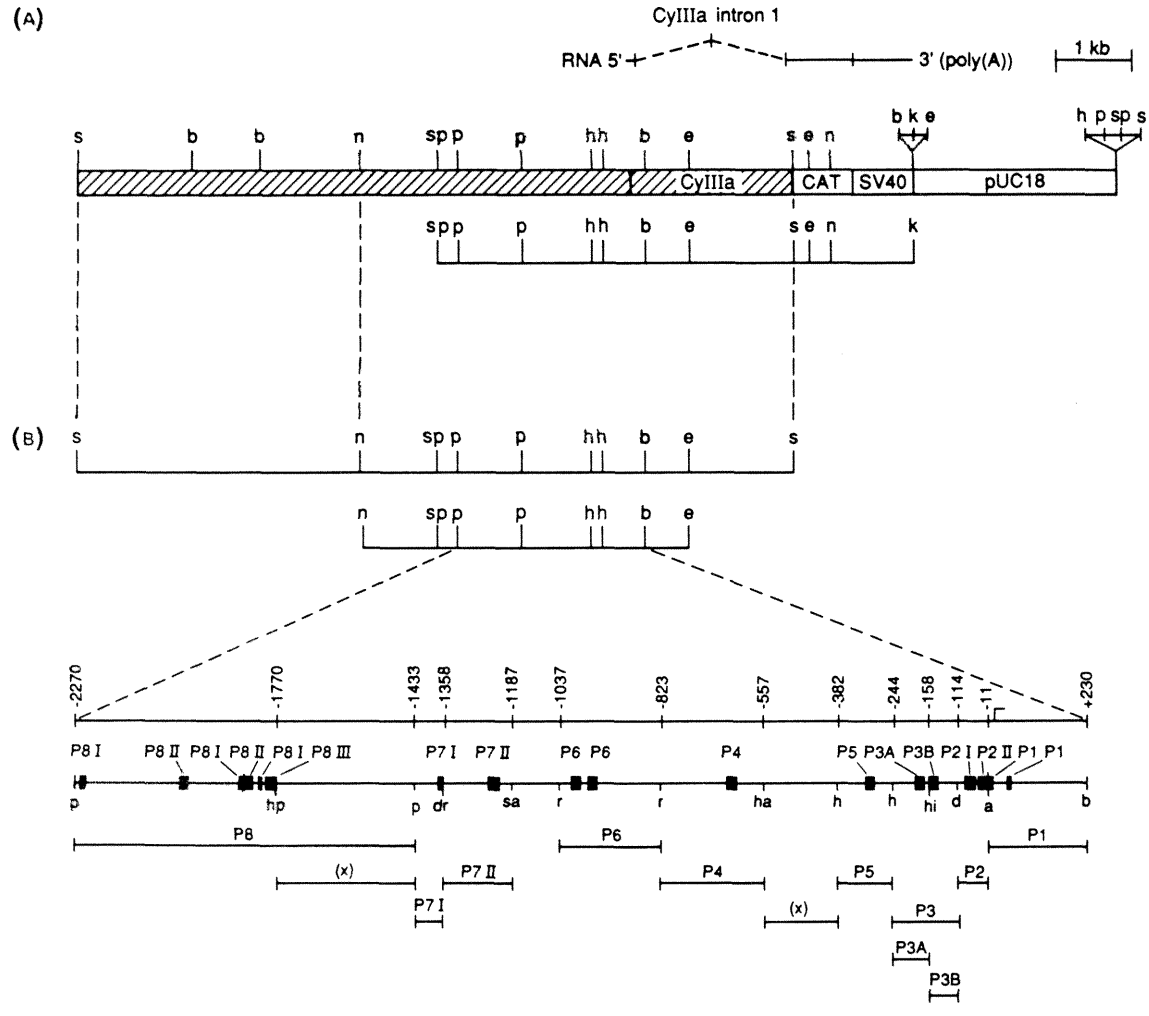


Fig. 1. Structure of the pCyIIIa · CAT · 1 gene fusion, locations of sites of *in vitro* interaction between nuclear factors and CyIIIa 5′-flanking sequence, and competitor DNA fragments. (A) Map of pCyIIIa · CAT · 1. Hatched areas represent sea urchin sequences, including the 2.2 kb first intron contained within the CyIIIa primary transcript (Akhurst *et al.* 1987). The fusion point is a *Sal*I site located 11 codons following the start codon of pCyIIIa · CAT · 1. An SV40 poly(A) addition site is included, together with plasmid sequences. The 6.3 kb *Sph*I–*Kpn*I fragment derived from the pCyIIIa · CAT · 1 fusion gene was used for microinjection. Alternatively, the CyIIIa · CAT fusion gene (not shown) described by Flytzanis *et al.* (1987) was linearized at the *Sph*I site located 2.5 kb upstream of the CyIIIa transcription start site, and used for microinjection. (B) Map of the CyIIIa *cis*-regulatory domain, and competitor DNA fragments. The locations of all known sites of *in vitro* interaction between sequence elements in the 5′-flanking regulatory region of the CyIIIa gene and nuclear factors are depicted by black boxes within the 2.5 kb *Pst*I–*Bam*HI fragment. The protein–DNA interactions described by Thézé *et al.* (1990) for each binding site are indicated (P1–P8). The P1 binding region actually extends into the 5′ end of the first exon of the CyIIIa gene. The 9.5 kb CyIIIa *Sal*I fragment and the 4.3 kb *Nco*I–*Eco*R1 fragment (indicated above the map), each of which contains the entire regulatory domain, were used as competitor in the *in vivo* competition experiments shown in Fig. 2. DNA subfragments of the regulatory region (shown below the map) that contain the indicated factor-binding site(s) (or no binding site, x) were released from pUC subclones (Calzone *et al.* 1988) by digestion from the polylinker with *Eco*R1 and *Hin*dIII (P1–P3, P4–P6, x) or *Hin*dIII (P5), or from pBluescript subclones (A. Cutting, E. Davidson, unpublished data) by digestion with *Bam*HI and *Hin*dIII (P3A, P7I). Fragment P3B was derived from fragment P3 by digestion with *Hin*fI. These fragments were ligated to sea urchin carrier DNA molecules bearing termini homologous with the specific fragment and prepared for microinjection as described in Materials and methods, and were subsequently used for the *in vivo* competition experiments shown in Fig. 3 and Table 1. Restriction sites shown are: a, *Ava*II; b, *Bam*HI; d, *Dde*I; dr, *Dra*I; h, *Hin*dIII; ha, *Hae*III; hi, *Hin*fI; hp, *Hpa*II; k, *Kpn*I; n, *Nco*I; p, *Pst*I; r, *Rsa*I; e, *Eco*RI; s, *Sal*I; sp, *Sph*I.

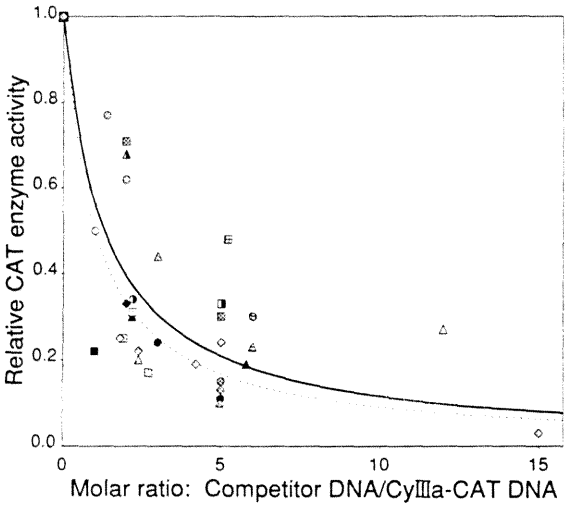


Fig. 2. Competition *in vivo* between the CyIIIa 5′-regulatory domain and coinjected CyIIIa · CAT reporter genes. At least 1200 molecules of a CyIIIa · CAT gene fusion were injected at each point, together with varying numbers of either the 9.5 kb *Sal*I DNA fragment or the 4.3 kb *Nco*I–*Eco*RI DNA fragment, containing all known sites of *in vitro* interaction between the CyIIIa 5′-regulatory region and nuclear factors (see Fig. 1B) as competitor. The molar ratio of competitor sequence to CyIIIa · CAT sequence incorporated at each point was measured by slot-blot hybridization of DNA extracted from samples of pooled transgenic embryos, using ^{32}P-labeled probes specific for each sequence, as described in Materials and methods. A second aliquot of embryos was pooled and assayed for CAT activity at each point. The relationship between CAT activity per embryo relative to the CAT activity per embryo (normalized to 1.0) in the control in which no competitor DNA was included (y), and molar ratio, (x), of competitor DNA molecules to CyIIIa · CAT genes, can be approximated as $y=(1+\alpha x)^{-1}$, where α is the competition efficiency coefficient, and provides an estimate of the fraction of competitor fragments actively functioning as competitor (Livant *et al.* 1988; see legend to Table 1). The dashed curve indicates the ideal form, i.e. where $\alpha=1.0$. The solid curve is the least squares fit to the experimental data, and yields a value for α of 0.76 [Livant *et al.* (1988) obtained $\alpha=0.44$ in similar experiments; considering the scatter there may be no real difference between these results, nor any real discrepancy from an α value of unity]. Twenty-six experiments are shown. Each unique symbol represents an experiment carried out with a single batch of eggs. CyIIIa · CAT gene fusion and competitor DNA fragment used are: solid symbols (black or white), 6.3 kb *Sph*I–*Kpn*I pCy111a · CAT · 1 fragment plus 9.5 kb *Sal*I competitor DNA fragment; cross-hatched symbols, *Sph*I-linear CyIIIa · CAT (Flytzanis *et al.* 1987) plus 9.5 kb *Sal*I competitor DNA fragment; right-shaded symbols, *Sph*I-linear CyIIIa · CAT plus 4.3 kb *Nco*I–*Eco*RI competitor DNA fragment.

mass of competitor DNA added to obtain the nominal competitor/fusion gene molar ratio desired. The total DNA mass injected was about 0.2 pg/egg.

As in our previous studies (op cit) average CAT enzyme content per embryo in the reference samples that contained no competitor was generally $1.0–3.0\times10^7$ enzyme molecules per embryo. To pool the data shown in Fig. 2 these values have been used for normalization of the CAT enzyme content in the embryos within each experimental series. The CAT enzyme was measured in samples of 24 h embryos, and in Fig. 2 is expressed as a function of the molar ratio of competitor to CyIIIa · CAT genes actually incorporated per average embryo. These ratios were directly measured in aliquots of the same embryo samples by slot-blot hybridization, using competitor DNA and CAT DNA sequence probes. The experimental data fit by the solid line in Fig. 2 indicate that CAT activity falls off at a rate that is almost stoichiometric, with respect to the number of competing 5′ regulatory sequences present. The best fit value from the pooled data of Fig. 2 yields $\alpha=0.76$ (see legend of Fig. 2). Thus we have used $\alpha=0.76$ as a normalization factor in most of the following experiments. These data clearly confirm that endogenous factors required for the expression of the CyIIIa · CAT fusion gene can be titrated away from the fusion gene by introduction of excess DNA binding sites.

Competitive titration in vivo *by subelements of the CyIIIa regulatory domain*

To determine the effects on CyIIIa · CAT gene activity of titrating away those factors binding to individual subregions of the large regulatory domain, we coinjected various molar excesses of the subfragments identified in Fig. 1B. Many of these subfragments were also utilized for the gel-shift experiments of Calzone *et al.* (1988) by which the sites of interaction were initially located, and we have adopted the same subfragment nomenclature in the present account. Several include only a single site of DNA–protein interaction. However, the P1, P2 and P6 subfragments contain two sites of interaction, and the P8 subfragment, which is particularly complex, includes at least 8 protein-binding elements of three different classes (Thézé *et al.* 1990). Two subfragments that do not contain sequences that form DNA–protein complexes *in vitro* on reaction with embryo nuclear extracts are also shown on the map (X in Fig. 1B). Neither of these subfragments displayed any competitive function in our hands (results not shown), nor do subfragments extending 5.3 kb upstream of the *Sph*I site (see Fig. 1A). In initial studies we found that microinjection of small DNA fragments (<1 kb) does not reliably result in measurable replication during embryogenesis, probably due to failure to form concatenates that are sufficiently large to promote stable incorporation. To circumvent this problem each of the competitor DNA subfragments shown in Fig. 1B, with the exception of the large subfragment containing the whole P8 binding region, was first ligated *in vitro* to carrier genomic DNA. The actual concentration of the specific competitor sequence in each ligated DNA

preparation was measured by slot-blot hybridization, using a ^{32}P-labeled RNA probe.

In each injection series, eggs obtained from a single female were coinjected with at least 1200 molecules of a CyIIIa · CAT gene, together with increasing quantities of competing sequence that were always in molar excess with respect to the CyIIIa · CAT DNA. Competition curves obtained *in vivo* for each of nine nonoverlapping subfragments that were derived from the CyIIIa regulatory domain are shown in Fig. 3. The average CAT enzyme content and the average competitor DNA:CyIIIa · CAT DNA ratio were measured in pooled embryo samples at 24 h. The average exogenous DNA amplification in the various injection samples *within* a given experiment was similar, and in different batches the values measured in the 24 h embryos ranged from about a 10-fold to an 80-fold increase over the amount of DNA injected per egg. Normalized experimental data were fitted to the same form describing the competition stoichiometry as above, and for each experiment a value for the competition efficiency coefficient α was derived by nonlinear least squares. In Fig. 3, competition functions are shown for each of 10 subfragments that contain known sites of DNA–protein interaction identified *in vitro*. These are presented in order of descending α values; i.e. the best competitors first, and the least effective last. The curve representing perfect competition stoichiometry ($\alpha=1$) is shown as a dotted line in each panel for comparison. Fig. 3A–F demonstrates immediately that a majority of the nonoverlapping subfragments tested function individually as competitors against the whole CyIIIa regulatory domain, and the P4, P5 and P8 subfragments *each* appears to compete as well as does the whole regulatory domain. Each of these regions apparently contains an essential positive regulatory site(s) bound by a factor or factors of limited availability relative to the number of competitor fragments incorporated. Fig. 3G–J shows that subfragments containing the binding regions for P6, P3, P3A and P7II, respectively, displayed relatively insignificant competitive activity.

Quantitative data for these and some additional experiments are listed in Table 1. The values given by the parameter α' in the second column of Table 1 estimate the competitive effect of the indicated subfragment in each experiment, relative to the competition obtained with the *complete* regulatory region. Pooled data for multiple experiments with given subfragments were used to assess the best overall value of the competition coefficient, given as α'' in the third column of Table 1. Each of five different regions, i.e. those containing binding sites for P2, P4, P5, P7I and P8, compete between 10 % and 100 % (or better) than does the whole regulatory domain (i.e. $\alpha' \geq 0.1$). An α' value of 1.0 would indicate that the competitive effect of the subfragment is equal to that of the fragment containing the entire regulatory domain. The α' values for DNA subfragments containing the binding sites for P3, P3A, P6 or P7II are below 0.1, and that for P1 is just above 0.1. As discussed below, this result is unlikely to be due simply to high prevalence of the respective factors.

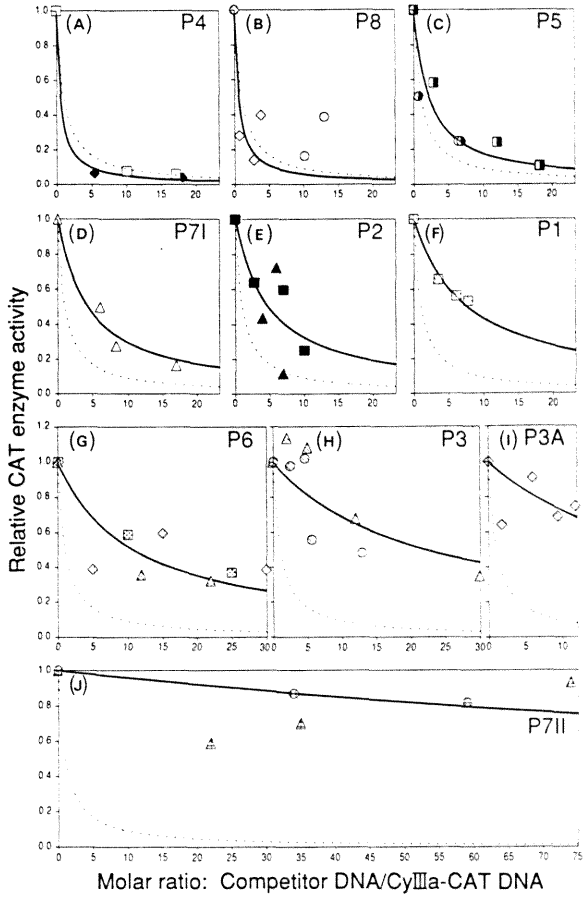


Fig. 3. *In vivo* competition by subelements of the CyIIIa 5′-regulatory region in transgenic embryos. At least 1200 molecules of a CyIIIa · CAT gene fusion were injected at each point, together with varying numbers of competitor fragments that contain the binding regions for the indicated set of DNA-binding factors (Thézé *et al.* 1990; see Fig. 1B). The relationship between average CAT activity per embryo and the average molar ratio of competitor DNA molecules to CyIIIa · CAT molecules incorporated is expressed as in Fig. 2. The competition coefficient α (i.e. where ideal stoichiometric competition would generate a value for α of 1.0) calculated for the whole regulatory domain in Fig. 2 was used as a normalization factor to correct the individual data points in these experiments (see legend to Table 1), so that the competitive effect of the individual fragments would be compared directly to the effect of the whole domain. Thus, the dashed curves in A–J indicate the competitive behavior of the entire regulatory region. The solid curves, fit by least squares, indicate the behavior of the individual subfragments. Competition is near stoichiometric in A–C. There is little, if any, competition in F–J. Panels A–C, E, G, H, J show the least squares fit to the pooled experimental data obtained from more than one experiment. The data for the eighteen experiments shown here were derived from a subset of those experiments used to generate the value of α for the whole regulatory domain shown in Fig. 2 (data points from a single experiment are described by matching symbols in these figures).

Table 1. *Competitive effect* in vivo *of subfragments of the CyIIIa 5′-regulatory domain*

Competitor CyIIIa subfragment*	Experiment	α'†	α''†
P2	1	0.17 (0.11)	0.21 (0.19)
	2	0.25‡ (0.24)	
P4	1	2.5 (0.01)	1.8 (0.02)
	2	1.1 (0.1)	
P5	1	0.29 (0.04)	0.47 (0.12)
	2	0.92 (0.10)	
P7I	1	0.24 (0.7)	
P8	1	0.24 (0.14)	1.7 (0.22)
	2	2.2 (0.18)	
P1	1	0.13 (0.03)	
P3	1	0.02§ (0.04)	0.04 (0.18)
	2	0.14§ (0.09)	
	3	0.04 (0.17)	
	4	0.06 (0.17)	
P3A	1	0.04 (0.16)	
P6	1	0.09‡ (0.21)	0.09 (0.15)
	2	0.07‡ (0.00)	
	3	0.12 (0.05)	
P7II	1	0.01 (0.23)	0.04 (0.17)
	2	0.003 (0.06)	

* Each competitor subfragment is derived from the entire CyIIIa 5′-regulatory domain, and contains the binding regions for the sets of DNA-binding factors indicated in Fig. 1B (Thézé *et al.* 1990; Calzone *et al.* 1988).

† The relative activities of the CyIIIa · CAT reporter gene in the presence of molar excesses of the indicated competitor DNA subfragment in each experiment were normalized, usually by the value for the competition efficiency coefficient estimated from the pooled control data shown in Fig. 2. This value was $\alpha=0.76$. To carry out the normalizations the following procedure was used: y_i is the CAT activity measurement (average molecules CAT enzyme/embryo in the competition sample, relative to the average molecules CAT enzyme/embryo in the control sample in which no competitor DNA was included; see Materials and methods); x_i is the molar ratio [subfragment:CyIIIa · CAT] actually incorporated, by direct measurement (see Materials and methods); N is the competition efficiency coefficient indicating the fraction of competitor functioning in a manner quantitatively equivalent to the respective site(s) on the incorporated CyIIIa · CAT molecules. In general for this competition, $y=(1+\alpha x)^{-1}$ (Livant *et al.* 1988). The normalized values denoted $y(c:x_i)$ are calculated as follows:

$$y(c:x_i) = y_i \cdot \frac{y(1:x_i)}{y(N:x_i)}$$

where N is the value of α used for the normalization, and $y(1:x_i)+(1+x_i)^{-1}$; $y(N:x_i)=(1+Nx_i)^{-1}$. As noted above N was usually 0.76, except for a few cases, indicated (see below) in which there were sufficient internal control data, so that N was derived from that experiment rather than from the pooled data. The resulting data reduction in these cases was in fact never significantly dependent on whether the internal control or the pooled value 0.76 were used for the normalization. Internal controls were routinely used as a qualitative screen for occasional batches of eggs in which for some reason even the control competition functioned not at all or very poorly, probably because these eggs did not adequately support exogenous DNA amplification. The normalized $y_c(x_i)$, x_i values were fit to $y_c(x_i)=(1-\alpha' x_i)^{-1}$ by least squares, and the values of α' are those reported in the Table for the individual subfragment experiments. For the last column of Table 1 all of the $y_c(x_i)$ values were pooled, the data were refit, and an α'' value established by least squares. Root mean square errors are reported in parentheses.

‡ For these experiments, we used individually determined efficiency coefficients to obtain α', rather than 0.76 (i.e., individually determined values of N, as in note †; here the value for α for the CyIIIa · CAT control with the whole regulatory domain competitor in each of these particular experiments). For P2, experiment 2, N=0.96; for P6, experiment 1, N=0.73; P6, experiment 2, N=0.31.

§ Measurements of CAT DNA and competitor DNA content per embryo were not made in these experiments. The values calculated for α' were estimated based on the subfragment/CAT DNA molar ratios introduced into each egg. These experiments were not utilized in calculating α''.

Lethal effects of competitive titration of the P3B binding site

The site labeled P3B in Fig. 1 includes a canonical octamer binding site (Thézé *et al.* 1990). Uniquely among all the subfragments tested, competition with a subfragment containing only this site results in the death of 60 % to over 95 % of the recipient embryos, by the pluteus stage, in different experiments. In controls for these particular experiments, 95–100 % of eggs injected only with a saturating number of CyIIIa · CAT molecules plus carrier DNA developed through hatching blastula stage, and 70–80 % completed embryogenesis normally. The same P3B injection solution that is lethal to recipient embryos is rendered harmless if treated with DNase prior to injection.

The unusual lethality of P3B competition is particularly interesting because as shown earlier the adjacent P3A site, when injected on a fragment by itself, does not display competitive activity (Fig. 3I; Table 1), and when the *Hin*dIII–*Dde*I 'P3' subfragment, which includes *both* the P3A and P3B sites (see Fig. 1A) is injected, the results are the same as when the P3A subfragment alone is injected (see Table 1). That is, there is no significant competitive activity (α' is about 0.04 for both subfragments), and no lethality is observed. The P3B factor is evidently required for the expression of genes that are necessary for viability, a not unlikely possibility since octamer binding sites have been reported in a great many different genes in mammals, and in several sea urchin histone genes as well (Barberis *et al.* 1987). That the presence of the P3A site on the same competitor fragment with the P3B site 'rescues' the embryos from the lethal effects of P3B competition suggests that, when P3A is bound at the adjacent sites, the essential P3B octamer factor is precluded from binding, and thus is not sequestered away on the competitor fragments.

Discussion

Competitive titration of positively acting factors within the embryo nuclei

Livant *et al.* (1988) showed that as the number of CyIIIa · CAT genes per embryo nucleus is increased, over a range from a few to several thousand, the amount of CAT enzyme expressed attains a *plateau*, after an initial phase in which CAT activity rises proportionally with incorporated gene number. The kinetics of the response can be fitted to the expected form for a saturation phenomenon. This is an interesting result which it is necessary to consider briefly before taking up the nature of the competitions that we present in this paper. About 90 % saturation of CyIIIa · CAT expression occurs at 1050 specific sites per nucleus, implying that the limiting factors required for CyIIIa · CAT expression are 90 % bound to specific sites at this level (D. Livant and E. Davidson, unpublished). In the present experiments, we tried to approach saturation conditions by injecting ≥1100 molecules of CyIIIa · CAT per egg, *plus* in the competition samples, molar

excess of the individual sites being tested. On the saturation plateau, the excess sites compete with one another for factor molecules. As the site concentration increases in the nuclei, these factors progressively repartition from the non-specific sites with which they are reversibly associated to the added specific sites, according to the particular values of the partition function K_r that obtain for each factor *in vivo* (K_r is the ratio of equilibrium constants for reaction of a given factor with its specific sites to that for reaction with non-specific DNA sites). Cooperative binding amongst the many different factors required for productive expression of CyIIIa · CAT, or additional stabilization of active transcription complexes, is also implied. Otherwise, as excess copies of the regulatory domain are introduced, the factors would distribute independently amongst them. The number of completely loaded regulatory domains would thus decrease, and a decline in activity rather than the saturation plateau observed convincingly by Livant *et al.* (1988) would result.

The experiments shown in Fig. 2 confirm the evidence of Livant *et al.* (1988) that when the whole regulatory domain is introduced in molar excess with respect to saturating levels of CyIIIa · CAT, competition for necessary factors occurs *in vivo*. This is reflected by an almost stoichiometric decrease in CAT expression, as a function of competitor:CyIIIa · CAT molar ratio. When similar competition is observed following introduction of molar excesses of a given site, as in the experiments of Fig. 3, these competitor sites therefore must effectively titrate away the factor species that recognizes them specifically. However, we have not shown explicitly that the competitive effects are direct. For example the striking P5 competition shown in Fig. 3C could be an indirect result, were the sequestration of the P5 factor actually to affect some regulatory gene, that in turn produces some other factor that affects CyIIIa · CAT expression. The improbability of such an interpretation depends on the frequency with which factors required for regulation of the CyIIIa gene are also required for regulation of genes whose products are necessary for CyIIIa expression, e.g. these same factors, or polymerase components, aspects of processing mechanism, etc. Needless to say, that frequency cannot yet be assessed. But, it is worth noting that in two cases where there is relevant deletion evidence, the same result is obtained by both deletion and competition methods. Flytzanis *et al.* (1987) showed that severe depression of CyIIIa · CAT expression results from excision of the P8 sites (their ΔSP, ΔSPP deletions), or of the P5 site (their ΔH deletion, in which exactly the same fragment was removed from CyIIIa · CAT as was used for the experiments of Fig. 3C). It follows that the strong competitions shown for the P5 and P8 sites in Fig. 3B and 3C are indeed direct rather than indirect effects. More generally, however, it is useful to recognize that excess site competition, and site deletion followed by gene transfer, are complementary rather than redundant ways of probing the significance of regulatory interactions *in vivo*. Deletion does indeed reveal site function directly, but in the context of more or less disturbed *cis*-regulatory DNA structure. While this may usually be of secondary or no importance, interaction between bound factors could depend in given cases on regulatory domain structure. In one example, a cardiac myosin light chain gene in which the relevant sites are closely apposed, intranuclear competitive titration yielded consistent results where deletion did not (Braun *et al.* 1989). Recently competitive titration has been used in studies of the regulatory molecular biology of sea urchin embryo histone genes (Lai *et al.* 1989), human actin genes (Muscat *et al.* 1988) and a gut-specific esterase gene of *C. elegans* (McGhee, personal communication).

Identification of positively acting CyIIIa regulatory sites

The surprising result from the experiments shown in Fig. 3 and Table 1 is that out of ten subfragments tested five include sites that compete effectively at low molar excess with respect to CyIIIa · CAT expression. These are, in order of their competitive potency, the subfragments containing the P4, P8(I+II+III), P5, P7I and P2(I+II) sites. In addition, the P1 factor may compete weakly and P3B octamer factor is clearly necessary, though its role in CyIIIa gene regulation remains unspecified, since its sequestration in a minor fraction of cells is lethal to the embryos. The positively acting factors listed here are probably entirely different and independent of one another, since the sequences to which they bind are wholly distinct (Thézé *et al.* 1990), and *in vitro* measurements indicate that these factors are of different prevalences (Calzone *et al.* 1988). Furthermore, these particular factors accumulate in the embryo with different kinetics (Calzone *et al.* 1988), and interact with different unique sets of genes (Thiebaud *et al.* 1990). The three most potent competitors are those subfragments bearing sites for factors P4, P8(I+II+III), and P5. In all of these cases, the individual subfragment competes within a factor of two as efficiently as does the whole regulatory domain. Considering the statistical nature of these observations, and the failure of the whole domain to compete ideally (α=0.76; Fig. 2), it seems reasonable to conclude that the competitive efficiencies of these subfragments cannot be distinguished from that of the whole regulatory domain. Thus, each of these subfragments must be the site of absolutely required, positive interactions, without which CyIIIa · CAT activity is minimal. The prevalence of the P4, P8 and P5 factors in the embryo nuclei is close to that of the CyIIIa · CAT gene introduced in our competition experiments (i.e. ~500 molecules/nucleus of P4 and P5, and about 1500 of the P8 factors). This prevalence will be rapidly exceeded as molar excesses of competitor sites are introduced. P4 is a CCAAT factor and the essentiality of this interaction is thus not surprising, though it may be noted that there are probably several different CCAAT binding factors in sea urchin embryos (Barberis *et al.* 1987). Although molar excesses of sites should also have been attained *in vivo* for the relatively rare P7I factor (~200 copies/

nucleus) and for P1 (~1100 copies/nucleus), competition in these cases is only around one-fourth to one-sixth as efficient as with the whole domain. Unless the P1 and P7I factors are distributed to only a small subset of cells, in which their concentrations per nucleus would be 5–10× higher, it can be concluded that these interactions are not individually absolutely required for CyIIIa · CAT function, as are those mediated by the sites on the P4, P8 and P5 subfragments. Yet they are necessary for maximum activity.

There can be two reasons for failure of competition when competition sites are introduced: factors with relatively low K_r will not partition to the added sites efficiently; and for factors that are very prevalent it is more difficult to introduce a true excess of specific sites. High endogenous factor prevalence (~6000 copies/nucleus) could easily explain the relatively weak competition displayed by the P2 subfragment. The number of CyIIIa · CAT genes plus competitor would only begin to exceed the number of P2I and P2II factor molecules at the highest molar excesses utilized (cf. Fig. 3E). Thus it cannot be excluded that one or both of the P2 factors is not as essential as at least one of the P8 and the P4 and P5 factors; that is, that the efficiency of P2 subfragment competition is depressed to about one-fifth of the competition efficiency of the whole domain only because of failure to achieve site/factor ratios equivalent to those in the other experiments shown. It is not feasible to introduce greater masses of DNA without incurring risk of embryo lethality irrespective of the sequence injected (McMahon *et al.* 1985). On the other hand, the almost complete failure of the P3A, P6 and P7II subfragments to compete effectively cannot be explained simply in the same way. The average numbers of molecules of these factors per nucleus are not very different than those for which clear competition was obtained (P3A, ~720 molecules/nucleus; P6, ~2800 molecules/nucleus; P7II, ~1480 molecules/nucleus; Calzone *et al.* 1988). Though the K_r values of these factors differ (Calzone *et al.* 1988), they are all sufficiently high so that competition should have been observed at the molar ratios actually achieved, were these positively acting factors required for embryonic CyIIIa · CAT expression.

The CyIIIa regulatory domain is a logic switch

Interference with the interactions so far discussed depresses CyIIIa · CAT function *in vivo*, but in all cases the residual expression is spatially correct. As shown in the accompanying paper (Hough-Evans *et al.* 1990), in embryos with reduced CyIIIa · CAT function due to *in vivo* titration, CAT mRNA is still confined to aboral ectoderm cells. While these interactions are all positive in sense, the P3A and P7II subfragments, which do not competitively titrate down CyIIIa · CAT activity (Fig. 3 and Table 1), function *in vivo* as negative regulators of spatial expression. Viewed as a functional unit, the CyIIIa regulatory domain can be seen to require that a number of distinct factor-site interactions take place in order for the gene to function properly in time, in space (i.e. cell lineage) and in amplitude. The inference from what we are learning about this developmental gene regulatory system is that the multiple interactions it requires are not, as may appear superficially, redundant in a functional sense. They exist to carry out *diverse biological functions*. Perhaps each bears a specific functional relation to the biological status of the cell, that is, to developmental stage, state of specification or lineage, etc. In the case of the P8 and P5 factors, for example, which seem at first sight to function redundantly, i.e. as essentially positive regulators (this paper), we have good reason to suspect quite different biological roles. The P8 factors are present in the embryo before the CyIIIa or CyIIIa · CAT genes are activated, and their concentration remains unchanged as development proceeds (Calzone *et al.* 1988). Thus they are unlikely to be responsible for temporal activation, but to be needed as amplitude controllers. On the other hand, the same study shows that the P5 factor rises dramatically over the same period in which this and other aboral ectoderm genes are turned on (Davidson, 1989). P5 could control this important temporal developmental function, since it too is essential for expression, and this factor is probably utilized in the regulatory domains of several other genes that are activated at the same stage of development (Thiebaud *et al.* 1990). Thus we think of the CyIIIa regulatory system as a developmental logic switch, in the sense that it converts the multielement pattern of individual, functionally diverse interactions within it, into a single scalar function, *viz.* increase (or decrease) in the rate of transcriptional expression.

This research was supported by NIH Grant HD-05753. R.A. is supported by an NIH Training Grant (GM-07616).

References

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

Barberis, A., Superti-Furga, G. and Busslinger, M. (1987). Mutually exclusive interaction of the CCAAT-binding factor and of a displacement protein with overlapping sequences of a histone gene promoter. *Cell* **50**, 347–359.

Bond-Matthews, 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.

Braun, T., Tannich, E., Buschhausen-Denker, G. and Arnold, H.-H. (1989). Promoter upstream elements of the chicken cardiac myosin light-chain 2-A gene interact with *trans*-acting regulatory factors for muscle-specific transcription. *Mol. Cell Biol.* **9**, 2513–2525.

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 and Dev.* **2**, 1074–1088.

Casey, J. and Davidson, N. (1977). Rates of formation and thermal stabilities of RNA:DNA and DNA:DNA duplexes at high concentrations of formamide. *Nuc. Acid Res.* **4**, 1539–1552.

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.

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.

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. *Devl Biol.* **108**, 431–442.

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

HICKEY, R. J., BOSHAR, M. F. AND CRAIN, W. R. (1987). Transcription of three actin genes and a repeated sequence in isolated nuclei of sea urchin embryos. *Devl 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. *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* **110**, 41–50.

LAI, Z.-C., DEANGELO, D. J., DILIBERTO, M. AND CHILDS, G. (1989). An embryonic enhancer determines the temporal activation of a sea urchin late H1 gene. *Mol. Cell Biol.* **9**, 2315–2321.

LEE, J. J. (1986). The genomic organization and expression of the *Strongylocentrotus purpuratus* actin gene family. Ph.D. thesis, California Institute of Technology, Pasadena, California.

LEE, J. J., SHOTT, R. J., ROSE, S. J., THOMAS, T. L., BRITTEN, R. J. AND DAVIDSON, E. H. (1984). Sea urchin actin gene subtypes: Gene number, linkage, and evolution. *J. molec. Biol.* **172**, 149–176.

LIVANT, D., CUTTING, A., BRITTEN, R. J. AND DAVIDSON, E. H. (1988). Titration of the activity of a fusion gene in intact sea urchin 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.

MUSCAT, G. E. O., GUSTAFSON, T. A. AND KEDES, L. (1988). A common factor regulates skeletal and cardiac α-actin gene transcription in muscle. *Mol. Cell Biol.* **8**, 4120–4133.

SHOTT, R. J., LEE, J. J., BRITTEN, R. J. AND DAVIDSON, E. H. (1984). Differential expression of the actin gene family of *Strongylocentrotus purpuratus*. *Devl Biol.* **101**, 295–306.

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.

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

(*Accepted 7 June 1990*)