Proc. Natl. Acad. Sci. USA
Vol. 93, pp. 9322–9328, September 1996
Colloquium Paper

This paper was presented at a colloquium entitled "Biology of Developmental Transcription Control," organized by Eric H. Davidson, Roy J. Britten, and Gary Felsenfeld, held October 26–28, 1995, at the National Academy of Sciences in Irvine, CA.

Modular cis-regulatory organization of developmentally expressed genes: Two genes transcribed territorially in the sea urchin embryo, and additional examples

(*Strongylocentrotus purpuratus*/cis-regulatory systems/*CyIIIa*/*Endo16*/spatial information processing)

CARMEN V. KIRCHHAMER, CHIOU-HWA YUH, AND ERIC H. DAVIDSON*

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

ABSTRACT **The cis-regulatory systems that control developmental expression of two sea urchin genes have been subjected to detailed functional analysis. Both systems are modular in organization: specific, separable fragments of the cis-regulatory DNA each containing multiple transcription factor target sites execute particular regulatory subfunctions when associated with reporter genes and introduced into the embryo. The studies summarized here were carried out on the *CyIIIa* gene, expressed in the embryonic aboral ectoderm and on the *Endo16* gene, expressed in the embryonic vegetal plate, archenteron, and then midgut. The regulatory systems of both genes include modules that control particular aspects of temporal and spatial expression, and in both the territorial boundaries of expression depend on a combination of negative and positive functions. In both genes different regulatory modules control early and late embryonic expression. Modular cis-regulatory organization is widespread in developmentally regulated genes, and we present a tabular summary that includes many examples from mouse and *Drosophila*. We regard cis-regulatory modules as units of developmental transcription control, and also of evolution, in the assembly of transcription control systems.**

Early development of the sea urchin, as of most other invertebrate forms, begins with the rapid division of the egg into polyclonal territories, each of which gives rise to specific embryonic cell types and structures [Type I embryogenesis (1)]. In these embryos, the blastomeres of the early territorial lineages begin to express some downstream differentiation genes even during cleavage. Differential control of the transcription of such genes typically depends on a complex array of interactions occurring within their cis-regulatory domains, between transcription factors presented in the various blastomere nuclei, and their target sites, which are "hard wired" into the regulatory DNA. There are two different aspects of developmental transcription factor function, of which in this review we focus on the first: Transcription factors bring regulatory information to the gene and once bound by means of interactions with one another, with ancillary proteins, and with the basal transcription apparatus (2), they execute their respective biochemical transcription control functions. The distribution of transcription factor activities and concentrations in the embryo reflects the specification events by which the blastomeres assume their territorial identities (3). Occupancy of target sites is the primary mechanism by which each copy of the gene receives the inputs that convey to it the spatial component of the embryo in which it resides, and the current stage of development. After gastrulation, regulatory requirements change. The embryo becomes a mosaic of differentiated cell types, and the expression of batteries of genes encoding cell type-specific proteins is stepped up.

Here we summarize recent studies from our laboratory that illuminate the cis-regulatory systems controlling expression of two genes during embryogenesis in the sea urchin *Strongylocentrotus purpuratus*. The *CyIIIa* gene, which encodes a particular form of cytoskeletal actin, is initially expressed in 11 clones of blastomeres that give rise to the aboral ectoderm of the embryo (4, 5). Activation of this gene in late cleavage serves as a marker of aboral ectoderm specification (6, 7). After gastrulation and throughout larval life *CyIIIa* is actively expressed in all cells of the differentiated squamous epithelium, which constitutes the definitive aboral ectoderm (6–9). The *Endo16* gene encodes a polyfunctional secreted glycoprotein (10). This gene is initially expressed in the eight clones of blastomeres constituting the vegetal plate of the blastula stage embryo; then, during gastrulation, throughout the archenteron to which the vegetal plate gives rise by invagination. After gastrulation *Endo16* transcription is shut off in the delaminating mesenchyme cells, in the foregut, and in the midgut, but is stepped-up in the definitive midgut (11, 12). The cis-regulatory systems of both *CyIIIa* and *Endo16* display a regional as opposed to dispersed or interspersed functional organization, in which given subelements of the regulatory DNA sequence perform specific developmental subfunctions. We consider the subelements of these cis-regulatory systems as control modules, from which the overall pattern of developmental gene expression is assembled.

Modular cis-Regulatory Organization

An experimental definition of a cis-regulatory module is a fragment of DNA containing multiple transcription factor target sites, which when tested in a gene transfer protocol produces some particular subelement of the overall pattern of expression of the gene. The cis-regulatory modules with which we are here concerned should be separable from the basal promoter of the gene and should work with heterologous basal promoter elements, as well as in various combinations with other modules. cis-regulatory modules are discrete subelements of the control system. Thus, the overall developmental pattern of expression of the gene is a direct function that can be experimentally determined of the activities of the individual modules that constitute the total cis-regulatory system. In the simplest case modular functions will be additive, e.g., where one module produces expression in tissue A and another in tissue B and the gene runs in A + B, but in others the relations are more complex, particularly where modular functions are negative.

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

*e-mail: davidson@mirsky.caltech.edu.

Since it receives different inputs in the form of the individual transcription factors that bind within it, the cis-regulatory module can be regarded as an information processing control element. For example, it may receive temporal and spatial information from several different transcription factors interacting at its target sites, and occupancy of these sites might be required for the module to generate its output. The output could be a signal to the basal apparatus causing transcription in a certain spatial domain of the embryo at a certain time. There is a clear implication in the concept of the cis-regulatory module that transcription factors bound adjacently within it interact locally, among themselves. The transcription factors may bind cooperatively, they may jointly mediate interactions with positive or negative adapter proteins (13), or in the case of a negative factor they may interfere with the function of a contiguous positive factor, etc. Intramodule biochemistry is likely to be as variable as the identities of the relevant transcription factors. Individual transcription factor target sites, if studied in isolation or if multimerized, may sometimes reproduce some of the functions of a module, at least qualitatively. Such a result would not vitiate the definition of the module of origin unless it were the case that none of the other interactions occurring within this module have any function, but we think this possibility unlikely. As we note below, in a gene where we have tested the functional significance of every detectable DNA/protein interaction, module function can be shown to be affected in some way by each and every one.

As we discuss in more detail later in this paper, there are three major reasons why it is important to identify and understand the modular functional organization of developmental cis-regulatory systems. (*i*) Many developmentally expressed genes appear to have a modular regulatory organization, so this is a major feature of gene regulation molecular biology. (*ii*) Because modular regulatory elements can be combined and can be used with diverse basal promoters, regulatory systems that produce novel patterns of developmental gene expression can be constructed in the laboratory. (*iii*) If we can do this experimentally, the same kind of process must have occurred in evolution.

Modular Functions in the *CyIIIa* Gene

Embryonic expression of the *CyIIIa* gene is controlled through interactions with at least 9 different transcription factors that bind at over 20 specific sites (14, 15). These are distributed nonrandomly over a 2300-bp cis-regulatory domain, as indicated in the diagram at the top of Fig. 1. This domain has been shown to be necessary and sufficient to generate the complete spatial, temporal, and quantitative pattern of *CyIIIa* gene expression in the embryo (16–19). Seven of the nine transcription factors have been cloned (the eighth is known to be a CCAAT-binding factor). An extensive series of gene transfer experiments, revealing the functions of a different subregion of the regulatory system by means of deletions and mutations, has been carried out by Kirchhamer and Davidson (16), and additional key observations are described in other recent studies (20, 21). As discussed in detail in ref. 16, we now know the functional significance of each specific interaction, except for some of the many individual SpGCF1 interactions (Fig. 1); these we discuss separately below.

The *CyIIIa* cis-regulatory system consists essentially of two complex modules that fulfill the criteria just outlined. These are the "proximal" and "middle" modules of Fig. 1 (16). In addition there is a distal region, which so far as we are aware, consists only of clustered SpGCF1 sites. Put most simply, the proximal module interprets the specification functions by which the oral and aboral ectoderm and vegetal territories are established, and its function is to activate the *CyIIIa* gene in the aboral ectoderm late in cleavage (6, 7). This module is largely responsible for the whole early pattern of *CyIIIa* expression. The middle module assumes the major role in controlling *CyIIIa* expression from the gastrula stage onward, driving the rate of expression to higher levels as the aboral ectoderm differentiates. The major positive input to the proximal module early in development is the yet uncharacterized "P1" factor, as in the absence of the P1 target sites the proximal

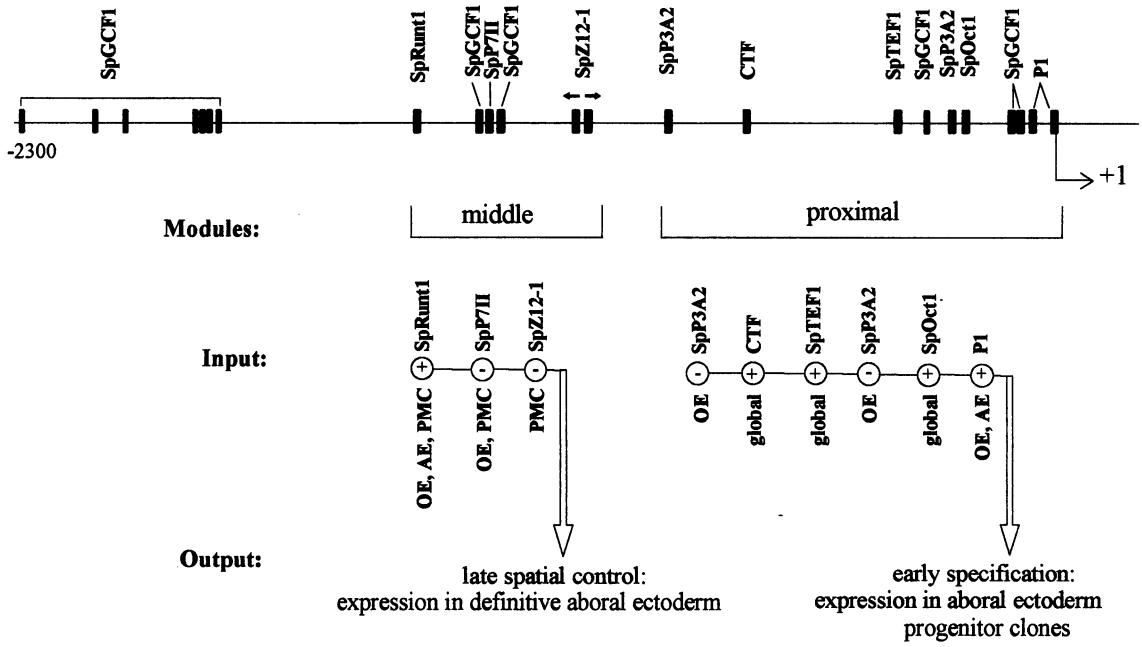

FIG. 1. Abstract summary of transcriptional organization within the *CyIIIa* cis-regulatory system. At top is a map, shown to scale, of the 2300-bp cis-regulatory domain, with transcription factor target sites indicated by solid rectangles. The broken arrow is the transcription start site. Transcription factors indicated above the line representing the DNA are all cloned except for the CCAAT-binding factor (CTF) and the P1 factor (see ref. 16 for transcription factor identities and individual references). The modules discussed in the text are indicated by horizontal brackets. Inputs are transcription factors that bind in each module, and the territories within the embryo where they bind are indicated below. OE, oral ectoderm; AE, aboral ectoderm; PMC, primary or skeletogenic mesenchyme; global, everywhere in embryo. Positive and negative functions are indicated by + and − symbols.

module is functionless. The P1 factor is apparently active throughout the ectoderm (oral + aboral). Expression is confined to the aboral ectoderm by the negatively acting SpP3A2 factor, which must become functional specifically in the oral ectoderm lineages (16, 17). The other positively acting factors that bind in the proximal module (i.e., besides the P1 factor; see Fig. 1) determine the level of expression and convey temporal information (16, 22). The middle module uses an entirely different set of transcription factors. Here the major positive input is provided by the SpRunt1 factor. SpRunt1 mRNA rises rapidly in the embryo after gastrulation as the middle module becomes the dominant *CyIIIa* regulatory element (20), and the P1 sites are no longer required for function. Two different negatively acting factors, SpZ12-1 and SpP7II (Fig. 1) that bind within the middle module, are required to confine expression to aboral ectoderm from the gastrula stage onward (16, 21). Like the SpP3A2 interaction within the proximal module, the SpZ12-1 and SpP7II interactions exercise their negative effects locally, and they are not required to prevent ectopic expression unless the remainder of the middle module is included in the reporter gene constructs (16). Thus, as shown diagrammatically in Fig. 1, each module integrates the inputs it receives and produces a certain regulatory output.

In the complete cis-regulatory system, the proximal and middle modules of the *CyIIIa* gene display at least three forms of interdependence, mainly quantitative, even though the functions of these modules can be separately analyzed. (*i*) Some of the positively acting factors of the proximal module, particularly the CCAAT binding and the SpTEF1 factors, boost the output of the middle module if their target sites are also present in the construct (ref. 16; see also ref. 22). (*ii*) Some site in the proximal module, yet undefined (but other than the P1 or SpP3A2 sites), is necessary for middle module function in synthetic transgenes. (*iii*) The SpRunt 1 interaction in the middle module boosts the output of the proximal module early in development (16, 20). In addition, the distal region of the *CyIIIa* cis-regulatory system increases the output of either or both of the two modules if included in expression constructs (16, 22, 23). This observation focuses attention on the role of the SpGCF1 interactions, for which sites occur in the proximal and middle modules, as well as distally. The location of these sites is highlighted in Fig. 2*A*. Zeller *et al.* (24) discovered that SpGCF1 is a protein that multimerizes once it binds to DNA. Hence it is capable of looping the regulatory DNA, as demonstrated *in vitro* by electron microscopic visualization. When recombinant SpGCF1 protein is added to the *CyIIIa* regulatory DNA sequence, every conceivable loop predicted by

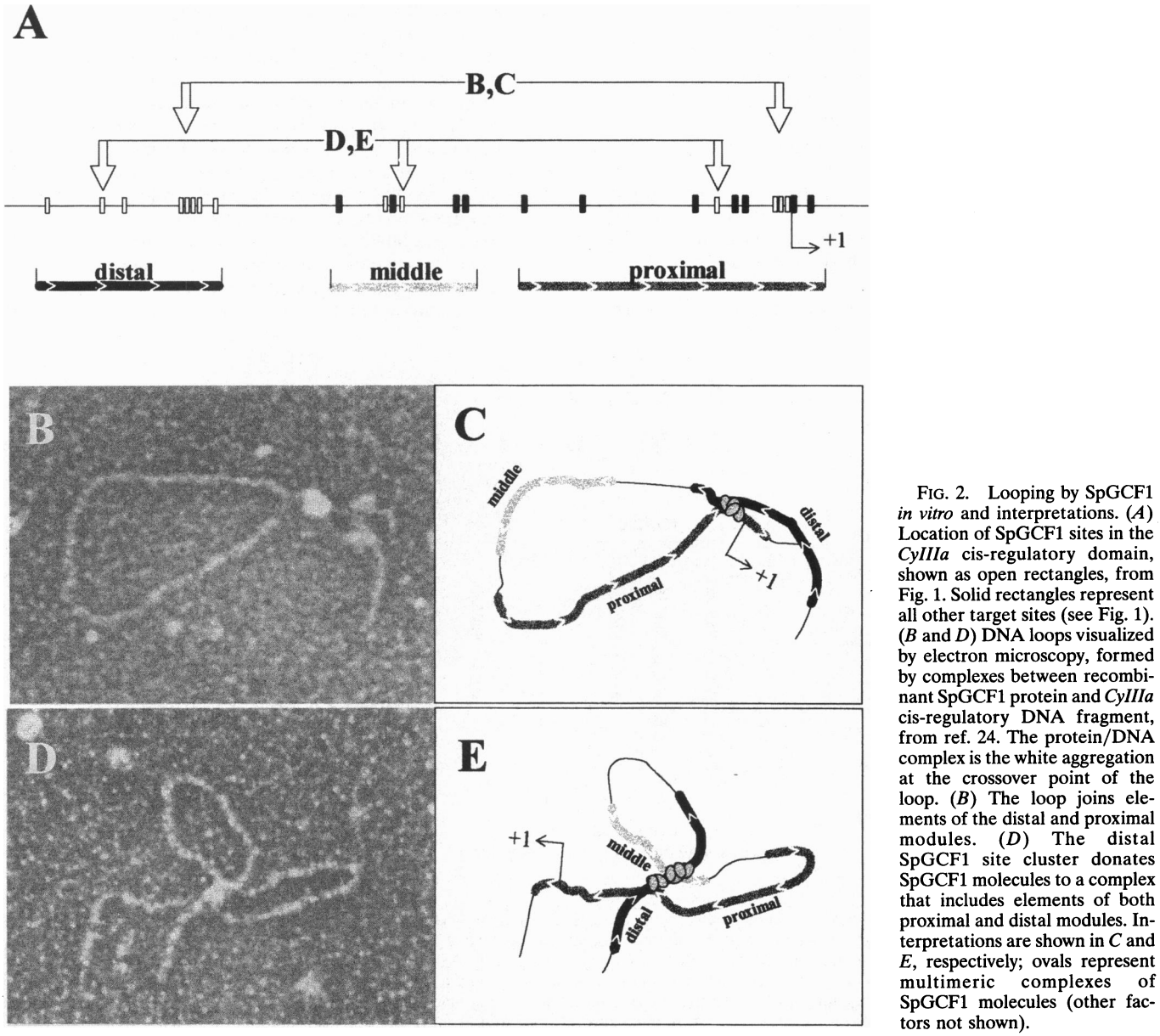

FIG. 2. Looping by SpGCF1 *in vitro* and interpretations. (*A*) Location of SpGCF1 sites in the *CyIIIa* cis-regulatory domain, shown as open rectangles, from Fig. 1. Solid rectangles represent all other target sites (see Fig. 1). (*B* and *D*) DNA loops visualized by electron microscopy, formed by complexes between recombinant SpGCF1 protein and *CyIIIa* cis-regulatory DNA fragment, from ref. 24. The protein/DNA complex is the white aggregation at the crossover point of the loop. (*B*) The loop joins elements of the distal and proximal modules. (*D*) The distal SpGCF1 site cluster donates SpGCF1 molecules to a complex that includes elements of both proximal and distal modules. Interpretations are shown in *C* and *E*, respectively; ovals represent multimeric complexes of SpGCF1 molecules (other factors not shown).

the locations of SpGCF1 target sites is formed (24). Examples and interpretations are shown in Fig. 2 *B* and *C*, and *D* and *E*. SpGCF1 could function by facilitating direct physical contact between elements of each module and between every module and the basal transcription apparatus. This may account for the positive function of SpGCF1 interactions, and explain how the distal cluster of SpGCF1 sites stimulates expression, i.e., by donating SpGCF1 factors to the complexes formed and thus increasing complex stability and the probability of productive interactions.

Modular Functional Organization of the *Endo16* cis-Regulatory System

In its initial phase of expression, in the early blastula, the *Endo16* gene is active in a ring consisting of about 64 cells, which are the immediate descendants of the 8 sixth cleavage "veg_2" blastomeres. These lineages constitute the progenitors of the archenteron. The ring surrounds the progenitors of the skeletogenic mesenchyme, and their sister cells, the small micromeres, in both of which lineages the *Endo16* gene is silent. The upper boundary of the ring of *Endo16* expression is the interface with the overlying ectoderm cells, i.e., the descendants of the sixth cleavage veg_1 lineages. A glance at a lineage map (e.g., refs. 4 and 9) shows that the blastomeres of the contiguous expressing and nonexpressing territories are close relatives: the initial veg_2 and veg_1 blastomeres are sister cells and the veg_2 blastomeres are "second cousins" of the sixth cleavage micromeres, i.e., they share the same "great grandparents" (the third cleavage vegetal blastomeres). When we examined the *Endo16* cis-regulatory system we found that an elaborate and complex mechanism is used to establish these early boundaries of *Endo16* expression.

As done earlier with the *CyIIIa* gene (14, 15), analysis of the *Endo16* regulatory system was begun by mapping all of the specific sites of DNA/protein interaction within the *Endo16* cis-regulatory domain that is necessary and sufficient to provide a complete and normal pattern of embryonic *Endo16* expression (25). This domain turns out to be about 2200 bp long, extending upstream of the start site. Within this sequence are at least 30 target sites for 13 different factors. The sites are discontinuously distributed as shown at the top of Fig. 3, in addition to a large number of SpGCF1 sites. The *Endo16* factors are not yet cloned, and so we know a lot less about the individual inputs into each regulatory module than we do about the *CyIIIa* cis-regulatory modules. The outputs of each subregion of the *Endo16* cis-regulatory domain are revealed by a detailed series of reporter gene transfer experiments carried out by Yuh and Davidson (26). These functional outputs are shown in simplified form in the diagram of Fig. 3. Here it can be seen that except for the most distal region, module G, and the proximal cluster of sites called module B, the whole regulatory system appears to be engaged specifically in producing the correct early pattern of *Endo16* expression in the vegetal plate and preventing incorrect expression in the adjacent skeletogenic and ectodermal territories.

As with the *CyIIIa* gene, the spatial domain of the embryo in which positive *Endo16* regulatory functions are active in early development exceeds the correct domain, and negative interactions are required to confine expression to the appropriate embryonic territories. Module A, the most proximal cluster of sites, by itself promotes reporter gene expression in the vegetal plate and in the structure descended from it, the archenteron, but also in the skeletogenic mesenchyme and the overlying ectoderm (26). These are, as indicated above, the tissues deriving from the two blastomere territories abutting the ring of vegetal plate cells to which *Endo16* expression should be confined. Modules E and F (independently in our experimental tests) act specifically to shut off ectopic expression in veg_1 ectoderm, whereas module DC extinguishes ectopic expression specifically in skeletogenic mesenchyme. The early specification of the vegetal plate territory depends on positive inductive signaling from the skeletogenic to the vegetal plate territory (27–29), and is likely to involve negative interterritorial signaling at both skeletogenic and ectodermal boundaries as well (25). Modules E, F, and DC are negative transcriptional control subsystems, and the experiments of Yuh and Davidson (26) indicate that they act by precluding the positive output of module A in ectoderm and skeletogenic mesenchyme progenitors, respectively. Thus, they are likely to be among the regulatory termini of the signal transduction pathways that determine the boundaries of the vegetal plate. We are able to perturb these signaling functions by treatment with LiCl. This teratogenic agent extends the boundary of the vegetal plate, and the domain of *Endo16* expression, at the expense of adjacent ectoderm (12). Modules DC, E, and F all are converted into positively acting regulatory elements by

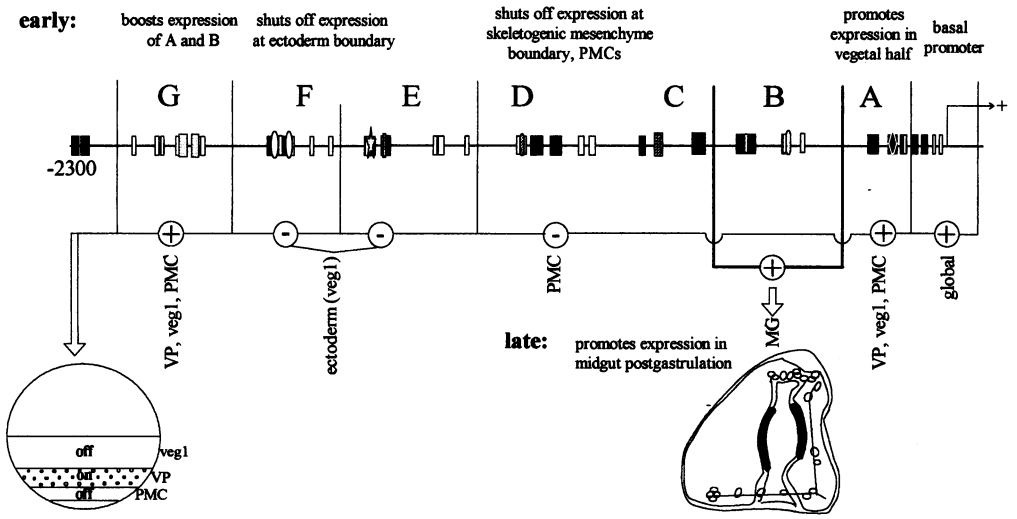

FIG. 3. Modular organization of the *Endo16* cis-regulatory system. The sites of specific DNA/protein interactions are indicated as rectangles, as in Fig. 2*A*, with SpGCF1 sites shown as open rectangles. Sites that are unique in each region are indicated by distinctive symbols. As shown by Yuh *et al.* (25), 13 different factors bind at these sites (9 uniquely), in addition to SpGCF1. The modular elements described in the gene transfer experiments of Yuh and Davidson (26) are indicated by the vertical lines and the uppercase letters above the horizontal lines representing the DNA. Below this line the spatial domains of the embryo where each module is active are indicated: MG, midgut; veg_1, tier of ectoderm cells overlying VP (vegetal plate); PMC, primary or skeletogenic mesenchyme territory. The functions of each module are summarized in the drawings in lower region of the figure.

LiCl treatment, and, just as they work via module A in controlling ectopic expression, LiCl response also requires module A (26). All three of the DC, E, and F spatial repressor elements function across the hundreds of base pairs separating them from module A. However, their absolute requirement for the presence of module A means that they do not interact directly with the basic transcription apparatus, nor do they function as do the repressors of the *CyIIIa* middle module discussed above, which evidently interact with activators bound at a nearby upstream position. It may be significant for the intermodule communication required by this mode of function that E, F, DC, and A all include SpGCF1 sites (Fig. 3), as well as target sites for other factors that occur in several different modules (25), and that could function similarly to SpGCF1.

The function of the early *Endo16* control system is to process the positive and negative spatial information that is generated as the vegetal plate is specified. In contrast, the late control system that promotes stepped-up *Endo16* transcription in the midgut includes no negative cis-regulatory interactions, at least none that we have been able to detect, even though *Endo16* expression is made to disappear from all regions of the archenteron except the midgut in the course of gut regionalization. Perfect midgut-specific expression is mediated by module B alone, the output of which is boosted by module G; however,

Table 1. *Drosophila* and mouse examples of modular spatial control elements

Gene	No. of modules studied	Module controls expression in, or regulatory function (minimum length of sequence studied, kb)	Ref.
Drosophila			
deadpan	2	CNS (2.5) PNS (1.5)	30
scratch	2	CNS (5.5) PNS (4.5)	30
snail	3	CNS (0.58) PNS (0.6) mes (0.36)	31
rhomboid	2	NE (0.3) Midline (0.3)	32
even-skipped	3	Stripe 2 (0.7) Stripe 3/7 (0.5) Autoregulation (0.7)	33
Ultrabithorax	3	vis mes (1.0) ect (0.04) PS 6, 8, 10, 12 (0.5)	34–36
Krüppel	7	Central early expression (1.2, 1.7) A domain late blastoderm (1.7, 1.7) Polar expression (8) Muscle precursors (0.3) Bolwig organ (0.6)	37
hairy	4	Stripes 1/5 (2.8) Stripes 6 (3.9) Stripes 7 (1.2) Stripes 3/4 (1.4)	38–40
decapentaplegic	4	vis mes (3) vis mes, PS7 (0.67) Dorsal ect (0.8) Lateral ect, midgut end (0.48)	41–45
Yolk protein genes 1 and 2	2	Ovary (0.3, 0.1)	46
achaete/scute	5	Proneural expression for: A wing margin and two vein sensilla (3.7) A and P dorsoventral macrochaetae (5.7) Dorsal and tegula radii sensillae (1.2) Proximal tegula (3.8) A notopleural macrochaetae (6)	47
Delta	9	st5-st6 expression domains (0.86, 2, 1, 3) Pair rule gene pattern st9 (1) Segment polarity expression >st9 (1.8) SI and SII neuroblasts (1, 3.3) Midline (0.8)	48
Zerknüllt	3	Dorsal expression (0.4) Ventral repression (0.2) Dorsolateral repression (0.6)	49
Mouse			
Hoxa-1 and *2*	3	rhomb 2 (2.5) rhomb 4 (1.25) NC/FP/caudal NT/gut, epi (0.55)	50
Hoxa-7	4	Sets A boundary (0.11) Sets P boundary in mes, ect, restricts expression to ventral NT (0.13) Negative element restricts expression to prevertebrae (1.3) Positive global enhancer (1.59)	51
Hoxb-1	4	Early mes (0.8) Early NE (0.8) rhomb 3-5, neural crest (0.33) expression to rhomb 4; RARE (0.74)	52, 53
Hoxb-4	2	mes, P NT enhanced (3 mouse; 0.9 chicken) A NT boundary rhomb 6/7 (1.5 mouse; 1.4 chicken)	54, 55
Hoxb-8	3	A cervical NT boundary; P mes expression (1.1) Cells in NT, spinal ganglia (2.8) NT, mes enhancer (0.45)	56, 57
Hoxd-11	3	Early PS, somites, limb bud (0.3) Later trunk expression, limb (4) Negative regulatory elements (2.8)	58

In most cases, listed fragment sizes were those used in gene transfer experiments, and the actual size of the modules could be much smaller. In some cases, other, possibly modular functions were discovered as well as those included here, but were omitted from this table because the fragments investigated displayed multiple functions or overlapped one another. Overlapping modules were included in the case of the *Delta* gene, where the early expression modules overlap some of the later expression modules. Thus, the *number* of modules shown is minimal, both because of these omissions, and because other functional modules can be presumed yet undiscovered. Where multiple fragment sizes are shown each produced expression in the domain indicated. A, anterior; CNS, central nervous system; ect, ectoderm; end, endoderm; epi, epithelium; FP, floorplate; mes, mesoderm; NC, notochord; NE, neuroectoderm; NT, neural tube; P, posterior; PNS, peripheral nervous system; PS, parasegment; RARE, retinoic acid response element; rhomb, rhombomere; st, stage; vis, visceral.

the latter has no spatial control function *per se*, because it also boosts the output of the early control system. Therefore, we believe that module B includes a target site for a midgut-specific, positive transcription factor, spatial expression of which is controlled at the next higher level of regulatory hierarchy. This would be the cis-regulatory region of the gene encoding the midgut-specific factor. In recent unpublished work, we in fact identified a specific oligonucleotide that produces midgut-specific expression when combined with the basal promoter in a reporter construct, and its sequence probably includes the relevant module B target site. We will now be able to clone the factor and directly test this proposition.

Our image of the *Endo16* regulatory system as an assemblage of interactive, but discrete modular subelements extends to a quantitative level. For example, as will be reported elsewhere, the sum of the separate activities of the positive A and B modules, expressed mathematically as time functions of chloramphenicol acetyltransferase reporter enzyme production, equals the activity of a construct in which these modules are physically linked. Similarly, the activities of the negative modules can be treated as independent time functions that modify the output of the A module when physically combined with it.

Modular cis-Regulatory Control Systems Are Common in Genes Expressed During Development

In Table 1 we have collected examples of genes from *Drosophila* and mouse, the cis-regulatory systems of which have been shown to have a modular functional character. This is most easily seen in cases where a given gene is expressed in different spatial domains, sometimes at different times, so that different sets of transcription factors are required to establish each spatial subelement of the overall expression pattern. The criteria used to select the examples in Table 1 (which is scarcely intended to be inclusive) are: (*i*) A specific DNA fragment from within the cis-regulatory domain has been shown to generate a subelement of the overall pattern of expression in a gene transfer experiment; and (*ii*) More than one such regulatory module has been identified from within the same cis-regulatory system. In some cases the identities of the relevant transcription factors are known, but in many cases these have yet to be determined and only the modular regulatory outputs are reported. Nor are the actual sequence boundaries of the cis-regulatory modules often established. Table 1 shows, nonetheless, that there are many developmentally regulated genes in which specific control functions are modified by given subregions of the regulatory DNA that can be assayed independently.

Of course there are many cis-regulatory systems that are not characterized as completely modular. For example, in the *hairy* gene of *Drosophila* the elements responsible for each of the stages of expression seem closely intertwined and cannot be separated on different pieces of DNA (27). Often genes that lie downstream of more complex spatial information processing systems may look like "single module" genes. For instance, if one examined only late embryonic expression of *Endo16*, the B module alone would appear to suffice, as discussed above. The *ftz* gene, a "downstream" pair rule gene, appears to require only a small cluster of sites for its regulation, although an internally complex one at that (59, 60). Similarly, the *SM50* gene of the sea urchin, which encodes a skeletal matrix protein and which is activated in the autonomously specified skeletogenic mesenchyme lineage, is, so far as we know, a single module gene (61). Among genes expressed later in development, many genes encoding terminal differentiation products may also operate under the control of single cis-regulatory modules among which are many "tissue-specific enhancers."

Significance of Modular cis-Regulatory Organization

All modules of a given cis-regulatory system, but the one closest to the promoter and the basal transcription apparatus, must transmit the outcome of their internal regulatory interactions over some distance in terms of DNA sequence to communicate with the basal apparatus. The same is true for intercommunication between modules (see Figs. 1–3, for examples). Interaction over cis-regulatory sequence distance may involve adapter proteins or may be mediated directly by the transcription factors, but in either case it probably involves DNA binding and looping, a mechanism that is largely insensitive to the particular length of sequence that separates the interacting cis-regulatory elements. Thus, a general observation is that when cis-regulatory modules are combined in synthetic constructs, they work irrespective of exact spacing and sometimes even order. This basic feature vastly facilitates the experimental combination of diverse modular regulatory elements for the purpose of building new regulatory systems. Thus, it becomes easy experimentally to invent cis-regulatory systems that have novel developmental patterns of expression. For example, Levine and colleagues (62) have combined the *eve* stripe 2 module with the *rhomboid* neuroectodermal lateral stripe module, producing a synthetic regulatory system that generates an A/P + D/V "cross" of reporter gene expression on each side of the *Drosophila* embryo. We have carried out similar experiments (unpublished data) combining skeletogenic and vegetal plate patterns of expression by inserting cis-regulatory modules from the *SM50* and *Endo16* genes in the same vector. Such synthetic modular combinations can of course be used not only to study cis-regulatory functions, but also to drive expression of any desired gene product in novel but predictable patterns. A whole new horizon of research strategies in developmental gene expression is thus within view.

Twenty-five years ago Britten and Davidson (63) argued that transpositional insertions of control elements from elsewhere in the genome into given cis-regulatory domains is a major process in the evolution of novel developmental regulatory systems, and hence of novel metazoan forms. Combination of modular regulatory elements (as we now do in the laboratory) must indeed be the mechanism underlying the evolutionary assembly of regulatory systems such as those discussed above, and those listed in Table 1. Elsewhere in this Colloquium, Britten presents a list of apparently transposed cis-regulatory elements that now affect the regulatory behavior of their host genes. There are two sea urchin examples that are particularly germane. The cis-regulatory regions of several of the Ca^{2+} binding *Spec* genes have been shown by Klein and colleagues (64, 65) to contain what they term "RSR" modules, which in reporter constructs confer aboral ectoderm expression, where R is a repeated sequence and S is a sequence that contains transcription factor binding sites. In another case a "cassette" or module that includes many of the *CyIIIa* target sites has been identified in an intron of the sea urchin metallothionein gene, and has been shown to confer aboral ectoderm specific expression to this gene (66, 67).

To summarize in one sentence, cis-regulatory modules can be considered both the information processing units of developmental gene regulation and also the building blocks from which complex developmental regulatory systems have been assembled during evolution.

The research summarized in this review was supported by National Institutes of Health Grant HD-05753.

1. Davidson, E. H. (1991) *Development (Cambridge, U.K.)* **113,** 1–26.
2. Tjian, R. & Maniatis, T. (1994) *Cell* **77,** 5–8.
3. Davidson, E. H. (1990) *Development (Cambridge, U.K.)* **108,** 365–389.
4. Cameron, R. A., Hough-Evans, B. R., Britten, R. J. & Davidson, E. H. (1987) *Genes Dev.* **1,** 75–85.
5. Cameron, R. A., Fraser, S. E., Britten, R. J. & Davidson, E. H. (1990) *Dev. Biol.* **137,** 77–85.
6. Cox, K. H., Angerer, L. M., Lee, J. J., Davidson, E. H. & Angerer, R. C. (1986) *J. Mol. Biol.* **188,** 159–172.

7. Lee, J. J., Calzone, F. J. & Davidson, E. H. (1992) *Dev. Biol.* **149,** 415–431.
8. Cameron, R. A., Britten, R. J. & Davidson, E. H. (1989) *Mol. Reprod. Dev.* **1,** 149–155.
9. Cameron, R. A. & Davidson, E. H. (1991) *Trends Genet.* **7,** 212–218.
10. Soltysik-Espanola, M., Klinzing, D. C., Pfarr, K., Burke, R. D. & Ernst, S. G. (1994) *Dev. Biol.* **165,** 73–85.
11. Nocente-McGrath, C., Brenner, C. A. & Ernst, S. G. (1989) *Dev. Biol.* **136,** 264–272.
12. Ransick, A., Ernst, S., Britten, R. J. & Davidson, E. H. (1993) *Mech. Dev.* **42,** 117–124.
13. Komachi, K., Redd, M. J. & Johnson, A. D. (1994) *Genes Dev.* **8,** 2857–2867.
14. Calzone, F. J., Theze, N., Thiebaud, P., Hill, R. L., Britten, R. J. & Davidson, E. H. (1988) *Genes Dev.* **2,** 1074–1088.
15. Theze, N., Calzone, F. J., Thiebaud, P., Hill, R. L., Britten, R. J. & Davidson, E. H. (1990) *Mol. Reprod. Dev.* **25,** 110–122.
16. Kirchhamer, C. V. & Davidson, E. H. (1996) *Development (Cambridge, U.K.)* **122,** 333–348.
17. Hough-Evans, B. R., Franks, R. R., Cameron, R. A., Britten, R. J. & Davidson, E. H. (1987) *Dev. Biol.* **121,** 576–579.
18. Hough-Evans, B. R., Franks, R. R., Zeller, R. W., Britten, R. J. & Davidson, E. H. (1990) *Development (Cambridge, U.K.)* **110,** 41–50.
19. Livant, D., Cutting, A., Britten, R. J. & Davidson, E. H. (1988) *Proc. Natl. Acad. Sci. USA* **85,** 7607–7611.
20. Coffman, J. A., Kirchhamer, C. V., Harrington, M. G. & Davidson, E. H. (1996) *Dev. Biol.* **174,** 43–54.
21. Wang, D. G.-W., Kirchhamer, C. V., Britten, R. J. & Davidson, E. H. (1995) *Development (Cambridge, U.K.)* **121,** 1111–1122.
22. Franks, R. R., Anderson, R., Moore, J. G., Hough-Evans, B. R., Britten, R. J. & Davidson, E. H. (1990) *Development (Cambridge, U.K.)* **110,** 31–40.
23. Flytzanis, C. N., Britten, R. J. & Davidson, E. H. (1987) *Proc. Natl. Acad. Sci. USA* **84,** 151–155.
24. Zeller, R. W., Griffith, J. D., Moore, J. G., Kirchhamer, C. V., Britten, R. J. & Davidson, E. H. (1995) *Proc. Natl. Acad. Sci. USA* **92,** 2989–2993.
25. Yuh, C.-H., Ransick, A., Martinez, P., Britten, R. J. & Davidson, E. H. (1994) *Mech. Dev.* **47,** 165–187.
26. Yuh, C.-H. & Davidson, E. H. (1996) *Development (Cambridge, U.K.)* **122,** 1069–1082.
27. Davidson, E. H. (1989) *Development (Cambridge, U.K.)* **105,** 421–445.
28. Ransick, A. & Davidson, E. H. (1993) *Science* **259,** 1134–1138.
29. Ransick, A. & Davidson, E. H. (1995) *Development (Cambridge, U.K.)* **121,** 3215–3222.
30. Emery, J. F. & Bier, E. (1995) *Development (Cambridge, U.K.)* **121,** 3549–3560.
31. Ip, Y. T., Levine, M. & Bier, T. (1994) *Development (Cambridge, U.K.)* **120,** 199–207.
32. Ip, Y. T., Park, R. E., Kosman, D., Bier, E. & Levine, M. (1992) *Genes Dev.* **6,** 1728–1739.
33. Harding, K., Hoey, T., Warrior, R. & Levine, M. (1989) *EMBO J.* **8,** 1205–1212.
34. Bienz, M. & Lawrence, P. A. (1988) *Cell* **53,** 567–576.
35. Qian, S., Capovilla, M. & Pirrotta, V. (1991) *EMBO J.* **10,** 1415–1425.
36. Zhang, C.-C., Muller, J., Hoch, M., Jäckle, H. & Bienz, M. (1991) *Development (Cambridge, U.K.)* **113,** 1171–1179.
37. Hoch, M., Schröder, C., Seifert, E. & Jäckle, H. (1990) *EMBO J.* **9,** 2587–2595.
38. Riddihough, G. & Ish-Horowicz, D. (1991) *Genes Dev.* **5,** 840–854.
39. Howard, K. R. & Struhl, G. (1990) *Development (Cambridge, U.K.)* **110,** 1223–1231.
40. Langeland, J. A. & Carroll, S. B. (1993) *Development (Cambridge, U.K.)* **117,** 585–596.
41. Masucci, J. D. & Hoffmann, F. M. (1993) *Dev. Biol.* **159,** 276–287.
42. Huang, J.-D., Schwyter, D. H., Shirokawa, J. M. & Courey, A. J. (1993) *Genes Dev.* **7,** 694–704.
43. Jackson, P. D. & Hoffmann, M. (1994) *Dev. Dyn.* **199,** 28–44.
44. Capovilla, M., Brandt, M. & Botas, J. (1994) *Cell* **76,** 461–475.
45. Hursh, D. A., Padgett, R. W. & Gelbart, W. M. (1993) *Development (Cambridge, U.K.)* **117,** 1211–1222.
46. Logan, S. K., Garabedian, M. J. & Wensink, P. C. (1989) *Genes Dev.* **3,** 1453–1461.
47. Gomez-Skarmeta, J. L., Rodriguez, I., Martinez, C., Culi, J., Ferres-Marco, D., Beamonte, D. & Modolell, J. (1995) *Genes Dev.* **9,** 1869–1882.
48. Haenlin, M., Kurisch, M., Kramatschek, B. & Campos-Ortega, J. A. (1994) *Mech. Dev.* **47,** 99–110.
49. Doyle, H. J., Kraut, R. & Levine, M. (1989) *Genes Dev.* **3,** 1518–1533.
50. Frasch, M., Chen, X. & Lufkin, T. (1995) *Development (Cambridge, U.K.)* **121,** 957–974.
51. Puschel, A. W., Balling, R. & Gruss, P. (1990) *Development (Cambridge, U.K.)* **108,** 435–442.
52. Marshall, H., Studer, M., Pöpperl, H., Aparicio, S., Kuroiwa, A., Brenner, S. & Krumlauf, R. (1994) *Nature (London)* **370,** 567–571.
53. Studer, M., Pöpperl, H., Marshall, H., Kuroiwa, A. & Krumlauf, R. (1994) *Science* **265,** 1728–1732.
54. Whiting, J., Marshall, H., Cook, M., Krumlauf, R., Rigby, P. W. J., Stott, D. & Allemann, R. K. (1991) *Genes Dev.* **5,** 2048–2059.
55. Morrison, A., Chaudhuri, C., Ariza-McNaughton, L., Muchamore, I., Kuroiwa, A. & Krumlauf, R. (1995) *Mech. Dev.* **53,** 47–59.
56. Charité, J., de Graaf, W., Vogels, R., Meijlink, F. & Deschamps, J. (1995) *Dev. Biol.* **171,** 294–305.
57. Vogels, R., Charité, J., de Graaf, W. & Deschamps, J. (1993) *Development (Cambridge, U.K.)* **118,** 71–82.
58. Gérard, M., Duboule, D. & Zákány, J. (1993) *EMBO J.* **12,** 3539–3550
59. Dearolf, C. R., Topol, J. & Parker, C. S. (1989) *Genes Dev.* **3,** 384–398.
60. Topol, J., Dearolf, C. R., Prakash, K. & Parker, C. S. (1991) *Genes Dev.* **5,** 855–867.
61. Makabe, K. W., Kirchhamer, C. V., Britten, R. J. & Davidson, E. H. (1995) *Development (Cambridge, U.K.)* **121,** 1957–1970.
62. Gray, S., Szymanski, P. & Levine, M. (1994) *Genes Dev.* **8,** 1829–1838.
63. Britten, R. J. & Davidson, E. H. (1971) *Q. Rev. Biol.* **46,** 111–138.
64. Gan, L., Zhang, W. & Klein, W. H. (1990) *Dev. Biol.* **139,** 186–196.
65. Gan, L. & Klein, W. H. (1993) *Dev. Biol.* **157,** 119–132.
66. Nemer, M., Bai, G. & Stuebing, E. W. (1993) *Proc. Natl. Acad. Sci. USA* **90,** 10851–10855.
67. Nemer, M., Stuebing, E. W., Bai, G. & Parker, H. R. (1995) *Mech. Dev.* **50,** 131–137.