Intersecting batteries of differentially expressed genes in the early sea urchin embryo

Pierre Thiebaud,[1] Marcia Goodstein, Frank J. Calzone,[2] Nadine Thézé,[3] Roy J. Britten, and Eric H. Davidson

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

We determined the distribution of *cis*-regulatory sites, previously identified in the control domain of the *CyIIIa* gene, in three other genes displaying diverse spatial patterns of expression in the sea urchin embryo. Competitive gel-shift reactions were carried out using probes from the *CyIIIa* gene, with competitor fragments isolated from the previously defined control domains of the other genes. *CyIIIa* is expressed only in aboral ectoderm lineages; the other genes studied were *Spec1*, also expressed in aboral ectoderm; *CyI*, expressed in many different cell types; and *SM50*, expressed only in skeletogenic mesenchyme. All four genes are activated at about the same time in late cleavage. Where competitive interactions indicated a functionally comparable binding site (in vitro), a sequence homology was sought, and in most cases could be identified. An interesting pattern of putative regulatory site usage emerges: Of 10 *CyIIIa* interactions tested, three only were unique to the *CyIIIa* gene with respect to the set of four genes tested; one believed on previous evidence to be a temporal regulator was shared by all four genes, and the remainder were shared in various subsets of the four genes.

As initially applied by Morgan (1934), for example, in the phrase "different batteries of genes come into action as development proceeds" (p. 9), the term "battery of genes" indicates sets of genes that are activated in concert during ontogeny. In molecular terms, gene batteries can now be defined as sets of genes that share target sites for given DNA-binding regulatory factors, to which the members of the set respond. Viewed from the standpoint of individual factor–site interactions, the current literature includes many examples of gene batteries in this sense, i.e., of sets of genes that utilize homologous sites at which bind given cell type-specific factors, or cell cycle factors, or inducible factors, etc. However, *cis*-regulatory domains, as we are coming to know them, often include a large number of distinct, DNA–protein interactions. The states of the individual interactions apparently are integrated by the regulatory apparatus in specifying the transcriptional state of the gene in respect to time, space, and circumstance (cf. Davidson 1990). A gene controlled by such a regulatory domain would therefore belong to many different gene batteries, if complex *cis* regulatory systems are organized combinatorially. Thus, for example, suppose X and Y are regulators that appear in *different cell types*, recognizing *cis* sites x and y, and A and B are temporal regulators that appear at *different times* (recognizing *cis* sites a and b) in both cell types. There might be three genes expressed in the cell type of X, utilizing, respectively, the regulatory domains xa, xa, and xb, and three genes expressed in the cell type of Y, utilizing ya, yb, yb. Thus, we would distinguish four gene batteries, defined in terms of the individual regulatory interactions, which *intersect*, so that members of the cell type-specific X battery are also members of A or B batteries, while the temporally specific B battery includes members of both X and Y batteries, etc. Furthermore, in this example some members of these gene batteries share more than one element (xa, xa, or yb, yb) while others share only one element (xa, xb).

Are relationships such as these actually to be encountered as we confront real developmental regulatory systems? The early sea urchin embryo, divided by its lineage segregation into a small number of spatially discrete, polyclonal territories, each expressing a specific set of histospecific genes, offers an opportunity to approach directly these aspects of developmental regulatory architecture. Here we utilize four differentially expressed genes, each of which has been characterized by gene transfer and other studies sufficiently so as to establish at least the location of large portions of the four regulatory domains necessary for embryonic expression. These genes are listed in Table 1, together with relevant

Present addresses: [1]Laboratoire de Biologie et Génétique du Développement, Université de Rennes, 35042 Rennes, France; [2]Department of Developmental and Cell Biology, University of California, Irvine, Irvine, California 92717 USA; [3]Laboratoire de Génétique Moléculaire, Université de Rennes, 35042 Rennes, France.

Thiebaud et al.

Table 1. *Embryonic patterns of expression of four sea urchin genes*

Gene	Function	Onset of expression	Spatial territory of expression	
			initial	final[a]
CyIIIa[b]	cytoskeletal actin	late cleavage	aboral ectoderm lineages	aboral ectoderm
Spec1[c]	Ca^{2+} binding protein	late cleavage	aboral ectoderm lineages	aboral ectoderm
SM50[d]	skeletal matrix protein	late cleavage	skeletogenic mesenchyme	skeletogenic mesenchyme
Cy1[e]	cytoskeletal actin	late cleavage	all territories	oral ectoderm gut

[a]"Final" refers to pattern of expression at the late pluteus stage, at the end of embryogenesis. While the expression of *CyIIIa* is confined to embryo and larva (Shott et al. 1984; Cameron et al. 1989), two of the other three genes are expressed in adult sea urchins as well: for *CyI*, Shott et al. (1984); for *SM50*, Richardson et al. (1989).

[b]For *CyIIIa*, see review in Davidson (1986, pp. 232–233 and pp. 240–246); Cox et al. (1986); Lee et al. (1986); Lee (1986); Hickey et al. (1987).

[c]For *Spec1*, see Lynn et al. (1983); Hardin et al. (1985); Hardin and Klein (1987); Gan et al. (1990); Klein et al. (1990); Tomlinson and Klein (1990).

[d]For *SM50*, see Benson et al. (1987); Sucov et al. (1987, 1988); Killian and Wilt (1988).

[e]For *CyI*, see Davidson (1986, pp. 244–246); Cox et al. (1986); Lee et al. (1986); Lee (1986); Hickey et al. (1987); Katula et al. (1987).

literature citations. The *CyIIIa* gene, which codes for a cytoskeletal actin and is expressed only in embryo and larva (Shott et al. 1984; Lee et al. 1986; Cameron et al. 1989), is the best known from the standpoint of regulatory molecular biology, and this serves as the reference gene in the experiments we describe below. Thézé et al. (1990) mapped some 20 different sites of DNA–protein interaction in the regulatory domain of the *CyIIIa* gene. These interactions are mediated by 12 apparently distinct factors. We now possess sufficient information from gene transfer studies to assign a putative biological regulatory significance to a majority of the interactions in which these factors participate. Thus, some are required for histospecific spatial regulation (Hough-Evans et al. 1990) and some probably for temporal regulation, or positive control of level of expression under various circumstances (Calzone et al. 1988; Franks et al. 1990; for review, see Davidson 1989). The *CyIIIa* gene is expressed exclusively in the lineages constituting the aboral ectoderm. As shown in Table 1, this is also the case for the *Spec1* gene, which in all respects appears to be activated coordinately with *CyIIIa*. The other two genes utilized in the comparison, called *SM50* and *CyI*, are activated at the same developmental stage in the early embryo as *CyIIIa* and *Spec1*, but in wholly different spatial domains. *CyI* is expressed ubiquitously from late cleavage up through the mesenchyme blastula stage, the stage from which the nuclear extracts utilized for the present experiments were obtained. However, in the late embryo *CyI* transcripts are confined mainly to gut and oral ectoderm (Table 1). The *SM50* gene is expressed exclusively in skeletogenic mesenchyme, for which its activity serves as a histospecific molecular marker, from late cleavage throughout embryonic and larval development (Table 1).

To examine possible regulatory relationships among the four genes, fragments of the *CyIIIa* regulatory domain containing known sites of factor–DNA interac-

tion were labeled and used as probes in a quantitative gel-shift system (Calzone et al. 1988). The same unlabeled *CyIIIa* fragments were added in various excess as competitors for the specific reaction of the corresponding probe with factors present in the 24-hr embryo nuclear extracts (homologous competition). In parallel experiments, fragments from the regulatory domains of the *Spec1*, *SM50*, or *CyI* genes were tested for competitive activity against the *CyIIIa* probes (heterologous competition). The effectiveness of the heterologous competition, if any, provides a measure of the possibility that the factor detected by the probe is utilized by the test gene as well as by the *CyIIIa* gene. The competitions were carried out under conditions permitting only high-specificity interactions, and in all but one or two cases, everywhere that strong heterologous competitions were observed a reasonable sequence homology could be detected between the known *CyIIIa* binding site and a specific sequence element found in the competitor gene fragment. The regulatory interrelationships implied by these measurements could then be correlated with the spatial and temporal patterns of expression of the four genes during early development.

Results and discussion

Experimental approach

The object of the experiments described in the following section was to scan the regulatory domains of the *CyI*, the *SM50*, and the *Spec1* genes for sequence elements that could recognize and bind with high-specificity factors that react specifically in vitro with regulatory sites of the *CyIIIa* gene (Calzone et al. 1988; Thézé et al. 1990). To this end, we introduced subregions of the heterologous regulatory domains (or, in the case of *SM50*, the whole regulatory domain) as competitors in gel-shift reactions carried out with probes bearing the known

DNA-binding elements of the *CyIIIa* gene. The *CyIIIa* probes, and the heterologous regulatory domain fragments, are shown in Figure 1. The gel-shift competitions were carried out stoichiometrically as described earlier (see Calzone et al. 1988, and Materials and methods), so as to provide an evaluation for a parameter K_r, which is proportional to the stability of the specific DNA–protein complex formed (see Materials and methods for formal definitions). K_r thus indicates the *site specificity preference* of the protein(s) binding to the probe fragment, sometimes referred to as the "affinity" of the factor for its target sequence. Calzone et al. (1988) showed that for the various *CyIIIa* sites and factors treated in this work, K_r varied from about 3×10^4 to about 2×10^6. K_r values obtained from reactions including heterologous competitors were normalized to the K_r obtained with the homologous *CyIIIa* competitor (i.e., the unlabeled probe fragment), and the quotient is expressed in this work as the parameter F [Eq. (1) of Materials and methods]. In the presence of excess probe, values of F could also be estimated directly from the relative amounts of complex surviving when given amounts of heterologous competitor were also added [see Eq. (2) in Materials and methods]. In most cases F was obtained independently by both procedures, and as a convention in the following F is listed in roman type if derived by Eq. (1) and in italics if by the method of Eq. (2); in general the values agree excellently. When F = 1, the heterologous and *CyIIIa* sites are functionally equivalent, with respect to the stability of the DNA–protein complexes formed in vitro; where F << 1, the heterologous site binds much more weakly than does the *CyIIIa* site; where F > 1, it binds more tightly to the heterologous gene than it does to the site on the *CyIIIa* probe fragment. In this work, we have ignored all heterologous interactions in which F ≤ 0.2, not because these are necessarily unimportant biologically, but mainly because we could not usually detect sequence homologies that would explain such weak reactions. In contrast, as shown below, most cases, in which F ≥ 0.5, could be rationalized in terms of target sites in the competitor regulatory domains that display reasonable sequence homology with the known *CyIIIa* target sites. Where multiple sites for a given factor occur within a competitor fragment this may be reflected in the measured K_r value.
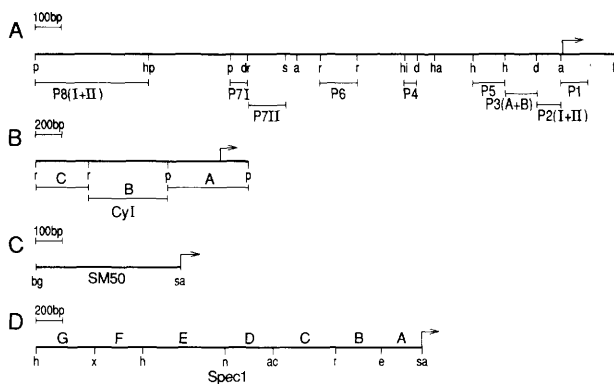

Figure 1. Map of the DNA fragments used as probes and competitors. The transcription initiation site is indicated (arrow), and the restriction sites are: (a) *Ava*II, (ac) *Acc*I, (b) *Bam*HI, (bg) *Bgl*I, (d) *Dde*I, (dr) *Dra*I, (e) *Eco*RI, (h) *Hin*dIII, (ha) *Hae*III, (hi) *Hin*fI, (hp) *Hpa*II, (n) *Nde*I, (p) *Pst*I, (r) *Rsa*I, (s) *Sau*3A, (sa) *Sal*I, and (x) *Xho*I. (*A*) *CyIIIa* 5'-flanking sequence probes (see Calzone et al. 1988; Thézé et al. 1990; for characterization of the gel shifts observed when these probes are reacted with embryo marker extracts). The probes are designated by abbreviations that refer to the regulatory factors for which binding sites are present on the probes, as mapped on the DNA sequence by Thézé et al. (1990). Probe P8 contains at least seven binding sites, viz. four sites for factor P8I, and three for factor P8II. Probes P7I and P7II contain the binding sites for these factors (two P7I sites may be present on the P7I probe); probe P6 contains two sites for factor P6, in inverted orientation; probe P4 contains a single CCAAT factor binding site, and probe P5 a single binding site. Probe P3 contains binding sites for two factors, P3A and P3B. These reactions are identified by distinct gel-shift bands. There may be two P3A sites on this probe fragment, and there is a single P3B site. P3B is an octamer binding factor. Probe P2 also contains sites for two different nonhomologous factors, interactions with which produce distinct gel-shift bands. These are P2I and P2II. Probe PI overlaps the transcription initiation site, as shown, and contains two copies of the binding site for factor PI. Canonical sequences for all these binding sites are given in Thézé et al. (1990). (*B*) Map of the *CyI* regulatory domain. The fragments used for the competition (A, B, C) reactions are indicated. (*C*) *SM50* 5' regulatory domain. The whole fragment was used for competition reactions. (*D*) Map of the 5' regulatory domain of the *Spec1* gene. The fragments used for the competition reactions (A–G) are indicated.

Distribution of functional sites among CyI, SM50, *and* Spec1 *genes for the* CyIIIa *spatial control factors P3A and P7II, for the putative temporal regulator P5, and for the CCAAT binding factor P4*

Before embarking on a general analysis of the results, it is useful to consider qualitatively the distribution among the test genes of four sites where interactions occur with factors the biological significance of which is known or suspected in the *CyIIIa* gene. Spatial control of *CyIIIa* gene expression in the embryo depends on interactions with factors called P3A and P7II (Franks et al. 1990; Hough-Evans et al. 1990; for review, see Davidson 1989). Thus, when eggs are injected with *CyIIIa* · CAT reporter gene fusions, CAT mRNA appears exclusively in aboral ectoderm lineages, as does *CyIIIa* mRNA (Hough-Evans et al. 1987, 1988, 1990). However, if interactions with either the P3A or P7II factors are prevented, by competition in vivo, a striking ectopic expression of the reporter transcript results, so that CAT mRNA accumulates in gut, oral ectoderm, and mesenchyme lineages. P3A and P7II factors are thus negatively acting spatial regulatory factors for the *CyIIIa* gene, though it should be noted that interference with the P7II interaction causes much stronger ectopic expression.

In Figure 2 are reproduced representative autoradiograms of heterologous gel-shift reactions carried out with a *CyIIIa* probe that contains the P3A and P3B sites

and with heterologous competitors from the other three genes. In Figure 2A it can be seen that the *Spec1A* fragment (cf. Fig. 1) competes strongly for the P3A factor. The K_r value calculated from these data is in fact several times that obtained earlier for the P3A reaction with the *CyIIIa* site included in the probe (F = 5.1 for this case). Figure 2B shows an example, from a directly comparable experiment, of a fragment that does not compete signifi-
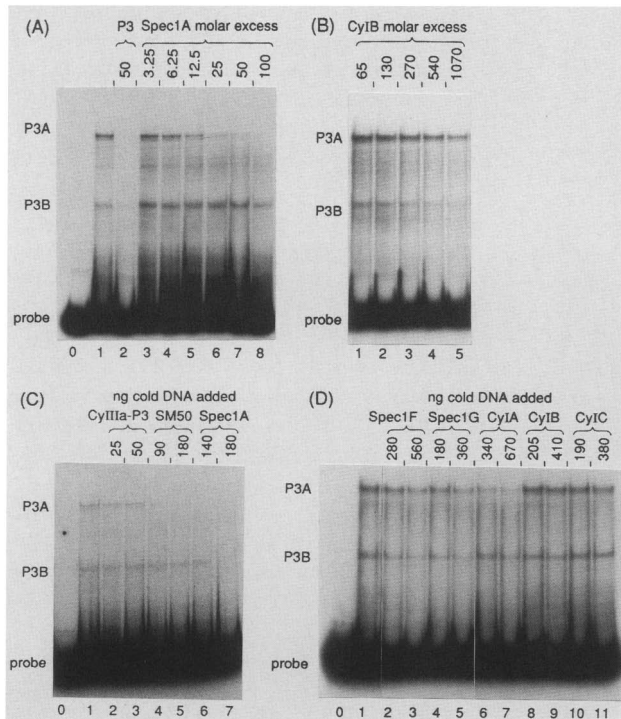
cantly for the P3A factor, in this case the *CyIB* fragment. This comparison illustrates the specificity of positive competitions such as seen in Figure 2A. Figure 2, C and D, provides examples of data from which F values were determined directly by the ratio method. Figure 2C, lanes 2 and 3, in which unlabeled *CyIIIa* P3 fragment is added, shows only small decreases in the amount of P3A complex, compared to lane 1. This is as expected, since in the ratio method all reactions are carried out near probe excess, and in this case all contain 75 ng of labeled probe. Thus, the probe specific activity in lane 2 is still 75% of that in lane 1, and in lane 3, 60%. In lanes 6 and 7, competition with the same *Spec1A* fragment as in Figure 2A is shown. The F value extracted from these experiments [see Eq. (2)] was 3.8, not significantly different from that derived by direct comparison of heterologous and homologous K_r values. Lanes 4 and 5 of Figure 2C show that the *SM50* regulatory region competes very strongly for the P3A factor(s) as well. In Figure 2D it can be seen in similar experiments that functional versions of P3A binding sites are present in *Spec1G* and to a lesser extent in *Spec1F* fragments (lanes 2–5) as well (though not in *Spec1B*, C, D, or E fragments; data not shown). A weak site is also seen in the *CyIA* fragment (lanes 6 and 7), but no P3A sites occur in *CyIB* or C fragments (lanes 8–11). F values for these and all other measurements are listed in Table 2. Figure 2 also displays the independent behavior of the octamer binding factor (P3B), for which a site also exists on the same probe fragment as used to detect P3A reactions. Functionally homologous P3B sites can be detected in several of the *Spec1* and *CyI* fragments tested as well (Fig. 2D). However, there is no evidence for other sites that compete very strongly for this factor in any of the other fragments tested, including *Spec1A* (Fig. 2B) or the *SM50* regulatory fragment, and in general no correlation exists in the distribution of competitive activity between the P3A and P3B sites.

Figure 3 shows similar observations with respect to the P7II interaction, which is also required for spatial regulation of the *CyIIIa* gene (Hough-Evans et al. 1990). Figure 3A displays a standard competition series, used for calculation of the heterologous K_r for the *Spec1E* fragment, and Figure 3B shows a ratio method experiment demonstrating that effective binding sites for the P7II factor exist in both the *Spec1B* and *Spec1E* fragments. F values were 2–3 for *Spec1B* and 3–4 for *Spec1E*, and in both cases there was again good agreement between results obtained by the ratio method and by direct K_r comparison (see Table 2). No other fragment in any of the test genes contained sites that significantly competed with the *CyIIIa* P7II site. Thus, the P7II spatial regulatory factor may serve as a specific controller of aboral ectoderm gene expression, and its gene battery would appear to include both the *CyIIIa* and *Spec1* genes. This cannot be said of the P3A factor(s), since strong sites are present in the skeletogenic *SM50* gene as well as in the *CyIIIa* and *Spec1* genes. There is also a significant, though rather weak site that binds P3A factors in the *CyIA* fragment (Fig. 2D).

Figure 2. Heterologous gel-shift competitions with a *CyIIIa* probe bearing sites for factors P3A and P3B. (*A*) Standard competition series used for K_r determinations (Eq. 1), with *Spec1* fragment A (see Fig. 1) as unlabeled competitor, and 0.22 ng of probe in all samples. (Lane *0*) Probe but no nuclear protein; (lane *1*) both but no competitor; (lane *2*) a 50-fold molar excess of unlabeled probe as competitor; (lanes *3–8*) the indicated excesses of *Spec1A* fragment. Note that the radioactive P3A complex is diluted by the *Spec1A* competitor but the P3B complex is not. (*B*) Competition series using *CyIB* fragment competitor, and 0.14 ng of the same probe as in *A*. Significant competition is not observed. (*C* and *D*) Heterologous gel shifts for estimation of *F* by ratio method (Eq. 2), carried out at relatively high probe concentration, near probe excess. The labeled probe is the same as in *A*, and each reaction contained 75 ng of unlabeled probe in C and 25 ng in D. The amounts of competitor indicated above were added in addition. Note that total *CyIIIa* probe fragment is 75 ng in lane *1*, 100 ng in lane *2* (25 ng unlabeled), and 125 ng in lane *3* (50 ng unlabeled). (*C*) Competition between the *CyIIIa* P3A and P3B sites and *Spec1A* and *SM50* sites. (Lane *0*) No protein; (lane *1*) protein and probe only; (lanes *2* and *3*) *CyIIIa* P3 fragment; (lanes *4* and *5*) *SM50* fragment; (lanes *6* and *7*) *Spec1A* fragment, in the amounts shown. (*D*) Competitions between the *CyIIIa* P3A and P3B sites and *Spec1F*, *Spec1G*, *CyIA*, *CyIB*, and *CyIC* fragments. (Lane *0*) No protein; (lane *1*) protein and probe only; (lanes *2* and *3*) *Spec1F*; (lanes *4* and *5*) *Spec1G*; (lanes *6* and *7*) *CyIA*; (lanes *8* and *9*) *CyIB*; (lanes *10* and *11*) *CyIC* (see Fig. 1).
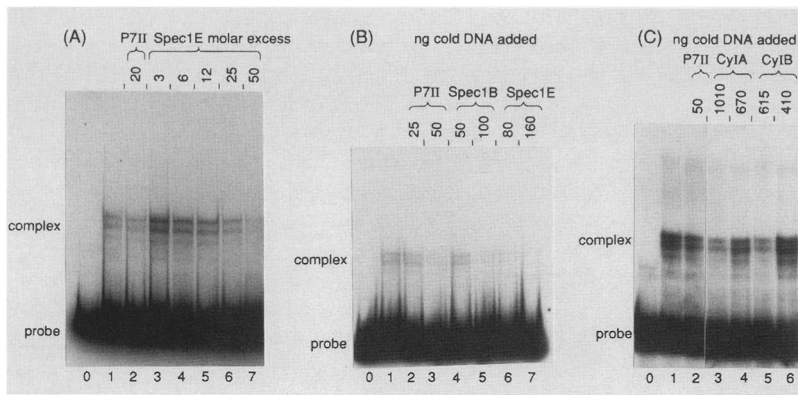
Figure 3. Heterologous gel shift competitions with a *CyIIIa* probe bearing sites for factor P7II. (*A*) Standard competition series used for K_r determination (Eq. 1) with the *Spec1*E fragment. The amount of probe in all samples was 0.47 ng. (Lane *0*) No protein extract; (lane *1*) extract but no competitor; (lane *3*) 20-fold molar excess of unlabeled probe fragment (see Fig. 1 for location of probe). (Lanes *3–7*) *Spec1*E competitions, molar ratios of competitors indicated. (*B*) Ratio method competitions, with *Spec1*E and *Spec1*B fragments. All samples (except lane *0*) contained 0.35 ng of labeled probe and 40 ng of unlabeled probe, plus the indicated quantities of the various competitor fragments. (*C*) Ratio method competitions with *CyI*A and *CyI*B fragments. All samples (except lane *0*) contained 0.30 ng of labeled probe and 32 ng of unlabeled probe, plus the indicated amount of competitor). No significant competition is observed. (Lanes *0*, *1*, and *2*) As above; (lanes *3* and *4*) *CyI*A; (lanes *5* and *6*) *CyI*B. The decrease in band intensity seen in lanes *3–6* is not significant; note the very large amounts of competitor fragments in these samples (e.g., <1000 ng and <600 ng in lanes *3* and *5*, respectively). *F* values from this experiment were 0.003 for *CyI*A and 0.008 for *CyI*B (Table 2).

Figure 4 illustrates some of the ratio method competitions carried out with the *CyIIIa* P4 and P5 binding sites. P5 is a factor that is likely to be involved in temporal regulation of the *CyIIIa* gene, based on the low concentration of this factor prior to activation of the *CyIIIa* gene in late cleavage, and the sharp accretion of this factor over the period when this gene is activated and expressed most highly. All of the genes included in the present study are activated at about the same embryonic stage, as shown in Table 1. Furthermore, P5 is known to serve as a positive regulator, interaction with which is absolutely required for expression of the *CyIIIa* · CAT fusions, according to both the in vivo competition experiments of Franks et al. (1990) and a deletion experiment of Flytzanis et al. (1987). The P4 factor is a positively acting CCAAT binding factor (Calzone et al. 1988; Thézé et al. 1990), and interaction with it is also required for function of the *CyIIIa* · CAT fusion (Franks et al. 1990). Neither P4 nor P5 effect spatial regulation (Hough-Evans et al. 1990). Nor is the P4 CCAAT-binding factor likely to be a temporal regulator, since this factor is present in the embryo at essentially the same concentration before the gene is activated as after (Calzone et al. 1988). Figure 4A displays an extremely potent P4 binding activity in the *SM50* promoter, which is about 10× greater than that of the *CyIIIa* CCAAT binding site included in the probe fragment (Table 2). A strong site can also be seen in the *CyI*B fragment, a weak site in the *CyI*A fragment, and insignificant binding activity in the *Spec1*F, G, and *CyI*C fragments (Fig. 4A). No other regions of the test genes bind this factor (Table 2). Figure 4B demonstrates that the *CyI* gene shares with the *CyIIIa* gene a P5 binding factor, a strong site for which is present in the *CyI*B fragment (lanes 1–5; F is 1.8 for this fragment; Table 2). No significant reaction was seen with the *CyI*C fragment (lanes 6–10). Sites binding the P5 factor are indeed present in all the genes tested, viz. in the *Spec1*F and G fragments, though there is no reaction with the *Spec1*A, B, C, D, or E fragment;

and in the *SM50* gene as well (data not shown here; see Table 2). We conclude that P5, a putative temporal regulator, may indeed be utilized by these four different genes that all display the same temporal pattern of developmental expression (Table 1). P5 could thus serve as a pleiotropic and general driver of temporal gene activation in the early embryo. The P4 CCAAT factor is less general in this set of genes, in that while sites for it are clearly present in the *CyIIIa*, *SM50*, and *CyI* genes, no

Table 2. *F values for heterologous gel shift competitions using* CyIIIa *gene probes and competitors from* CyI, SM50, *and* Spec1 *genes*

	Spec1A	Spec1B	Spec1C	Spec1D	Spec1E	Spec1F	Spec1G	SM50	CyIA	CyIB	CyIC
P1	*0.07*	*0.04*	*0.02*	*0.02*	*0.46*	*0.20*	*0.26*	*0.22*	*0.03*	*0.07*	*0.05*
					0.56	0.10	0.10	0.20			
P2I, II						*0.23*		*0.20*	*0.02*	*~0*	*~0*
	0.60	0.80	0.01	0.01	0.93	0.12	0.16	0.10			
P3A	*3.8*							*2.0*	*0.32*	*0.12*	*0.07*
	5.1	0.05	0.04	0.03	0.06	0.43	0.56	1.2	0.34	0.30	0.24
P3B	*~0*	0.01	0.02	0.01	0.01	0.17	0.58	0.22	0.21	*1.0*	0.33
P4						*0.23*	*0.35*	*10.0*	*0.50*	*2.6*	*0.04*
	0.08	0.06	0.05	0.05	0.03	0.35	0.29	10.7	0.21	1.2	0.06
P5	0.09	0.10	0.07	0.05	0.05	0.53	1.0	0.58	---	*1.8*	0.50
P6	No competition observed										
P7I	0.10	0.05	0.07	0.05	0.57	0.03	0.17	0.14	0.11	~0	0.20
P7II		*2.6*			*3.7*	*0.15*	*0.15*	*0.10*	*~0*	*~0*	*~0*
	0.05	2.3	0.04	0.04	3.8						
P8I, II	No competition observed										

F is the parameter describing the relative binding preference of a given factor for a site in the heterologous regulatory domain compared to the probe site in the reference *CyIIIa* gene (see text for formal definition). **F** values were obtained, as described in Materials and methods, either by the "standard competition" approach of Eq. (1), here presented in roman type, or by the "ratio method" of Eq. (2), here presented in italics. **F** values ≥0.5 are shaded, and all **F** values ≥0.2 are presented in darker type.

Figure 4. Heterologous gel-shift competitions carried out with the probes containing the P4 CCAAT binding site, and the P5 site of the *CyIIIa* gene. (*A*) P4 competitions by the ratio method. (Lane *0*) No protein, 0.15 ng probe (see Fig. 1); (lane *1*) protein and 0.15 ng probe, but no competitor; (lane *2*) 12.5 ng unlabeled *CyIIIa* P4 fragment (i.e., total probe fragment 12.65 ng); lanes *3–15* all contain 12.65 ng of P4 probe. (Lanes *3* and *4*) *Spec*1F; (lanes *5* and *6*) *Spec*1G; (lanes *7* and *8*) *SM50*; (lanes *9* and *10*) *CyI*A; (lanes *11* and *12*) *CyI*B; (lanes *13* and *14*) *CyI*C; in the amounts indicated. (*B*) Standard P5 competition series. Each reaction contained 0.03 ng of probe. Lanes *1–5* indicated excesses of *CyI*B fragment; lanes *6–10* indicated excesses of *CyI*C.

Only the major (*lower*) complex band was used in calculation of F values shown in Table 2. The experiment shows that the *CyI*C fragment does not compete significantly, while the *CyI*B fragment contains a P5 binding site.
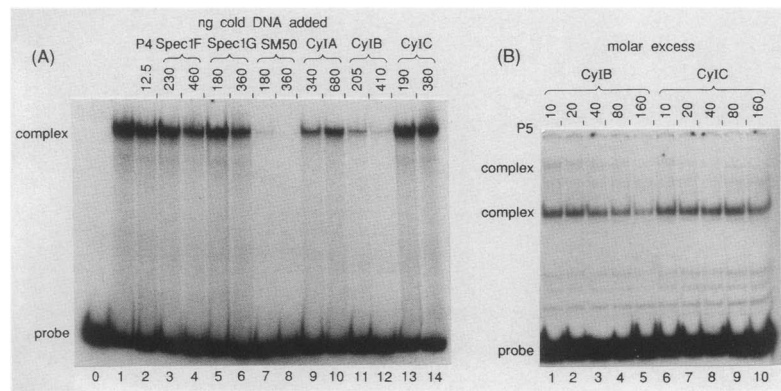
reaction was observed with any region of the *Spec1* regulatory domain tested.

Quantitative estimation of specific site preferences displayed by CyIIIa *factors for sites in* SM50, CyI, *and* Spec1 *gene regulatory domains*

In Table 2 we have collected the F values measured by both the ratio and the standard competition procedures using all of the *CyIIIa* probes mapped in Figure 1A. Table 2 includes the measurements obtained for the P3A, P3B, P7II, P4, and P5 factors as shown in Figures 2–4 (and from other experiments not shown). In addition, measurements are presented for the PI factor; for the P2I and P2II factors, which bind adjacently in the *CyIIIa* gene; and the P7I factor. In Table 2, F values $\geqslant 0.5$ are shaded, and those <0.2 are given in light type. As discussed by Calzone et al. (1988) the actual prevalence of many of the *CyIIIa* factors in 24-hr embryo nuclear extracts has been found to be approximately of the magnitude required for binding to their target sites, given the K_r values measured for these factors in vitro. Furthermore, Livant et al. (1988) showed by a direct in vivo titration that the prevalence of at least the limiting factors for *CyIIIa* · CAT expression in living embryo nuclei is about what is deduced from measurements made on nuclear extracts. Therefore, an argument can be made that in general an F value for a heterologous site less than one-fifth or one-tenth of that measured for the *CyIIIa* sites would probably preclude extensive binding to that site, i.e., at the factor concentrations that actually exist. However, it is most important to bear in mind that there is no direct evidence, for any of the heterologous genes, either that interactions displaying high F values in Table 2 are actually functional, or that interactions displaying low F values are actually nonfunctional.

Table 2 shows that with respect to our small set of heterologous test genes, the two P8 factors [I and II; Thézé et al. (1990) reported a third factor, III, but this has been discovered to bind to a site artifactually created in cloning], and the P6 factor, react uniquely with the

CyIIIa gene. No competitively active sites for these factors were detected, and the experiment does not determine whether these sites are shared by the other genes. However, all of the other factors display binding sites for which $F \geqslant 0.5$ in one or more of the test genes. Considering the small size of the gene sample we have chosen for study, and its diversity in terms of expression, this is in itself remarkable. The implication is that all of these "shared" factors are pleiotropic regulators, in that they affect multiple genes. Thus, there exists a battery of genes defined by each. Furthermore, there might be a limited, relatively small number of regulatory factors operative, that is, relative to the total number of genes operative at this stage (estimated at ~8500 for slightly later sea urchin embryos; for review, see Davidson 1986, pp. 165–170). Yet, with the possible exception of P5 (see below), there is no ubiquitous binding site for any of the factors considered among this small set of test genes. Instead, most of the factors whose distribution is assessed in these experiments appear in *different combinations* in each of the four regulatory domains.

Homologous sequence elements in regions of the heterologous regulatory domains that compete for CyIIIa *DNA-binding factors*

Functions are known or implied for a majority of the factors that interact with the *CyIIIa* gene (for review, see Davidson 1989), and these interactions were initially identified on the basis of their high K_r values. Thus, where in this work $F \geqslant 0.5$, indicating a similar high K_r value in the heterologous gene, a priori expectation would predict a functional, sequence-specific binding site, homologous with that in the respective *CyIIIa* gene probe. The results of sequence homology searches carried out in the heterologous DNA fragments that we scanned are summarized in Table 3. Here we have made use of preexisting knowledge of the *CyIIIa* binding sites (and where relevant, other known binding sites for given factors), as identified by oligonucleotide mapping, foot-

Table 3. *Sequence homologies with CyIIIa target sites in SM50, CyI, and Spec1 regulatory domains*[a]

| P1:[b] | (−12) | TGGTCCCCCACAGT + |
| | (+35) | TGGTGTCATCCAGT + |

Possible Strong Sites:

| Spec1E | (−2293) | GGGTGTTgCACAGT − |
| | (−2306) | TGaTGCCCTtCAaa − |

Possible Weak Sites:

| Spec1F | (−2865) | TcaTCCCATAatGT − |
| | (−2792) | TctTGCTATACAaT + |

Apparently Inactive Sites:

| Spec1B | (−607) | aGtTCCCCTCtAGT + |
| | (−489) | cGGTGCCACgCAtg − |

Important nucleotides: $\frac{T}{G}$GGTGC$\frac{T}{C}\frac{T}{C}$X$\frac{T}{C}$XCAGT

| P2I:[c] | (−71) | AACCCTGGACATAACTCTCGCTTG + |

Possible Strong Sites:

Spec1A	(−161)	tAatgattACATAACTtTCGtgaG +
Spec1B	(−475)	ttCaCTGaACATGACctaattTTt −
Spec1E	(−2245)	AAaatTaGACATAAGaCagGCTTt −

Apparently Inactive Sites:

| Spec1F | (−3085) | tgtCtTGGgaATAACTtTtatcTt + |

Important nucleotides: ACAT$\frac{A}{G}$A$\frac{C}{G}$

| P2II:[c] | (−34) | AAGGAAGGGGTAGTACTTTACTTGG + |

Possible Strong Sites:

Spec1A	(−315)	tcGctccaacTcaTgaTTTCCTTGG −
	(−404)	gAtaAccaaaataaAgTTTCATTGG −
Spec1B	(−670)	tgGaAAGGGTTAcattTcTttTTtG −
Spec1E	(−2385)	gttagAGGGTTAGTACagTAAaaGG −

Apparently Inactive Sites:

| Spec1F | (−2983) | cAcaAtGGGGcAGTgCgaTtAaaGt − |

Important nucleotides: AGGG$\frac{G}{T}$TA or TTT$\frac{AC}{CA}$TTGG

P3A:[d]	(−818)	gcgGCGCA +
	(−190)	TtTGCGCT −
	(−113)	gagGCGCT +

Possible Strong Sites:

Spec1A	(−82)	TcTGCGCA +
	(−80)	CaTGCGCA −
	(−62)	TcCGCGCA +
SM50	(−354)	CcCGCGCA +
	(−248)	TaTGCGCT +
	(−128)	TcTGCGCA +
	(−114)	CaCGCGCA +
	(−34)	CtCGCGCA −

Possible Weak Sites:

Spec1F	(−3006)	TtTaCGCT −
	(−2937)	TtTGtGCT −
CyIA	(−108)	TcgGCGCA −

Apparently Inactive Sites:		
Spec1E	(−2513)	TgTGgGCT +
Spec1D	(−1730)	TgTtCGCT +
	(−1579)	TtCtCGCT −
	(−1571)	TaTGaGCT −

Important nucleotides: $\frac{C}{T}$X$\frac{C}{T}$GCGC$\frac{A}{T}$

| P3B:[e] | (−154) | CTCATTTGCATATC + |

Possible Strong Sites:

| CyIB | (−825) | aatATTTGCATAaa − |

Possible Weak Sites:

| Spec1G | (−3328) | aTaAcTTGCATtTC + |

Apparently Inactive Sites:

| Spec1D | (−1863) | tgtATTTGCATtca + |

Strong Site Consensus: ATTTGCATA

| P4:[f] | (−654) | CTCTGATTGGACCA + |

Possible Strong Sites:

| SM50 | (−163) | CTCTGATTGGACCA + |
| CyIB | (−800) | gTaTGATTGGTCCc + |

Apparently Inactive Sites:

Spec1A	(−408)	gTtTcATTGGTCgt −
Spec1B	(−825)	taCgaATTGGTCtA −
Spec1C	(−1095)	CTCTGATTGaTtCA +
Spec1E	(−2617)	acCcaATTGGACCc −

Strong Site Consensus: CTCTGATTGG$\frac{T}{A}$CCA

| P5:[g] | (−311) | TTCATTGTCGCGACATACTTGTAG + |

Possible Strong Sites:

Spec1G	(−3285)	TTCATataattataAaAaTTagtt −
Spec1F	(−2823)	TTCATTGaaGgattAaAtcctatt −
SM50	(−3)	TTCATTcgaGtttggTggTaGTcG +
	(+91)	TTCATgGctGCttCtcAaaTGTgt −
CyIB	(−771)	TaCATTGTgttGgtATtaaTaaAc +

Apparently Inactive Sites:

Spec1E	(−2662)	TTCATTcTCtCctaAggaggGTtt +
	(−2180)	TTCATTGctttGtgAcgaTTtaAa +
Spec1C	(−1006)	TcgATgtTgatttaATACTTGTAt +

Important nucleotides: TTCATTG?

| P7I:[h] | (−1375) | GGTTGTGGTCATGGTT + |

Possible Strong Sites:

| Spec1E | (−2031) | aGTTGTaCTTATaGca + |
| | (−1989) | caTTGTtGGTATgaga − |

Apparently Inactive Sites:

| Spec1B | (−798) | atTTGgGGTCATatga − |
| CyIA | (−198) | acTTGTGCGaATcaga + |

Important nucleotides: TTGTX$\frac{GTT}{CGC}$AT

Table 3. (*Continued*)

P7II:[i]	(−1234)	ACCGTTACCCTTAACCT	+
Possible Strong Sites:			
*Spec1*E	(−2376)	gtaCTAACCCTCtAaCa	+
*Spec1*B	(−660)	AatGTAACCCTTtcCag	+
Apparently Inactive Sites:			
*Spec1*C	(−1151)	ACatTcACCCTTttCaT	+
Strong Site Consensus:		$\frac{G}{C}\frac{A}{T}\frac{A}{T}ACCCT\frac{T}{C}$	

tion that would include the strongly interactive fragments and exclude those that did not react. Here we have included the best matches, in both the strong and inactive categories.

[h]The two sequences we show for *Spec1*E are the best homologies we found in this region. A consensus sequence derived from these and the *CyIIIa* target sequence turned out to provide a sequence that occurs uniquely in *Spec1*E. We show a few similar, but imperfect, sequences in inactive regions of *Spec1* and *CyIA* genes.

[i]The consensus shown for the **P7II** site occurs only in *Spec1*E and B, the two fragments that compete strongly with the *CyIIIa* probe. A similar sequence in the inactive region of *Spec1*C is missing a crucial nucleotide.

[a]A "possible strong site" is one that occurs in a fragment for which the competition experiments yielded a value of $F \geq 0.5$ (Table 2). A "possible weak site" is one that occurs in a fragment for which F is 0.2–0.5. "Inactive sites" show no significant competition, i.e., $F < 0.2$. In each panel, the top sequence is that derived for the *CyIIIa* gene by Thézé et al. (1990). By convention, we label the orientation of this sequence as +; these sequences are given 5' to the left. The orientation of other sequences shown is the same if also labeled +, but opposite if labelled −. Sequence positions are given relative to the transcription start site. Boldface indicates nucleotides that agree with the previously established *CyIIIa* target site (except for **P3A**; see note). Lightface capital letters indicate nucleotides that disagree with the *CyIIIa* site but agree with the derived "strong site consensus" or "important nucleotides." Lightface lowercase letters indicate nucleotides that agree with neither. Shading indicates regions of the listed sequence that appear important.

[b]The two sequences we show for the **P1** strong sites are the best homologies we found in *Spec1*E, where **P1** reacts. There are two occurrences in every region where **P1** competes. The important nucleotides are based on the first *Spec1*F site, which appears closer to the *CyIIIa* site.

[c]Independent searches for the best homologies in **P2I** and **P2II** against *Spec1* revealed that the *Spec1* homologies were often relatively close to one another; note indicated positions. **P2I** and **P2II** reactions can be distinguished by different footprints (Thézé et al. 1990).

[d]Thézé et al. (1990) found that all **P3A** sites share the **GCGC (A/T)** core sequence. The importance of this is seen in the weak sites, which differ by only one nucleotide in the core sequence. We derived the consensus sequence from the *SM50* and *Spec1* sequences, which bind the **P3A** factor more strongly than do the *CyIIIa* target sequences. In this instance, bold letters reflect agreement with the derived consensus sequence.

[e]Thézé et al. (1990) observed that **P3B** is an octamer-binding protein. However, because of the octamer sequence in *Spec1*D, it is clear that sequence surrounding the core site is also important. The sequences shown in *CyIB* and *Spec1*G are the only other occurrences of octamer-like sequences that would explain the observed competition.

[f]Thézé et al. (1990) determined **P4** to be a **CCAAT** binding factor. The inactive sites we show are specifically chosen because they include the core sequence **CCAAT**, showing that external sequence is necessary. We derived the consensus shown using the *CyIB* sequence, which is the best homology to the *CyIIIa* **P4** site that we found. The inactive sequence in *Spec1*C emphasizes the importance of the core sequence as well as the external nucleotides.

[g]We analyzed **P5** extensively and could find no unique explana-

printing, and various other procedures (Calzone et al. 1988; Thézé et al. 1990). In each panel of Table 3, the actual or core *CyIIIa* site is given, as derived from these sources; then listed is the most likely homologous site, if any, in fragments that compete with a high F value (i.e., ≥ 0.5); followed by weak sites, if any, in other fragments; and by the best examples discovered of sequences that may appear homologous, but which derive from fragments displaying no actual competitive competence ($F = \leq 0.2$). Where a unique, though often partially degenerate sequence element can be inferred that, upon search of all the heterologous DNA fragments tested, turns out to be distributed only in those fragments that display significant competition, this sequence is listed at the bottom of the panel ("important nucleotides"). Table 3 shows that such a solution can in fact be derived for all of the factors tested for which there were found competitive interactions, except for P5.

A brief précis of the results in Table 3 follows.

P1. There are two P1 binding sites in the *CyIIIa* gene about 35 nucleotides apart that straddle the transcription start site. A striking observation shown in Table 3 is that two nearby homologous sites can also be observed in the *Spec1*E fragment which competes well for the P1 factor ($F \sim 0.5$). Thus, the double nature of this site is confirmed; however, the importance of its location is not, since in the *Spec1* domain these sites are distant from the transcriptional origin. The "important nucleotide" consensus shown for this site differs slightly in the internal region from that derived by Thézé et al. (1990) from the *CyIIIa* sites alone, but agrees in the significance of the relatively invariant terminal regions of the site [i.e., the $^T/_G$GGT(N)$_6$CAGT elements].

P2. The fragment used for P2 competitions in this work included two sites that react with different proteins, called P2I and P2II (Thézé et al. 1990). Table 3 suggests that the P2I and P2II sites are not linked adventitiously in the *CyIIIa* gene. Thus, in all three locations in the *Spec1* gene where significant P2 competitions occur, elements at least weakly homologous to *both* P2I and P2II sites are found. Thus, significant sequence matches are found for the P2II site (8 of 9 or 8 of 11 in *Spec1*A; 8 of 10 in *Spec1*B; 10 of 11 in *Spec1*E) within

100–200 nucleotides of P2I homologies. Thézé et al. (1990) had no comparisons with which to derive consensus sequences for the P2I and P2II elements, and at least for P2I the strong sites shown in Table 3 improve our understanding of the important regions of this target site.

P3A. Table 2 indicates very strong competitive reactions for this factor in the *Spec1*A and *SM50* fragments (F > 1; see Fig. 2). The explanation for these is obvious in Table 3, where it can be seen that the *Spec1*A fragment has two perfect sites conforming to the canonical target sequences for this protein ($^C/_TN^C/_TGCGC^A/_T$) (Thézé et al. 1990). Similarly the *SM50* fragment has five such perfect sites. Other sequences, in which this core is violated by even one nucleotide, produce relatively weak or no competitive reactions (cf. Tables 2 and 3).

P3B. Table 2 shows that a strong interaction with the "octamer" factor that binds to the *CyIIIa* P3B site was obtained with the *CyI*B fragment (F = 1.0). This fragment shows a perfect 9-nucleotide homology with the *CyIIIa* site. Homology outside the octamer site itself is important, since a perfect octamer sequence also appears in the *Spec1*D fragment, but without peripheral homology, and this site is competitively inactive (F = 0.01. On the other hand, the *Spec1*G region displays an imperfect octamer but with some surrounding homology, and a weak interaction is here observed.

P4. Table 2 shows that the *SM50* regulatory fragment reacts very strongly with the CCAAT factor that binds to the P4 site of the *CyIIIa* gene (F = 10), and a fairly strong reaction (F ~1.9) is also obtained with the *CyI*B fragment (Fig. 4A). No other obviously significant reactions were obtained. In Table 3 it can be seen that sites in the *SM50* and the *CyI*B fragments match the *CyIIIa* site at 14/14 and 11/14 positions, respectively, surrounding the ATTGG element. This 5-nucleotide core sequence is per se insufficient, as shown by the inactive *Spec1*A, B, and E sites in Table 3. However, a quantitative mystery remains. Since the *SM50* and *CyIIIa* sites are identical, it is impossible to account simply for the 10-fold difference in complex stability. Thus, either additional flanking sequence is important, or very strong cooperative interactions with adjacent factors greatly stabilize the *SM50* P4 factor complex (this fragment contains many very apposed, diverse binding sites).

P5. According to Table 2 and Figure 4B, significant binding sites for the P5 factor are present in the *Spec1* gene F and G fragments, the *SM50* regulatory sequence, and both the A and B fragments of the *CyI* regulatory region. As noted earlier, this observation is consistent with independent evidence suggesting that P5 may be a positive temporal regulatory factor (Calzone et al. 1988), since all of these genes, and *CyIIIa* as well, are activated at about the same time, in late cleavage. Thus, it is disappointing that we could not derive a common P5 binding site by using the footprinted sequence element of the *CyIIIa* gene (see Table 3; Thézé et al. 1990) which is unique to the active competitors. No consistent pattern emerges that would explain the *inactivity* of the *Spec1*C and E sites shown in Table 3.

P6. No significant intergenic competition was obtained.

P7I. This factor may also function as a temporal regulator (Calzone et al. 1988), and Table 2 shows the only significant competitive reaction in these experiments was with the E fragment of the *Spec1* gene. Table 3 displays a possible basis for this interaction, since *Spec1*E includes a homologous element that may define the core P7I binding site sequence, and that is not found in the other regions surveyed.

P7II. Table 2 shows strong competitive interactions with this spatial regulatory factor (F > 1) for two regions of the *Spec1* regulatory domain, and for none other of the tested fragments (see Fig. 3). A simple core sequence element can be derived (Table 3), which occurs in the appropriate *Spec1* fragments, but not elsewhere.

P8. No functional homology appeared to exist among the test genes for the two P8 factors required by the *CyIIIa* gene (Table 2).

Implications

We return to an issue with which we began. We have defined a gene battery as a set of genes that share a functional site for a given transcriptional regulator (Britten and Davidson 1969, 1971). Do the batteries subject to each given regulatory factor intersect, so that the *same* interactions appear in genes that display *different* spatial and temporal patterns of expression? It is useful here to consider two distinct meanings of the concept of regulatory *combinatoriality*. In literal terms, or in a strict biochemical sense, any *given cis*-regulatory domain that consists of multiple sites of interaction and that operates by means of biochemical interactions between the diverse bound factors present in a given domain can be said to function by a mechanism that is combinatorial, particularly if the interactions are cooperative. However, it is with the sense of this term that applies to regulatory *systems* that we are here more concerned. Regulatory *combinatoriality* requires—and follows from—an additional feature, viz., the specific form of organization referred to here as "intersecting gene batteries." Thus, in a combinatorial regulatory *system* the output of a given *cis*-regulatory domain would indeed depend on all or many members of a set of diverse interactions, but in addition the gene battery defined by any one of these intersections will also include other genes that are controlled by complex regulatory domains of different constitution, and producing different overall patterns of expression. The functional value of any particular combination will also of course be affected by the affinity of the factor for the individual site, with respect to the factor concentration at any given stage of life.

In Figure 5, the results of this study are presented as an abstract battery diagram (see legend for identification of symbols). Each box represents a gene battery in the primary sense that there occurs within the regulatory domains of the genes included a sequence-specific interaction with a given factor or factors. Where the box is solid, we infer this both from the competition results

(Table 2) and from sequence homology data (Table 3); the dashed box indicates that in the case of P5 we could not detect a consistent sequence homology among the strongly competing fragments. One conclusion that emerges immediately from Figure 5 is that the *Spec1* regulatory system is much more similar to the *CyIIIa* regulatory system than are the *SM50* or *CyI* regulatory systems. The regulatory system of this gene is under intensive investigation (Gan et al. 1990; Tomlinson and Klein 1990), but to our knowledge the possible functional roles of upstream sequences are not yet determined. One of the regulatory factors for which there are sites in both *CyIIIa* and *Spec1*, P7II, is known to function for the *CyIIIa* gene as a powerful negative regulator of aboral ectoderm specific spatial expression (Hough-Evans et al. 1990). The second known spatial regulator of *CyIIIa*, the P3A factor, is clearly utilized in other kinds of gene batteries as well (F. Calzone and E. Davidson, unpubl.; Table 2; Table 3; Fig. 5). The functions of the PI, P7I, P2I, and P2II factors are yet unknown, except that the P2 and the P7I factors function positively (Franks et al. 1990). Since all these factors interact with the two aboral ectoderm-specific gene regulatory domains, but not with *SM50* or *CyI* regulatory domains, it will be interesting to determine whether the spatial distribution of active forms of these factors is confined to

aboral ectoderm lineages. At present we know only that interference in vivo with P2 (I and II) or P7I interactions does not induce ectopic spatial expression (Hough-Evans et al. 1990). These (and P1) might be positively acting factors that modulate gene expression within the aboral ectoderm lineages in ways we have not tested, for example, in response to growth rate, inductive influences, or later in larval development, etc.

A second result of the arrangement shown in Figure 5 is that one can see clearly several examples of the intersection of different gene batteries. Thus, the putative P5 battery, consisting of a set of genes activated at the late cleavage–early blastula stage, includes genes that are expressed in three diverse spatial patterns (see Table 1). Neither the P4 CCAAT binding factor battery nor the P3B octamer factor battery is ubiquitous (Table 2; Fig. 5), and thus each may serve some specific biological function. The P4 battery intersects both the aboral ectoderm specific battery, and a likely skeletogenic mesenchyme gene battery (only the one member, *SM50*, is here included), while the P3B battery includes the *CyI* actin gene as well as the aboral ectoderm-specific gene battery.

We have in this study looked at only four genes, and all of the gene batteries we observe are almost certainly incomplete, as might also be some of the regulatory domains examined. Nonetheless, Figure 5 defines five different gene batteries, four of which *intersect* the aboral ectoderm gene battery. It must follow that the canonical patterns of spatial and temporal gene expression in the early sea urchin embryo are indeed controlled in a combinatorial manner, sensu strictu. This is a hopeful as well as fascinating inference, in that it suggests that a significant fraction of the early embryonic regulatory system might be encountered on examination of a modest number of genes that are expressed differentially at the onset of development.

	PI	P2(I, II)	P7I	P7II	P3B	P3A	P5	P4	P6	P8(I, II)
CyIIIa	+(2)	+	+	+	+	+	+	+	+(2)	+(8)
Spec1	+	+	+	+	+	+	+	o	o	o
CyI	o	o	o	o	+	o	+	+	o	o
SM50	o	o	o	o	o	+	+	+	o	o

Figure 5. Intersecting gene batteries detected by competition with the factors (or sets of factors) identified in the *CyIIIa* regulatory domain are shown across the top of the gene, in an arbitrary order chosen to facilitate the argument of the figure. Site order is in fact probably not functionally important, although immediate contiguity of specific sites may in some cases be. *Rows* indicate the strong interaction sites (F > 0.5) in each regulatory domain. The *CyIIIa* domain gene used as the reference gene in this study includes all of the necessary regulatory sites for correct specification in early development (for review, see Davidson 1989) and only those sites in the other regulatory domains are shown that are shared with the *CyIIIa* domain. Numerals indicate the number of times (>1) a given site appears in the regulatory domain estimated from Table 3. Each of the other domains may be expected to contain additional sites that could be either unique or members of additional gene batteries. (O)No strong interaction was observed (F < 0.2; see text); (+) a strong interaction was observed in the other regulatory domains. Boxes indicate gene batteries and shared sites, where *batteries* are defined as sets of genes sharing given functional binding sites. These also are incomplete, since there will be other genes that utilize the same sites, again indicated by the dotted lines at the bottom of each box. In each case inclusion in the battery is supported by *both* strong competitive binding in vitro (Table 2) and sequence homology (Table 3); the exception is P5 for which, in the absence of a simple, consistent sequence homology, the basis for the competitive binding reaction is unclear (*dashed line*).

Materials and methods

Probes and competitors

Probes were obtained from subregions of the *CyIIIa* regulatory domain (Calzone et al. 1988; Thézé et al. 1990), and were isolated by restriction enzyme digestion. A map indicating the respective probes is shown in Figure 1A. Probe fragments were labeled by the end-filling reaction using Klenow polymerase, with all four ^{32}P deoxynucleotide triphosphates (3000 Ci/mmole, Dupont NEN). The respective probes were then gel-purified. Figure 1, B (*CyI*), C (*SM50*), and D (*Spec1*), shows to scale the subfragments of the regulatory domains of the other three genes that were used as competitors in the gel-shift reactions. These were also gel-purified, eluted, and the quantity estimated by spectrophotometry or the DAPI fluorescence method of Brunk et al. (1979).

Gel shift assays

The 24-hr *Strongylocentrotus purpuratus* embryo nuclear extracts were prepared as described previously by Calzone et al. (1988). The gel-shift assays were also performed as indicated by Calzone et al. (1988), using poly(dI)/poly(dC) (Pharmacia) as nonspecific DNA for the P5 probe, and poly(dAT) (Pharmacia)

for all other probes. Each binding reaction (10 μl) contained 20 mM HEPES (pH 7.9), 0.5 mM DTT, 75 mM KCl, 5 mM MgCl$_2$, 5% glycerol, 1–4 μg of nuclear extract, and 0.1–0.5 ng of probe (5–10,000 cpm). After 10 min of incubation at 15°C, 2 μl of Gel Sample Buffer (15% Ficoll 400, 0.25% bromophenol blue, 0.25% xylene cyanol) was added to the reaction and the complexes were separated from unbound probe by electrophoresis in 5% acrylamide gels (30% acryl/0.8% bis) in 1× TBE [(1× TBE is 50 mM Tris borate (pH 8.3), 1 mM EDTA] at room temperature for 2–4 hr at 200 V, depending on the size of the probe. After electrophoresis, the gels were dried and the radioactive complexes and unbound probe located by autoradiography. The probe and the complexes were excised from the dried gel and counted in a scintillation counter.

Data reduction procedures and quantitative treatment

Our object was to determine the relative "affinity" of the heterologous competitor DNA for a given DNA-binding factor (or factors) for which known sites are present on the *CyIIIa* probe fragment, with respect to the affinity of this factor (or factors) for the *CyIIIa* site(s). K_r gives the preference of a factor for a given specific site compared to nonspecific DNA sites, where $K_r = K_s/K_n$ (Emerson et al. 1985; Calzone et al. 1988). Here the parameters K_s and K_n are the equilibrium constants for specific and nonspecific reactions, respectively. Thus, for this study the key parameter, F, is the ratio of K_r for reaction of the same factor(s) with the heterologous competitor, to K_r for the reaction of the factor(s) with the *CyIIIa* probe sites, or where H signifies the heterologous competitor gene,

$$F = K_{r(H)}/K_{r(CyIIIa)} \qquad (1)$$

In practice F was obtained (as noted in text) in two different ways:

i. $K_{r(H)}$ and $K_{r(CyIIIa)}$ were measured in separate reactions that included together with the labeled *CyIIIa* probe either the unlabeled heterologous competitor gene fragment, or the same *CyIIIa* fragment used as probe, except unlabeled. The K_r values were extracted from the competition gel-shift data as described in detail by Calzone et al. (1988): Briefly, where PD_s is the concentration of DNA–protein complex formed, D_s and D_n are the concentrations of unbound specific and nonspecific DNA (i.e., for nonspecific DNA, synthetic DNA, as above), and P_0 is the concentration of factor in the reaction,

$$\frac{PD_s \cdot D_n}{D_s} = K_r(P_0 - PD_s)$$

(Eq. 1 of Calzone et al. 1988; see Emerson et al. 1985); and rearranging,

$$PD_s = \frac{P_0 K_r D_s}{D_n + K_r D_s}$$

(Eq. 2 of Calzone et al. 1988). Using these forms, the K_r values were extracted by least squares from measurements of the amounts of complex formed as a function of quantities of specific DNA (probe + competitor, D_s) in the samples. Where the competitor and *CyIIIa* probe K_r values differ significantly, this ratio is a valid measure of the relative affinities only as PD_n becomes relatively insignificant, i.e., in practice beyond the initial few percent of the reactions.

ii. For the case where PD_s approaches P_0, i.e., sufficient labeled and unlabeled probe sequence $(D_s^* + D_s)$ is present so as to titrate out all or most of the factor present, F can be obtained more directly:

$$F = C(R - 1) \qquad (2)$$

Here $C = D_s^*/H_s$ where D_s^* is again the input molar concentration of labeled *CyIIIa* probe and H_s is the input concentration of *unlabeled heterologous* competitor sequence. R is measured as the ratio of radioactivity in probe–factor complex when there is no competitor added to when the heterologous competitor is present (in the ratio specified by C); thus where $PD_s^*_{(0)}$ is probe–factor complex *without competitor* (in or near probe excess), and $PD_s^*_{(H)}$ is probe–factor complex *in the presence of competitor*, $R = PD_s^*_{(0)}/PD_s^*_{(H)}$. This method requires only a few measurements of R at different C values.

Equation (2) follows from the definition of competition stoichiometry. Thus in the above symbolism for the *homologous* reaction,

$$\frac{PD_s^*}{PD_s^*_{(0)}} = \frac{1}{1 + D_s/D_s^*} \qquad (3)$$

and for the heterologous reaction,

$$\frac{PD_s^*_{(H)}}{PD_s^*_{(0)}} = \frac{1}{1 + PH_s/PD_s^*_{(H)}} \qquad (4)$$

From Eq. (4)

$$PH_s = PD_s^*_{(0)} - PD_s^*_{(H)} \qquad (5)$$

This is in any case required by the condition of saturation by probe excess. From Eq. (1) above,

$$F = \frac{PH_s}{H_s} \cdot \frac{D_s}{PD_s}$$

and since the ratio of the concentrations of the labeled and unlabeled moieties are equal $(D_s^*/PD_s^* = D_s/PD_s)$, we have by substitution of Eq. (5), Eq. (2).

Acknowledgments

References

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

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

———. 1971. Repetitive and non-repetitive DNA sequences and a speculation on the origins of evolutionary novelty. *Q. Rev. Biol.* **46:** 111–138.

Brunk, C.F., K.C. Jones, and T.W. James. 1979. Assay for nanogram quantities of DNA in cellular homogenates. *Anal. Biochem.* **92:** 497–500.

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

Cameron, R.A., R.J. Britten, and E.H. Davidson. 1989. Expression of two actin genes during larval development in the sea urchin *Strongylocentrotus purpuratus*. *Mol. Reprod. Dev.* **1:** 149–155.

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

Davidson, E.H. 1986. *Gene activity in early development*, 3rd ed. Academic Press, Orlando, Florida.

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

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

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

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

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

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

Hardin, S.H., C.D. Carpenter, P.E. Hardin, A.M. Bruskin, and W.H. Klein. 1985. Structure of the *Spec1* gene encoding a major calcium-binding protein in the embryonic ectoderm of the sea urchin, *Strongylocentrotus purpuratus*. *J. Mol. Biol.* **186:** 243–255.

Hardin, P.E. and W.H. Klein. 1987. Unusual sequence conservation in the 5' and 3' untranslated regions of the sea urchin *Spec* mRNAs. *J. Mol. Evol.* **25:** 126–133.

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

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

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

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

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

Klein, W.H., C.R. Tomlinson, W. Zhang, M. Xiang, M. Kozlowski, L. Gan, T. Nguyen, and G.M. Wessel. 1990. Ectoderm differentiation and the Spec gene family of sea urchins. *UCLA Symp. Mol. Cell. Biol.* **125:** 123–134.

Killian, C.E. and F.H. Wilt. 1988. The accumulation and translation of a spicule matrix protein mRNA during sea urchin embryo development. *Dev. Biol.* **133:** 148–156.

Lee, J.J. 1986. "Activation of sea urchin embryogenesis: Nuclear synthesis and decay rate measurements of transcripts from five different genes." Ph.D. Thesis, California institute of Technology.

Lee, J.J., F.J. Calzone, R.J. Britten, R.C. Angerer, and E.H. Davidson. 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., A. Cutting, R.J. Britten, and E.H. Davidson. 1988. An in vivo titration of regulatory factors required for expression of a fusion gene in transgenic sea urchin embryos. *Proc. Natl. Acad. Sci.* **85:** 7607–7611.

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

Morgan, T.H. 1934. *Embryology and genetics*. Columbia University Press, New York.

Richardson, W., T. Kitajima, F. Wilt, and S. Benson. 1989. Expression of embryonic spicule matrix gene in calcified tissues of adult sea urchins. *Dev. Biol.* **132:** 266–269.

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

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

Sucov, H.M., B.R. Hough-Evans, R.R. Franks, R.J. Britten, and E.H. Davidson. 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., F.J. Calzone, P. Thiebaud, R.L. Hill, R.J. Britten, and E.H. Davidson. 1990. Sequences of the CyIIIa actin gene regulatory domain bound specifically by sea urchin embryo nuclear proteins. *Mol. Reprod. Dev.* **25:** 110–122.

Tomlinson, C.R. and W.H. Klein. 1990. Temporal and spatial transcriptional regulation of the aboral ectoderm-specific Spec genes during sea urchin embryogenesis. *Mol. Reprod. Dev.* **25:** 328–338.

Intersecting batteries of differentially expressed genes in the early sea urchin embryo.

P Thiebaud, M Goodstein, F J Calzone, et al.

References	This article cites 29 articles, 9 of which can be accessed free at: **http://genesdev.cshlp.org/content/4/11/1999.full.html#ref-list-1**
Email Alerting Service	Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article or **click here.**