Development 124, 1851-1864 (1997)
Printed in Great Britain © The Company of Biologists Limited 1997
DEV5122

The hardwiring of development: organization and function of genomic regulatory systems

Maria I. Arnone[1,2] and Eric H. Davidson[2,*]

[1]Stowers Institute for Medical Research, Division of Biology, California Institute of Technology, Pasadena, CA 91125, USA
[2]Division of Biology, California Institute of Technology, Pasadena, CA 91125, USA

*Author for correspondence (e-mail: davidson@mirsky.caltech.edu)

SUMMARY

The gene regulatory apparatus that directs development is encoded in the DNA, in the form of organized arrays of transcription factor target sites. Genes are regulated by interactions with multiple transcription factors and the target sites for the transcription factors required for the control of each gene constitute its *cis*-regulatory system. These systems are remarkably complex. Their hardwired internal organization enables them to behave as genomic information processing systems. Developmental gene regulatory networks consist of the *cis*-regulatory systems of all the relevant genes and the regulatory linkages amongst them. Though there is yet little explicit information, some general properties of genomic regulatory networks have become apparent. The key to understanding how genomic regulatory networks are organized, and how they work, lies in experimental analysis of *cis*-regulatory systems at all levels of the regulatory network.

Key words: gene regulation, *cis*-regulatory element, regulatory module, transcription factor, gene network

INTRODUCTION

A defining property of metazoans is the regulatory program for development that is hardwired in their genomic DNA sequence. Developmental programs are largely encoded in *cis*-regulatory genomic sequences, through which the activities of individual genes are governed. As everyone knows, *cis*-regulatory target sites are recognized sequence-specifically by the transcription factors required to control the expression of each gene. The target sites specify and anchor the relevant transcription factors in appropriate positions with respect to one another, and to the basal transcription apparatus. These factors, and other proteins that in turn bind to them, determine the rate of transcription and mediate the accurate activation or repression of the gene in developmental time, and in morphological space, i.e., in the appropriate cell lineage(s), cell type(s) or region(s) of the developing organism. The identities of the genes encoding those transcription factors that, in terms of causality, lie directly upstream of any given *cis*-regulatory system are therefore determined by its target sites. Thus an initial point of departure: knowledge of *cis*-regulatory systems will indicate both their internal workings and also the specific interconnections amongst them, i.e., the structure of the gene regulatory network.

These statements are obvious but their consequences and applications are not. A decade ago no one predicted the actual complexity or character of *cis*-regulatory systems, nor for that matter the shape of the regulatory networks that are erected as genes encoding transcription factors themselves are differentially activated during development. In terms of direct experimental evidence, the forms of these networks are just beginning to be perceptible. Yet among the most fundamental objectives must be to understand the flow of regulatory information from the genome, since that is how the whole process of development is organized. From this 'genomic' standpoint, most current effort in developmental biology would seem focused on the inputs and outputs of genomic regulatory systems rather than on these systems themselves. A particularly important class of inputs includes the signaling interrelations, which ultimately affect gene expression in response to the identity and activity of neighboring cells. The output of the developmental regulatory system is the controlled expression of thousands of genes encoding proteins that endow cells with their various differentiated, communicative and morphogenetic competence. Knowledge of their inputs and outputs informs us about what genomic regulatory systems do and what they respond to. But to find out how they really work requires direct examination of the regulatory DNA sequence elements themselves, which is to say, of developmental *cis*-regulatory systems.

PROPERTIES OF *CIS*-REGULATORY SYSTEMS

Modular regulatory organization

The complexity of the job that a *cis*-regulatory system has to do is reflected in its primary structure. When a gene is to be expressed (or repressed) in a number of different circumstances in a developing organism, it is usually found that separate *cis*-regulatory subelements carry out different parts of the overall regulatory job. We refer to these subelements as regulatory modules. The experimental definition of such a module is a

fragment of *cis*-regulatory DNA that, when linked to a reporter gene and transferred into an appropriate cell, executes a regulatory function that is a subfraction of the overall combined regulatory function executed by the complete system. Individual modules are always found to contain multiple transcription factor target sites, and these contribute in various ways to the overall regulatory output. There are now a great many excellent examples of modular *cis*-regulatory organization, particularly in *Drosophila*. The most remarkable cases are found among genes encoding transcription factors that are expressed in complex spatial patterns and at different times. For instance, the *cis*-regulatory system of the *Kruppel* gene includes modules that direct expression specifically to certain regions of early or later blastoderm stage embryos; plus at least ten different elements that specify expression in Malpighian tubule precursors, Bolwig organ, amnioserosa, muscle precursors and various cells of the developing nervous system (Hoch et al., 1990). Similarly, in the *cis*-regulatory system governing expression of the proneural *achaete* and *scute* genes, nine different regulatory elements or modules are individually responsible for expression of these transcription factors in diverse proneural clusters of the developing wing imaginal disc alone (Gomez-Skarmeta et al., 1995). Modular *cis*-regulatory organization is the subject of a recent review (Kirchhamer et al., 1996a) in which many different cases from *Drosophila*, mouse and sea urchin are cited, so it is unnecessary to cite further examples here. Suffice it to say that modularity is a common feature of developmental *cis*-regulatory organization, observed particularly when genes must be expressed, or silenced, in multiple spatial domains of the organism. Put another way, target sites for the set of transcription factors required to generate a given spatial regulatory output are often found clustered together more or less contiguously, within a given sequence element of the *cis*-regulatory DNA. This is not always so, however. Some of the best examples of discrete *cis*-regulatory modules are the individual stripe elements found in the *evenskipped* (*eve*) and *paired* pair-rule genes (Harding et al., 1989; Gutjahr et al., 1994), but the regulatory sites that generate the similar stripes of the *hairy* pair-rule gene are largely interspersed with one another (Howard and Struhl, 1990; Riddihough and Ish-Horowitz, 1991). And, as we discuss below, many genes expressed in specific spatial domains of the organism appear to be controlled by smaller *cis*-regulatory systems that have the structural and functional features of single modules.

It is clear from thousands of gene transfer experiments that individual *cis*-regulatory modules have the intrinsic property that they can transmit regulatory outputs to the basal transcription apparatus (BTA), and/or to the modules to which they are linked. The communication is performed by transcription factors anchored within the modules, or by proteins that bind to these factors rather than directly to the DNA. Regulatory communication over hundreds or thousands of DNA base pairs almost certainly involves DNA looping. This follows from the great distances that often separate *cis*-regulatory modules from the basal transcription apparatus that they control. Among the most extreme examples is the *Ubx* gene of *Drosophila*, where the individual regulatory modules are located throughout the 70 kb gene (Bender et al., 1983; Martin et al., 1995). Modular *cis*-regulatory elements are conventionally found to work with heterologous promoters (here 'promoter' denotes the platform on which the BTA assembles), for instance, the commonly used SV40 promoter. Therefore, in general, it is the *cis*-regulatory elements, and not the BTA, that are responsible for the specificity of the developmental regulatory outputs observed in gene transfer studies. The most dramatic evidence is that two modular regulatory elements can be recombined in novel expression constructs, both servicing the same BTA, wherein they generate patterns of transgene activity that are spatially an accurate sum of the patterns mediated by the individual modular components. Two examples are the crossed pattern of lateral and dorsoventral stripes generated by constructs containing both the stripe 2 module from the *Drosophila eve* gene and the longitudinal neuroectodermal expression module from the *rhomboid* gene (Gray et al., 1994); and the combined gut and skeletogenic mesenchyme pattern of expression that is generated by transgenes in which regulatory elements from the gut-specific *Endo16* gene and the skeletogenic *SM50* gene of the sea urchin are linked (Kirchhamer et al., 1996b). Both these results require that each of the constituent modules individually direct the activity of the BTA when the modular target sites are engaged by transcription factors in the appropriate embryonic spatial domains.

The term 'enhancer' to some extent overlaps the meaning assigned here to the term module. We have avoided the use of 'enhancer,' however, because it has accumulated too many different denotations. Some of these refer merely to stepping up the rate of expression, others to causing the specificity of expression. 'Enhancer' sometimes also refers to single factors or target sites for single factors (e.g., as in 'enhancer protein'). As the examples that we now take up abundantly demonstrate, transcription factors never work alone and, in life, developmental *cis*-regulatory outputs never devolve entirely from *cis*-regulatory sites for a single species of factor.

Intramodular complexity

Fig. 1 displays diagrammatically the organization of eleven different *cis*-regulatory systems for which the following obtains: sequences of the specific transcription factor target sites have been determined by gel shift mapping or footprinting or other means; the factors have been identified or at least characterized sufficiently that the number of different DNA-binding factors and some of their properties are understood; and functional information has been obtained in detailed gene transfer studies. The examples in Fig. 1 derive from three very different metazoan phyla, viz echinoderms, chordates and arthropods, and yet at a glance one can see that in internal organization the *cis*-regulatory elements are all similarly complex. The size and scale of the *cis*-regulatory elements portrayed in Fig. 1 are indicated by the genomic positions of their termini, which are shown with respect to the transcription start sites. Most are single modules a few hundred base pairs long; note that the *eve* (Fig. 1H) and the myosin light chain (Fig. 1E) examples each include two different widely separated modules. The sea urchin examples (Fig. 1A,B) are several times larger, since they represent complete *cis*-regulatory systems, each of which includes several individual modules. The diversity of the factors binding in each element is indicated by the shapes of the symbols, and some aspects of their functional roles, where known, are given in Fig. 1 by the color coding (see legend). Briefly, factors present ubiquitously are shown in blue; positively acting factors that are spatially

localized in the organism are in green; repressors, which are here all spatially localized, are in red; and factors whose presentation and/or activity is known to be directly affected by signal transduction systems are shown hatched. All of the *cis*-regulatory elements included in Fig. 1 function differentially in the course of development.

Some fascinating generalizations emerge from Fig. 1, and the first of these to be considered is the insight this compilation provides into the complexity of interactions within the developmental *cis*-regulatory module. Complexity is here defined simply as the number of diverse interactions, i.e., of different transcription factors bound per module (whether or not the same factors interact in other modules of the same gene and irrespective of the number of target sites per factor). Of course, since not all of the regulatory elements have been examined in the same detail at the molecular level, this number is certain to be an underestimate; some factors must have been missed. Furthermore, the elements shown were in many cases isolated and defined as 'minimal' elements and, in general, we do not know where the actual boundaries of the regulatory modules are. All these caveats indicate that the real complexity is probably greater than that shown. Nonetheless, the examples in Fig. 1 turn out to be remarkably consistent: considering first the cases that represent single modular regulatory elements, the average number of diverse interactions is about 6.2 (Fig. 1C, 6; D, 8; G, 8; I, 5; J, 6; K, 4). Two modules are included in Fig. 1H, containing three and four different interactions; and also in Fig. 1E, containing four and five different interactions; and four modules are shown in Fig. 1F, containing 5, 4, 2 and 3 different interactions. The complete systems shown in Fig. 1A and B have an average of 4.5 interactions per module (for *CyIIIa*, Fig. 1A, 4 and 6; for *Endo16*, Fig. 1B, module A, 5; B, 4; DC, 5; E, 5; F, 5; G, 2). For all 22 modules included in Fig. 1, the average number of diverse interactions per module is 4.7. We may conclude that the complexity of specific interactions built into the DNA sequence of individual *cis*-regulatory modules usually falls in the range of four to eight different factors and that the average value is near five.

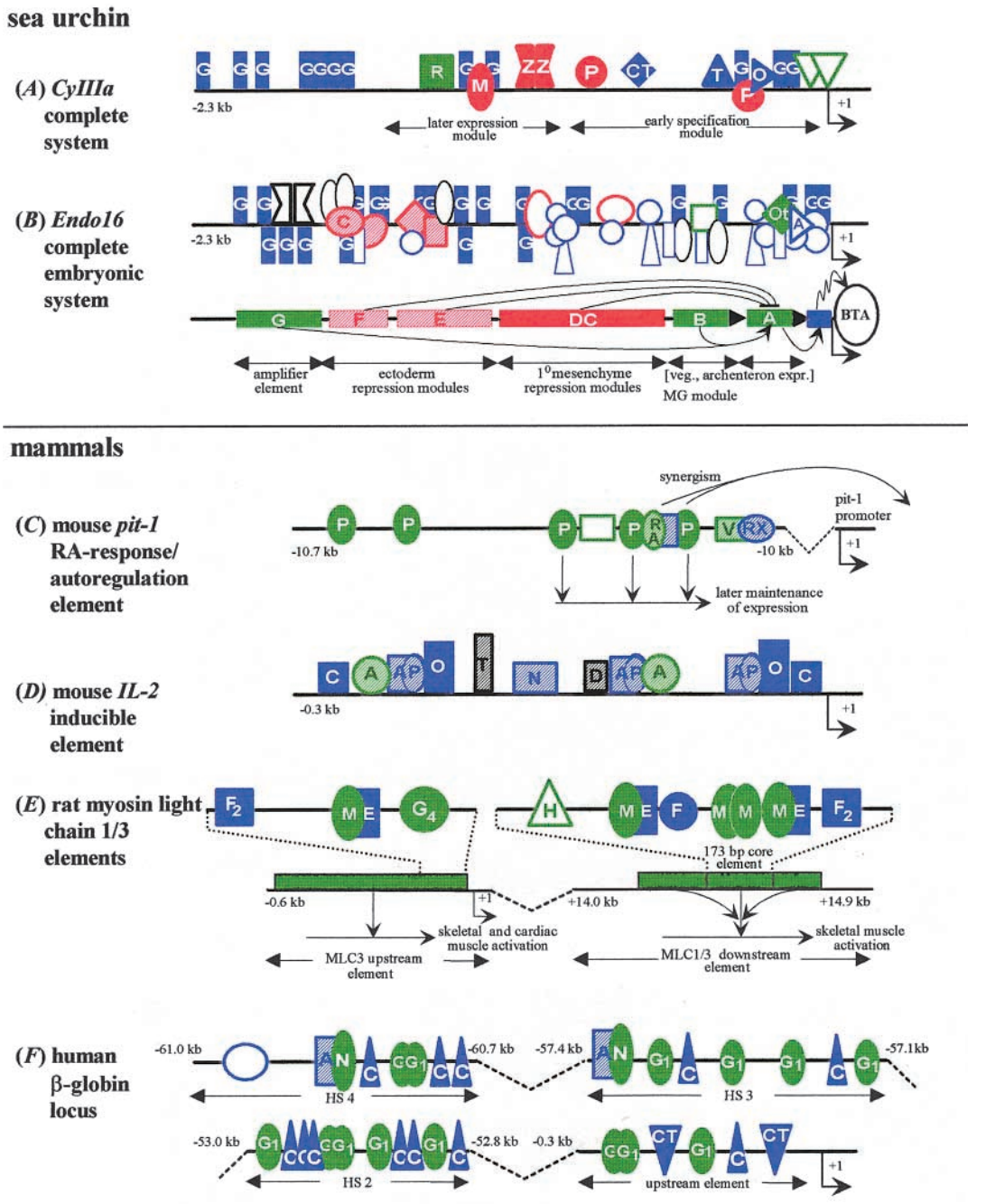
All of the examples of Fig. 1 show, furthermore, that factors of diverse chemical nature are utilized within each of the regulatory modules. There are no examples of regulatory modules serviced only by homeodomain proteins, or Zn finger proteins, and so forth. This suggests diversity in the nature of the protein:protein interactions that are required of the factors in order for each module to generate and communicate its regulatory output.

Significance of DNA-protein interactions detected in vitro

A question that must be faced is whether all the DNA-protein interactions that can be detected in vitro actually do anything. Just because a protein is bound specifically at its DNA target site, does that fact alone indicate a *cis*-regulatory function? There are various thermodynamic and probability arguments, all of which point obviously toward the functional significance of many specific DNA-protein interactions. Three such arguments are (i) that the equilibrium constants of some transcription factor-DNA complexes are such as to require the formation of multiple chemical bonds between protein side chains and the DNA bases, as has been amply confirmed by X-ray crystal structures of many such complexes; (ii) that where the affinity of a given factor for its target site is low, it is often found that highly specific cooperative interactions with other factors bound at adjacent target sites greatly increases both complex stability and site specificity and (iii) that the length of sequence protected by bound transcription factors suffices to specify these sites uniquely, even within the enormous animal genome. But convincing as they may be, these are all a priori arguments. The regulatory significance of transcription factor-DNA interactions is demonstrable experimentally in gene transfer experiments and, for the examples shown in color in Fig. 1, functions have been identified by this means for most of the factors indicated. In addition, for many of the *Drosophila* cases, e.g., the *eve* (Fig. 1H), *rhomboid* (*rho*) (I) and *knirps* (*kni*) (J) elements, there is supporting genetic evidence in the effects of mutations that eliminate the transcription factor or cause its misexpression (these comments refer to interactions by factor species, not to every one of the multiple target sites often found for each species of factors). However, *cis*-regulatory analysis cannot be done by genetics alone, but only by direct molecular manipulations, including characterization of DNA-protein interactions in vitro, combined with transfer of mutated or synthetic expression constructs into appropriate cells or embryos. Even for genes as well studied as *ftz* (Fig. 1G) direct examination revealed the participation of a number of new factors for which specific regulatory roles were demonstrated (Dearolf et al., 1989; Topol et al., 1991). Specific sites of DNA-protein interaction were mapped in vitro throughout the whole *cis*-regulatory domain of the *CyIIIa* gene of *S. purpuratus*, and here the issue of functional importance was explicitly examined for all the nine different species of interaction that were detected. All of these interactions (see Fig. 1A) have the property that the sequence preferences of the factor for its target site is at least $5\text{-}10\times10^3$ that for non-specific DNA polymers in vitro (Calzone et al., 1988, 1991; Thézé et al., 1990; Coffman et al., 1996), and this criterion turns out to be sufficient. Thus interference with every single one of these interactions, by in vivo competition (Franks et al., 1990; Hough-Evans et al., 1990) or by deletion and mutation, produces a distinct regulatory phenotype when introduced into embryos (Coffman et al., 1996; Kirchhamer and Davidson, 1996). So we can conclude that, at least at this level of sequence specificity, the target site code built into the *cis*-regulatory DNA is indeed functionally meaningful. Or, looking at the matter from the other side, if a sufficiently specific interaction is detected by physical methods, the observer had best assume that it means something, and try to find out what that something is.

Positive and negative functions of factors binding within *cis*-regulatory modules

Each of the modular elements portrayed in Fig. 1 consists of target sites for diverse factors, which execute diverse functions. Some serve as termini of signal transduction systems, a particularly prominent feature, e.g., of the *IL2* gene inducible element shown in Fig. 1D (Rothenberg and Ward, 1996). Other factors appear to control relations between the module and the BTA, or between it and other modules, a subject we take up below. Here we focus on interactions that have positive or negative developmental significance, that is, interactions that cause activation or repression of the gene in different spatial domains of the organism.



A very general rule of developmental gene regulation is that, except for autonomous domains of expression that are set up by localization of maternal factors in the very early embryo, the correct spatial boundaries of an embryonic territory are initially specified by both positive and negative *cis*-regulatory interactions. Thus in all the examples in Fig. 1 in which negative regulators are shown, deletion or mutation of their target sites causes ectopic expression across the normal borders of the transcriptional domain (see legend for references). These regulatory systems all cause expression in certain spatial territories of the embryo and shut off the genes that they control, or any reporter linked to them, in others. Examples can be found in Fig. 1 (A, B and G-K), and there are many more. Specific negative interactions determine the axial boundaries of Hox gene expression in vertebrates, for instance (e.g., see Gerard et al., 1993; Studer et al., 1994; Morrison et al., 1996). The spatial boundaries of expression of many *Drosophila* genes encoding transcription factors in addition to those in Fig. 1 have been shown to be controlled by specific negative interactions (e.g., Jäckle et al., 1992; Wimmer et al., 1995; Gray and Levine, 1996a). The *dpp* gene of *Drosophila* also includes negatively acting *cis*-regulatory elements required to generate a correctly confined spatial pattern of expression (Huang et al., 1993; Jackson and Hoffman, 1994). Negative spatial regulation has been discovered in *cis*-regulatory analyses of developmentally expressed genes in other organisms too: for instance, in sea urchins (in addition to the examples shown in Fig. 1A, B, see Niemeyer and Flytzanis, 1993; Frudakis and Wilt, 1995; Gan et al., 1990); in *C. elegans* (Egan et al., 1995) and in the snail *Patella* (Damen and Van Loon, 1996). The implication is that such negative regulators, or at least their activities, are confined to given spatial domains of the organism, and that is indeed what is generally found.

Developmental repressors work in a number of different ways, an interesting and important subject in itself that was recently reviewed by Gray and Levine (1996a). Sometimes negative regulators function as silencers, by shutting down the BTA so that it is impervious to any positive inputs, but often they work essentially by antagonizing activation functions mediated by positive regulators bound near them within the same module (‘short-range quenching’; Gray and Levine, 1996a). Short-range intramodular repression of the latter sort requires that the repressor bind within about 100 bp of the activator(s), the function of which it blocks. This has been shown for the *giant* repressor of the *eve* stripe 2 element (Gt

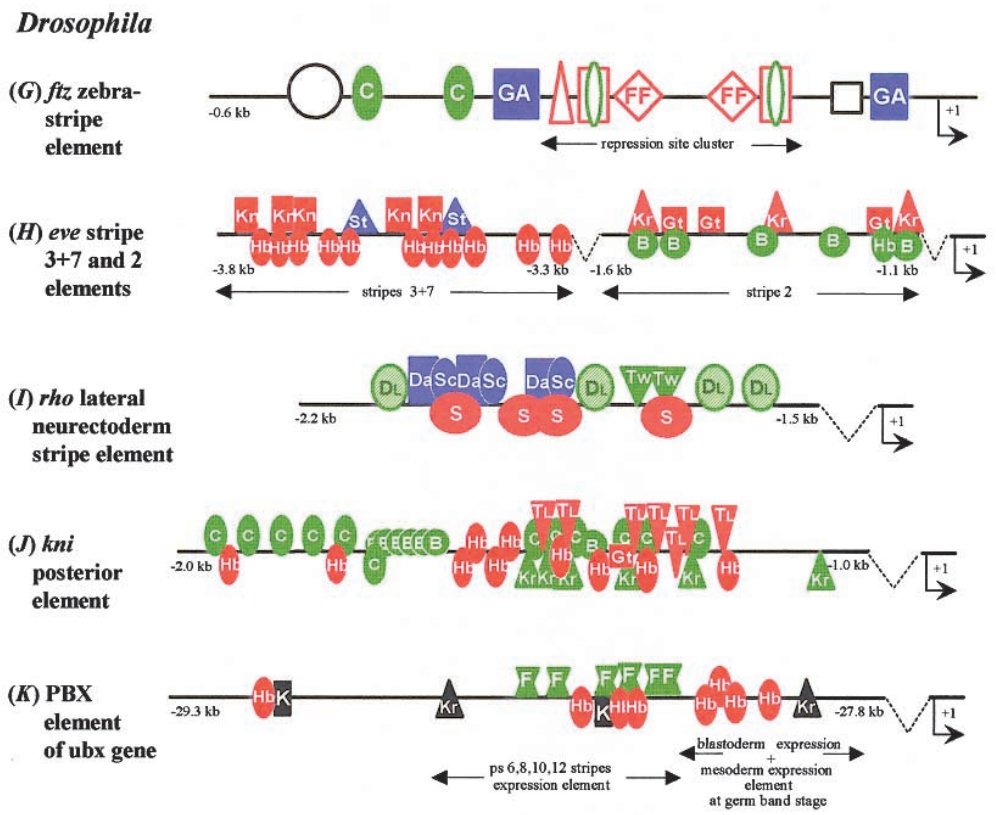


Fig. 1. Some metazoan *cis*-regulatory elements. *Cis*-regulatory elements are illustrated from various relatively well studied genes in sea urchin (A, B), mouse (C, D), rat (E), man (F) and *Drosophila* (G-K). Black horizontal lines represent the DNA. The scale and the size of each element are indicated by the positions of their termini (in kilobases, kb) with respect to the transcription start site (broken arrow). Each symbol represents an individual factor binding at its target site in the *cis*-regulatory DNA. The size of each symbol is not intended to be accurate in dimension, though symbols are larger in elements shown at higher magnification. Superimposed symbols imply only that specific sites are close together. Factors are shown for which interaction with the specific DNA element has been demonstrated by gel shift or footprint, or for which decisive genetic and sequence data provide identifications. The color coding indicates functional role where known: ubiquitous factors are in blue; factors demonstrated to work as repressors are in red; positively acting, spatially localized factors are shown in green. Factors for which none of these descriptions are applicable, or for which functions remain unknown are shown in black. Factors whose presentation or activity is known to be directly affected by signal transduction systems are shown hatched. Open, unlabeled symbols represent still unidentified factors; open but labeled symbols represent factors about which there is some functional or biochemical information which are either not cloned or are controversial. For detailed information on functional roles of factors when bound to their target sites see following references: (A) Modular *cis*-regulatory system of the *CyIIIa* cytoskeletal actin gene expressed in embryonic aboral ectoderm (Kirchhamer and Davidson, 1996 and references therein). The roles of the two individual regulatory modules are indicated below the map. (B) *Cis*-regulatory domain of the *Endo16* gene expressed in vegetal plate, archenteron and then midgut. At top is the map showing DNA-binding factors (from Yuh et al., 1994) and below is a summary of the modular functions and intermodular organization within the *Endo16 cis*-regulatory system. Here individual modules (A-G) are represented by filled or hatched boxes to which the color coding described above also applies. The function of each module is indicated beneath the horizontal double-headed lines. Arrows indicate positive interactions; terminally barred lines indicate negative interactions (from Yuh and Davidson, 1996; Yuh et al., 1996). Transcription factors portrayed in (A) and (B) are: A, AP-1-like factor; C, CREB-related protein; CT, CCAAT-binding factor; G, SpGCF1; M, SpMyb; O, SpOct-1, Ot, SpOtx; P, SpP3A2; R, SpRunt-1; T, SpTEF-1; Z, SpZ12-1. (C) Structure of the upstream element responsible for autoregulation and retinoic acid (RA) response of the *pit-1* gene (adapted from Rhodes et al., 1993). Straight and curved arrows indicate the role that specific Pit-1 (P)-binding sites have in maintenance of *pit-1* gene expression or interference with RA response. Abbreviations for other factors are: RA, retinoic acid receptor; RX, RXRα nuclear receptor coregulator; V, vitamin D receptor. (D) Stimulated T-cell configuration of the factors binding to an upstream element of the interleukin 2 (*IL-2*) gene (Rothenberg and Ward, 1996). Symbol abbreviations are: A, NF-AT; AP, AP-1; C, CACCC binding factor, Sp-1-like factor; D, CD28RC; N, NF-κB; O, Oct-1; T, TGGGC-binding factor. (E) C*is*-regulatory elements of the myosin light chain (MLC) 1/3 locus. On top are shown, at left, the *cis*-regulatory element proximal to the MLC3 gene transcription start site (McGrew et al., 1996) and, at right, the downstream element that controls both MLC1 and MLC3 gene expression (Rosenthal et al., 1992; Rao et al., 1996). Below the maps is a diagram of the genomic organization of these elements, represented as green boxes, for which the functional roles are also indicated: vertical and curved arrows indicate positive interactions, the outputs of which are represented by horizontal arrows. Factor abbreviations are: E, E12; F, MAPF-1 (muscle actin promoter factor 1); F2, MEF 2; G4, GATA-4; H, unidentified HOX protein; M, MDF (myogenic determination factor). (F) Partial organization of the *cis*-regulatory system of the β-globin locus. *Cis*-regulatory elements present in hypersensitive (HS) sites 4 (Pruzina et al., 1991), 3 (Strauss and Orkin, 1992) and 2 (Philipsen et al., 1990) are portrayed; in addition,the element proximal to the β-globin gene transcription start site is shown (deBoer et al., 1988). Abbreviations are: AN, AP-1/NF-E2 factors; C, CACC-binding factor(s); CT, CCAAT-binding factor(s); G1, GATA-1. (G-K) *Cis*-regulatory elements from five *Drosophila* genes. (G) The *fushi tarazu* (*ftz*) proximal element responsible for the 'zebra stripe' pattern of expression of *ftz* (Topol et al., 1991; and Carl S. Parker, personal communication). (H) Modular *cis*-regulatory elements responsible for stripe 2 (Small et al., 1993) and stripe 3+7 (Small et al., 1996) expression of the *even-skipped* (*eve*) gene. (I) *Cis*-regulatory element controlling the longitudinal neurectodermal pattern of expression of the *rhomboid* (*rho*) gene (Gray et al., 1994). (J) The *knirps* (*kni*) element for posterior expression (Rivera-Pomar et al., 1995). Many of the interactions shown are inferred only from the locations of target sites for regulators for which there is functional genetic evidence. (K) PBX control region of the *Ultrabithorax* (*Ubx*) gene (Zhang et al., 1991; Müller and Bienz, 1992). Where known the role of specific subelements is indicated below the map. Factor abbreviations used in *G -K* are: B, Bicoid; C, Caudal; Da, Daughterless; DL, Dorsal; F, Fushi tarazu; FF, ftz F1 steroid receptor; GA, GAGAG factor; Gt, Giant; Hb, Hunchback; K, Knirps; Kr, Kruppel; S, Snail, Sc, Scute; St, D-Stat; TL, Tailless, Tw, Twist.

in Fig. 1H; Arnosti et al., 1996) and for the *snail* repressor of the *rho* neuroectoderm lateral stripe element (S in Fig. 1I; Gray and Levine, 1996b). Short-range quenching is also found in the later expression module of the *S. purpuratus CyIIIa* gene where the repressor (M, in Fig. 1A) is required to prevent ectopic expression. However, if the target site for the activator of this module (R, in Fig. 1A) is also destroyed, the function of M is not needed, and no ectopic expression occurs, only a lower level of normal expression (Coffman et al., 1996, 1997; Kirchhamer and Davidson, 1996). Intramodular short-range quenching demonstrates the integrative function of *cis*-regulatory modules. The interplay of positive and negative functions within the module thus suffices to generate an output that can be interpreted by either the endogenous or a heterologous BTA to which it is linked, and that determines precisely bounded domains of expression in the embryo.

Programming communication within complex *cis*-regulatory systems

A requirement for interactions between *cis*-regulatory elements at more or less distant locations has been observed in many different systems. For instance, the *cis*-regulatory system of the myosin light chain (MLC) genes consists of the downstream positive regulatory elements shown in Fig. 1E, which are thought to service both the MLC1 and the linked MLC3 genes, and also individual regulatory modules upstream of each gene, of which that adjacent to the start site of the MLC3 gene is shown. Note that the proximal element contains target sites for muscle-specific factors and does not act simply as a docking site for the BTA. Hematopoietic expression of the globin genes is mediated by a distant 'locus control region,' (LCR), together with *cis*-regulatory elements in the immediate vicinity of the individual genes (see Fig. 1F). This is a particularly interesting case for the present context, due to an elegant in vivo demonstration of dynamic interaction between the LCR and the regions occupied by the genes themselves (Wijgerde et al., 1995). The dwell time for the complexes formed between these respective *cis*-regulatory modules ranges from 15-80 minutes. Complex formation requires DNA looping and transcription occurs only while the intermodular complex is extant. As Fig. 1F illustrates, both the LCR and proximal globin gene regulatory elements contain target sites for factors promoting expression of hematopoietic genes (G_1 in Fig. 1F), just as muscle factors are bound in both proximal and upstream regions of the myosin light chain *cis*-regulatory system.

Proximal *cis*-regulatory modules may perform specially important functions in processing the regulatory outputs of more distantly located modules. This has been clearly shown for the *Endo16* gene of *S. purpuratus*. Module A of this *cis*-regulatory system performs three communicative regulatory functions, in addition to mediating early vegetal plate-specific gene expression for which an Otx factor is required (Yuh and Davidson, 1996, Yuh et al., 1996; and unpublished data). The individual communicative functions are indicated by the arrows in the diagram shown in the lower portion of Fig. 1B. (i) Module A processes negative outputs from modules F, E and DC; i.e., an AP1-like protein bound in module A close to the Otx factor is required for all three upstream modules to execute their repressive functions. (ii) Module A synergistically amplifies the positive input of module B; again, a certain specific factor bound in module A (denoted by the open triangle) carries out this function. (iii) Module A transmits to the BTA the integrated output of all the modules (including its own). For this function, interactions in module A mediated by the factors shown as open circles in Fig. 1B are required. Needless to say, all these capacities of Module A are specified by the target sites encoded in its DNA. This type of organization of course increases the diversity of the control functions that can be handled by a *cis*-regulatory system, since it adds another level of integration upstream of direct interactions with the BTA.

How is the formation of looped complexes between factors bound in distantly located regulatory modules controlled? Some of the factors, or proteins bound to these factors, must have specific affinities for one another. Pursuing this line of thought a step further, there may be proteins the specific function of which is to dock in modules that are to interact, and then to recognize and bind to one another in order to facilitate or stabilize appropriate loop formation. Two candidates are already known, both initially identified as transcription factors, and both of which multimerize stably once bound to their specific DNA target sites. They are the mammalian factor Sp1 (Pascal and Tjian, 1991) and the sea urchin factor SpGCF1 (Zeller et al., 1995a). Target sites for SpGCF1 are found within many of the regulatory modules of both *CyIIIa* and *Endo16* (the factors labeled G in Fig. 1A,B) as well as in other genes (Frudakis and Wilt, 1995). Thus the possibility of specific intermodule communication could also be hardwired in the DNA, in the form of target sites for multimerizing proteins, and this type of regulatory coding could be much more widespread than is so far evident.

Cis-regulatory organization as an index of developmental role

A recurrent theme here has been the relation between internal *cis*-regulatory structure and function. Some day we shall perhaps know enough so that a map of *cis*-regulatory target sites will in itself provide a useful prediction as to the role that the regulatory system performs in development and its position in the overall regulatory network. There are already some interesting indications, several discussed earlier. For instance, the point is made above that *cis*-regulatory modules that have the function of interpreting embryonic spatial specification cues characteristically display target sites for both positively and negatively acting factors. It follows then that *cis*-regulatory elements that direct spatially confined patterns of expression but that use only positive regulators are likely to act downstream of the initial spatial specification systems. Pattern formation processes cause the presentation of transcription factors within given spatial domains and these factors may then serve as positive regulators, which suffice to activate downstream genes in particular regions, morphogenetic progenitor fields, or cell types of the embryo. Many interesting examples demonstrate the applicability of this concept, including some of those shown in Fig. 1. The *Endo16* gene of *S. purpuratus* provides one case: late in sea urchin embryo development, the gut is partitioned into foregut, midgut and hindgut, and at this point Module B alone suffices to produce correct expression, which is confined to the midgut (Yuh and Davidson, 1996). *Endo16* is now essentially governed by a one-module, positive-only regulatory system. Gut regionalization depends on signaling processes within the archenteron (McClay and

Logan, 1996). These apparently affect the presentation of the transcription factor represented in Fig. 1B by the green box in Module B, which is alone responsible for midgut expression (Yuh and Davidson, 1996; and unpublished data). At this late time, none of the negative modules required initially to specify expression in the vegetal plate are needed anymore. Another relevant *S. purpuratus* gene is that encoding the *CyIIa* cytoskeletal actin. In contrast to the *CyIIIa* system shown in Fig. 1A, the *CyIIa cis*-regulatory system is compact, apparently consisting of a single regulatory module and it lacks target sites for any negative regulators (unpublished data of the authors). However, also in contrast with *CyIIIa*, this gene turns on only following specification and initial differentiation of the cell types in which it is expressed, viz skeletogenic and secondary mesenchyme, and gut. Of course *cis*-regulatory systems that respond to already confined spatial regulators may operate by repression as well as by activation. Thus the *ftz* zebra element shown in Fig. 1G generates the complete 7-stripe pattern of expression by responding to the output of an upstream regulatory apparatus, which produces a striped distribution of repressors active in all the odd parasegments (Dearolf et al., 1989). Again, this is in contrast to the multimodule *eve cis*-regulatory system, wherein each module performs the job of integrating much broader patterns of transcription factor distribution into one or two single stripes (Harding et al., 1989; Small et al., 1992; Gray and Levine, 1996a; Fig. 1H).

A striking aspect of Fig. 1C-F is that none of the mammalian elements shown there include target sites for repressors. The explanation is that these are all cell-type-specific *cis*-regulatory elements that lie downstream of the prior genetic control functions by which the respective embryonic domains and thence the cell types arising within these domains are specified. For example, the gene encoding the POU domain Pit-1 transcription factor (Fig. 1C) is downstream of other transcription factors required for specification of pituitary cell types, including a *paired* homeodomain factor, Prop-1 (Sornson et al., 1996). A modular regulatory element controlling axial expression of the *HoxC8* gene in the mouse embryo provides a similar case (Shashikant et al., 1995; Shashikant and Ruddle, 1996). This spatial control system apparently also responds to a set of prior, positively acting developmental regulators. These may include a *caudal*-related factor, and a *forkhead* class transcription factor, as well as products of other *Hox* genes. Similarly, the genes encoding transcription factors that activate muscle-specific differentiation genes such as the MLC gene (Fig. 1E) are themselves downstream of the regulatory apparatus by which muscle precursors are initially specified, as has been elucidated in particular for *Drosophila* (see, e.g., Lilly et al., 1994; Staehling-Hampton et al., 1994; Bodmer, 1995; Carmena et al., 1995; Frasch, 1995; Maggert et al., 1995; Taylor et al., 1995; Ranganayakulu et al., 1996). There are of course many other examples of terminal differentiation genes that do include negative spatial control elements; for instance, in a number of neuron-specific genes (Schoenherr and Anderson, 1995a). As in the cases of the *S. purpuratus CyIIIa* and *Endo16* genes, these are examples in which some aspects of spatial control depend on *cis*-regulatory interactions at the level of the terminal differentiation gene.

One other developmental circumstance that probably indicates entirely positive, single-module *cis*-regulatory systems deserves mention. There is a large group of invertebrate embryos that utilize essentially similar processes to accomplish specification in early development, of which the best known examples are ascidian, sea urchin and nematode embryos. These embryos typically generate more or less invariant cell lineages, some founder cells of which are specified by interblastomere signaling during cleavage ('Type 1' embryos; Davidson, 1990, 1991, 1994; Davidson et al., 1995). However, Type 1 embryos typically generate some autonomously specified cell lineages as well. Autonomously specified founder cells apparently inherit spatially confined maternal factors that are localized in the cytoarchitecture of the cleavage-stage embryo. *Cis*-regulatory elements for three *S. purpuratus* genes that can reasonably be regarded as markers of early autonomous specification have been examined in some detail. One of these is *SM50* (Makabe et al., 1995), a gene encoding a skeletogenic matrix protein that is expressed very early in the autonomously specified lineage from which derives the skeletogenic mesenchyme. The other two, SpAn and SpHE (Wei et al., 1995; Kozlowski et al., 1996; SpHE is the hatching enzyme) encode metalloendoproteases that are expressed in animal half and equatorial cells, but not in the vegetal-most cells of the early blastula. All three of these genes display compact, single-module *cis*-regulatory systems and, in none of them, are target sites for negative regulators to be found. This makes sense: like those *cis*-regulatory elements that lie downstream of prior specification systems, these also have only to respond to already localized regulators.

GENE REGULATORY NETWORKS

The components that constitute gene regulatory networks consist of the linkages between different *cis*-regulatory systems together with the genes that they govern. The most concrete examples at present are batteries of genes encoding cell-type-specific differentiation proteins. Here the regulatory linkages are between the genes encoding certain transcription factors and individual genes encoding differentiation proteins, the *cis*-regulatory systems of which share target sites for these factors.

Gene batteries

In his original use of this term, Morgan (1934) thought of gene batteries as sets of genes expressed at different stages of development. Britten and Davidson (1969, 1971) transformed this into a molecular concept. Gene batteries were conceived as sets of genes that are coordinately expressed because their *cis*-regulatory sequences share homologous target sites for activators, and today's definition descends from this. However, the parallel input logic of *cis*-regulatory systems, as illustrated in Fig. 1, means that we have to deal with the problem of multiple regulatory inputs if we are to keep the original sense of a set of genes encoding proteins that together perform certain functions. Furthermore, even if the patterns of expression are cell-type-specific, many of the transcription factors for which the genes of a battery share sites are not. Reality differs from a priori conception in that cell type specificity is often not simply defined (particularly in mammalian lineages): it may in different cases depend on combinatorial interactions amongst factors; on cell-type-specific proteins that are bound to transcription factors anchored on the DNA; on signal-dependent modification of factors; on earlier developmental events than

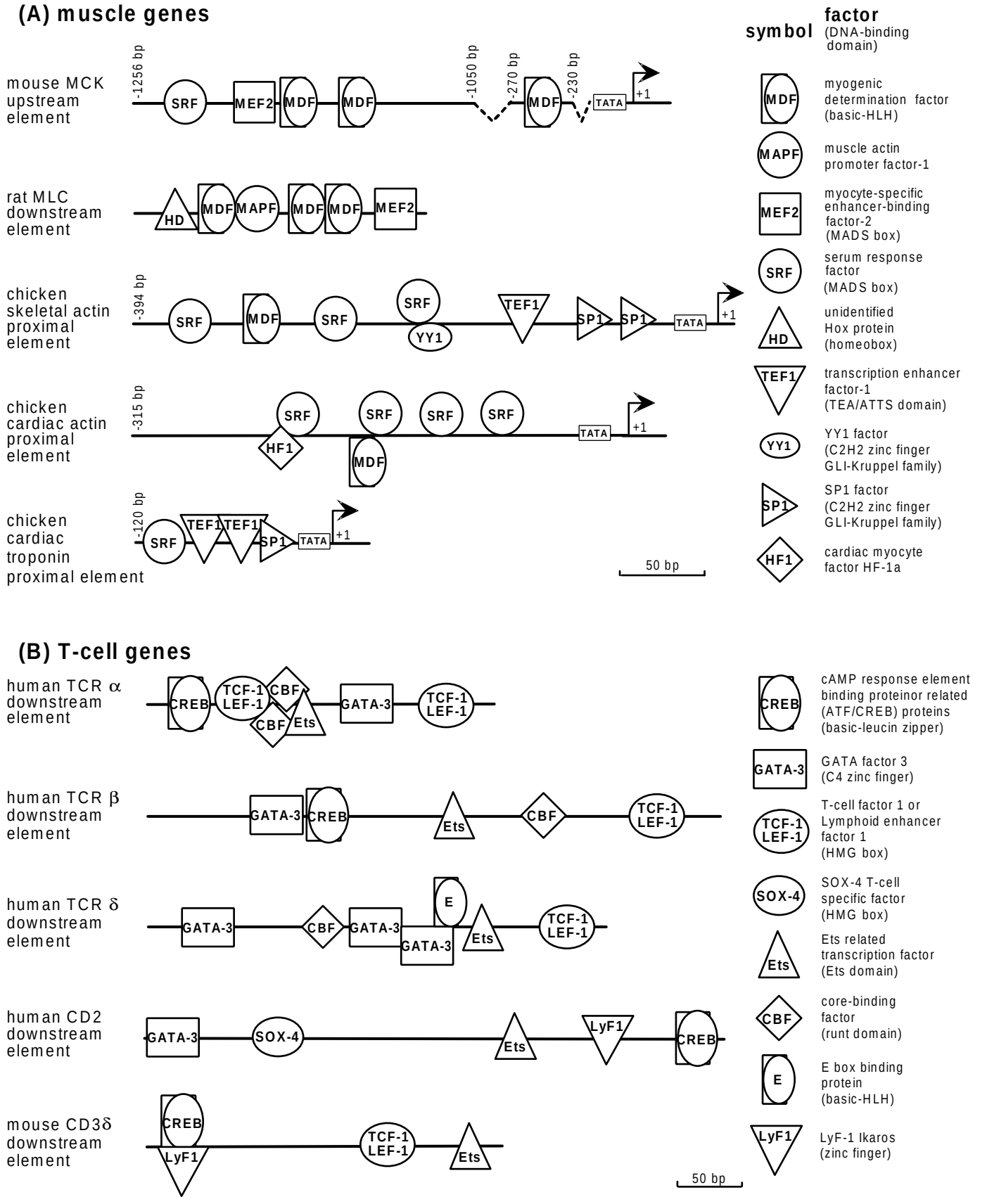

(A) muscle genes
symbol
factor (DNA-binding domain)
mouse MCK upstream element
-1256 bp
SRF
MEF2
MDF
MDF
-1050 bp
-270 bp
MDF
-230 bp
TATA
+1
rat MLC downstream element
HD
MDF
MAPF
MDF
MDF
MEF2
chicken skeletal actin proximal element
-394 bp
SRF
MDF
SRF
SRF
YY1
TEF1
SP1
SP1
TATA
+1
chicken cardiac actin proximal element
-315 bp
SRF
HF1
SRF
MDF
SRF
SRF
TATA
+1
chicken cardiac troponin proximal element
-120 bp
SRF
TEF1
TEF1
SP1
TATA
+1
50 bp
MDF
myogenic determination factor (basic-HLH)
MAPF
muscle actin promoter factor-1
MEF2
myocyte-specific enhancer-binding factor-2 (MADS box)
SRF
serum response factor (MADS box)
HD
unidentified Hox protein (homeobox)
TEF1
transcription enhancer factor-1 (TEA/ATTS domain)
YY1
YY1 factor (C2H2 zinc finger GLI-Kruppel family)
SP1
SP1 factor (C2H2 zinc finger GLI-Kruppel family)
HF1
cardiac myocyte factor HF-1a
(B) T-cell genes
human TCR α downstream element
CREB
TCF-1 LEF-1
CBF
CBF
Ets
GATA-3
TCF-1 LEF-1
human TCR β downstream element
GATA-3
CREB
Ets
CBF
TCF-1 LEF-1
human TCR δ downstream element
GATA-3
CBF
GATA-3
E
GATA-3
Ets
TCF-1 LEF-1
human CD2 downstream element
GATA-3
SOX-4
Ets
LyF1
CREB
mouse CD3δ downstream element
CREB
LyF1
TCF-1 LEF-1
Ets
50 bp
CREB
cAMP response element binding proteinor related (ATF/CREB) proteins (basic-leucin zipper)
GATA-3
GATA factor 3 (C4 zinc finger)
TCF-1 LEF-1
T-cell factor 1 or Lymphoid enhancer factor 1 (HMG box)
SOX-4
SOX-4 T-cell specific factor (HMG box)
Ets
Ets related transcription factor (Ets domain)
CBF
core-binding factor (runt domain)
E
E box binding protein (basic-HLH)
LyF1
LyF-1 Ikaros (zinc finger)

Fig. 2. Two examples of gene batteries. (A) *Cis*-regulatory elements controlling the expression of some striated muscle genes: muscle creatine kinase (MCK, Donoviel et al., 1996), myosin light chain (MLC, Rosenthal et al., 1992; Rao et al., 1996), skeletal α-actin (MacLellan et al., 1994), cardiac α-actin (Moss et al., 1994) and cardiac troponin T (Mar and Ordahl, 1988; Larkin et al., 1996). (B) *Cis*-regulatory elements controlling the expression of some T-cell specific genes: T-cell receptors (TCR) α, β and γ (Leiden, 1993; Giese et al., 1995); CD2 (Wotton et al., 1995) and CD3γ (Georgopoulos et al., 1990, 1992; Molnár and Georgopoulos, 1994). Symbol, name, abbreviation and nature of the DNA-binding domain of the transcription factors portrayed in A and B are indicated on the right side of each gene battery. The position of the TATA box and the position of transcription start site (broken arrow), when present in the element reported, are also indicated. The scales used for the genes grouped in A and B are indicated by bars at the bottom right of each panel, respectively (bp, base pairs).

those mediated by the *cis*-regulatory modules controlling expression at a particular time of observation; and so forth.

Two examples of gene batteries are shown in Fig. 2. Fig. 2A portrays *cis*-regulatory elements controlling expression of some contractile muscle proteins, and Fig. 2B illustrates *cis*-regulatory elements of some T-cell specific genes. The individual *cis*-regulatory elements of each of these batteries share most but often not all of the set of transcriptional regulators that could be used to define each battery. Furthermore, the order and spacing of the target sites for shared factors are in no two cases exactly alike. This largely reflects lack of functional constraint with respect to spacing and order, but it may also be that some genes are expressed in subtly different ways from others, i.e., at different rates and times in differentiation or under different forms of external inducement. But the fact remains that the *cis*-regulatory elements of each battery share linkages to more or less the same set of transcriptional regulators. There are not so many examples of *cis*-regulatory gene batteries defined so extensively as those in Fig. 2. However, there are many cases where target sites for at least one transcription factor have been found in the *cis*-regulatory domains of sets of differentiation genes which belong to some functional or developmental cohort, and we can be sure that in each case additional factors will be discovered to constitute a shared set of regulators. For mammals, examples of factors that regulate genes that are in some sense coordinately expressed include HNF-4 in liver genes (Sladek et al., 1990), C/EBP family members in genes required in fat cells (Christy et al., 1991; Yeh et al., 1995), Pit-1 in genes encoding various pituitary products (Andersen and Rosenfeld, 1994) and NRSF/REST, a negative regulator utilized in neuron-specific genes to repress expression in other cells (Schoenherr and Anderson, 1995b). Examples are also known in sea urchins for genes expressed in skeletogenic cells, gut cells and aboral ectoderm cells (Nemer et al., 1995; Kirchhamer et al., 1996a; Y.-H. Lee, C.-H. Yuh, M. Arnone and E. Davidson,

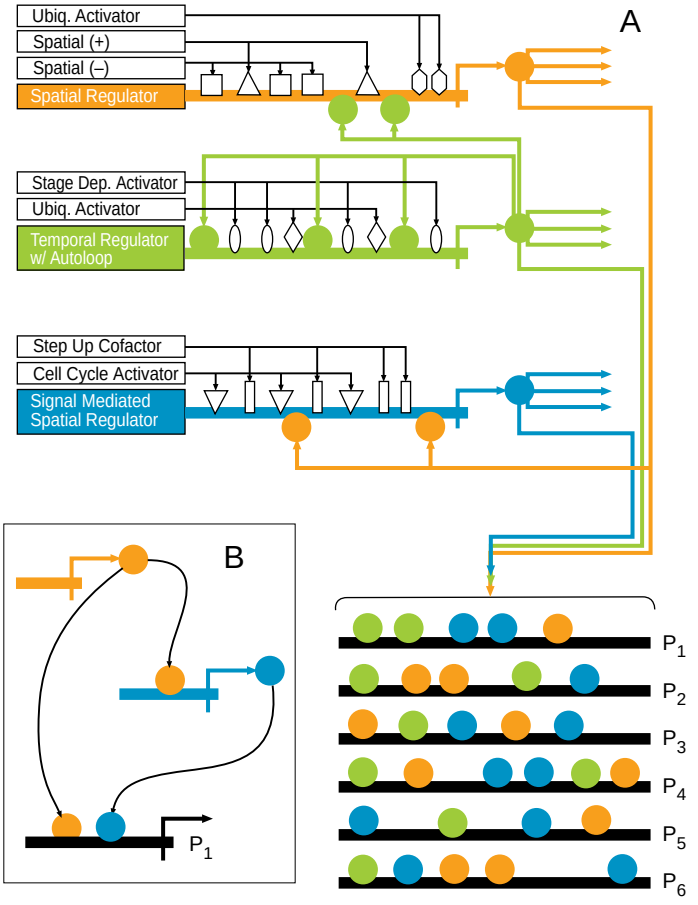


Fig. 3. A sector of an imaginary developmental gene regulatory network. (A) Network sector. Three genes encoding transcription factors are shown at the top. These are a spatial regulator (orange), a temporal regulator (green) and a signal-mediated regulator (blue). Genes encoding other transcription factors that originate off the diagram are indicated in black type on open backgrounds. A battery of six genes encoding some differentiation proteins (P1-P6) is shown below. Connections between the three genes encoding transcription factors and target sites in the P1-P6 genes are indicated by respectively colored bent arrows and the transcription factors as solid circles. The spatial regulatory gene is controlled by positive and negative interactions, which establish the limited spatial domain where it will be expressed, and it utilizes a ubiquitous ancillary activator to achieve an appropriate level of expression. This gene would be expressed only at certain stages due to requirement for the factor produced by the green temporal regulator, shown below the line binding to its target sites in the *cis*-regulatory DNA. The *cis*-regulatory system of the temporal regulator responds to its own transcription factor, and also depends on a factor appearing only after a certain stage of development, and on another ubiquitous ancillary activator. The signal-mediated regulator produces a factor that is activated by signals. For example, if this were a short-range signal produced by cells adjacent to the domain of expression of the spatial orange regulator, P1-P6 would be expressed only near the boundary. The *cis*-regulatory system controlling expression of the signal-mediated transcription factor includes target sites for the product of the orange spatial regulator, shown binding below the line representing the DNA, and also for two factors that work together to promote transcription during growth, one imagined as a regulator produced when cells are cycling, the other as a ubiquitous co-factor. The arrows at the right indicate that each of the three genes encoding transcription factors have many downstream targets besides the P1-P6 gene battery. Any resemblance between this network sector and a known regulatory element is purely coincidental. (B) A single relationship extracted from the network. A causal diagram is shown portraying the multilevel function of the orange spatial regulator, which controls both the gene encoding the blue signal-mediated regulator and the P1 gene; the latter, however, is also directly responsive to the spatial regulator.

unpublished data). The gene battery concept is relevant to sets of genes encoding transcription factors as well as to terminal differentiation genes. For example, the transcription factors encoded by the proneural *achaete* and *scute* genes of *Drosophila* directly regulate the expression of the genes of the *Enhancer of Split* complex, which encodes a set of transcription factors required for local spatial specification of peripheral nervous system elements (Singson et al., 1994).

Battery relations among *cis*-regulatory elements can only be elucidated by direct analysis of structure and function within these elements. Ultimately, this requires both gene transfer and the ability to clone out all the DNA-binding factors. At present, the only general approach is partial purification of the factor by means of affinity chromatography and protein microsequencing. In the context of embryogenesis, there are so far only two systems where this is possible, viz *Drosophila* (e.g., Biggin and Tjian, 1988; Perkins et al., 1988; Huang et al., 1995; Liaw et al., 1995) and sea urchin (Calzone et al., 1991; Coffman et al., 1992, 1996; other references in Kirchhamer and Davidson, 1996). Yet the isolation of developmentally active gene batteries that encode cohorts of differentiation proteins is of obvious importance. *Cis*-regulatory analysis of such sets of genes provides the most direct pathway upstream from a morphogenetically intelligible starting position, i.e., proteins of known function, into the interior of the gene regulatory network.

General considerations

(i) Peripheral and internal network elements

In considering linkages within the genomic regulatory network, one immediately realizes that there are two kinds of genes: those for which all linkages to specific *cis*-regulatory target sites are upstream and those for which there are both upstream and downstream linkages. The latter category consists exclusively of genes encoding transcription factors. Here 'upstream' is used in a causal sense to denote the genes encoding transcription factors that bind in a given *cis*-regulatory system, and hence 'downstream,' with respect to genes encoding transcription factors, denotes the *cis*-regulatory system(s) to which these factors bind. The linkages in which genes encoding transcription factors participate, and these genes, constitute the internal components of the regulatory network; genes encoding all other kinds of proteins constitute the peripheral or terminal elements. Of course many other kinds of causal relationship can be demonstrated, for instance, relationships that display signal systems causally upstream of genes encoding transcription factors. These involve chains of protein-to-protein interaction, however, and thus they describe regulatory connections beyond those immediately represented in the genomic DNA sequence.

(ii) Networks and causal relationships

Fig. 3 describes an entirely imaginary sector of a genetic regulatory network. The internal part of this network sector is shown at the top. It consists of three genes encoding transcription factors. The first, shown in orange, is a spatial regulator, controlled by the usual mix of positively and negatively acting transcription factors. The second, shown in green, is a stage-dependent temporal regulator with an autoregulatory loop. The third, shown in blue, encodes a signal-dependent transcription factor, which is wired up to the first, so that it is only expressed in the spatial domains where the orange regulator is presented. There are other inputs into the three internal genes, as shown. The peripheral genes are at the bottom, constituting a battery encoding six proteins. They will be expressed at certain places, according to the orange regulator; but only in certain subregions where the signal system activating the blue regulator is functional, and when cells are dividing; and only at certain stages of development, according to the green regulator. The point of this imaginary diagram is purely heuristic: it illustrates interesting principles. One such can be seen by comparing Fig. 3B to Fig. 3A. In Fig. 3B one of the causal relationships in Fig. 3A is abstracted. Experimental analysis has revealed a number of relationships of this kind, of which two interesting examples from the work of Levine and colleagues are reproduced in Fig. 4A. But such relationships should not be confused with regulatory networks as they must really exist, given what we already know of *cis*-regulatory organization. Isolated relationships trace simple lines of causality, while in life there are multiple parallel inputs into any *cis*-regulatory system, with at least some degree of informational integration to be expected to occur within each *cis*-regulatory module, whether it is located peripherally or internally.

(iii) Multilevel connections

Fig. 3A illustrates several kinds of connections: downstream connections from the internal genes to the peripheral genes, an autoregulatory connection and two connections amongst the internal genes. Autoregulatory connections are known in many genes that encode transcription factors. Fig. 1C in fact displays the internal anatomy of the *Pit-1* autoregulatory element and, for example, there are autoregulatory modules (separate from those portrayed in Fig. 1) in the *cis*-regulatory systems of the *ftz* gene (Hiromi et al., 1985; Dearolf et al., 1989) and the *eve* gene (Harding et al., 1989). Autoregulatory elements are also found in Hox genes (e.g., Pöpperl et al., 1995), and often occur

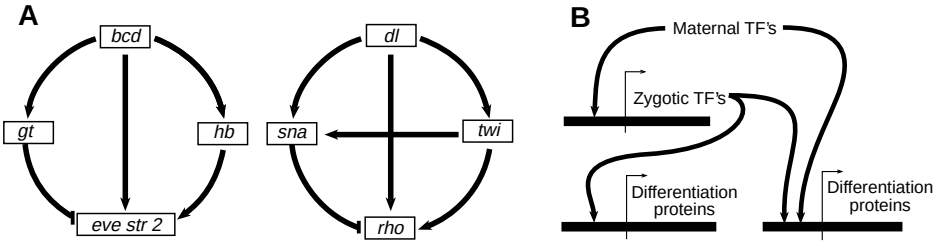


Fig. 4. Relationships between maternal transcription factors and zygotically expressed genes. (A) Upstream of the *eve* and the *rhomboid* genes of *Drosophila*, from Ip et al. (1992a); a 'view from the genome' (see text). The maternal factors are, respectively, *bcd*, which activates both the *hb* activator and the *gt* repressor, as well as the downstream *eve* stripe 2 module and *dl*, which activates both the *twi* activator and the *sna* repressor, as well as the downstream neuroectoderm module of the *rho* gene. An additional relationship is that *twi* also activates the gene encoding the *sna* repressor (Ip et al., 1992b). (B) Relationship diagram for early peripheral gene expression in sea urchin embryos. Arguments are discussed in text.

in genes encoding transcription factors of all classes, examples of which are too numerous to cite here. Regulatory linkages between genes encoding transcription factors are now a common theme. A famous example is the web of 'cross-regulatory' interactions amongst *HOM-C*/*Hox* genes (McGinnis and Krumlauf, 1992). But, as Fig. 3A illustrates, and the real relationships in Fig. 4A demonstrate, linkages between internal genes of the network may connect a given gene to others at several different levels or positions in the network. Therefore the network cannot be conceived in any simple sense as strictly hierarchical. That is, it cannot be composed simply of several horizontal 'layers.'

(iv) The 'view from the genome' and the 'view from the nucleus'

Figs 1, 2 and 3 of this paper portray what one might call 'the view from the genome.' They attempt to describe the organization of *cis*-regulatory systems in terms of all the interactions that occur under all circumstances, because that is what is determined by the information hardwired into the regulatory DNA. However, in life whether these interactions occur of course depends on which nucleus one considers, and at what stage of development. The 'view from the nucleus' describes a different picture for each developmental state in which the gene acts differently. For example, in vivo footprinting shows that the *IL-2* gene regulatory module portrayed in Fig. 1D is either occupied as shown, partially loaded or entirely devoid of the transcription factors indicated, depending on the developmental and stimulatory state of the T cells (Rothenberg and Ward, 1996). The same is true of other developmentally regulated genes studied in this way (e.g., Jackson et al., 1989; Mueller and Wold, 1989). Similarly, in a cell facing another across a boundary, a *cis*-regulatory element might have a negative regulator bound to it, while the view from the opposing nucleus might reveal only positive regulators in the same *cis*-regulatory element. The view from the genome is the sum of the views from all the relevant nuclei, integrated over time. Development is the process that partitions the total of potential interactions specified in the DNA sequence into the various states in which each *cis*-regulatory system exists in these nuclei at specific stages. Another way of saying this is that per-genome networks have no temporal dependence. Developmental per-nucleus networks may depend on time, for example progressive changes in chromatin structure in given lineages, as well as on the identity of the nucleus in the organism.

Complexity of developmental gene regulation

It is at present impossible to define fully the complexity of any sector of the genomic regulatory network for development. Here 'complexity' is interpreted in an elemental way, as the number of linkages upstream from the *cis*-regulatory system controlling a given phase of the expression of a developmentally regulated peripheral gene. We do not yet know enough about any system to construct a true network diagram that would cover even a small sector such as we concocted for Fig. 3A. At our present state of knowledge, the greatest complexity would seem likely to be required in the process of morphogenetic pattern formation. As discussed elsewhere (Davidson, 1990, 1991, 1993, 1994; Davidson et al., 1995) the process by which major body parts arise in postembryonic development of all bilaterian metazoans begins with regional expression of transcription factors in the anlage or progenitor field of the future structure, or of various transcription factors in the various portions of the field. After early embryonic stages, this always results in expression of spatially confined signals. These signals affect the activity of regulators controlling genes that encode new transcription factors essential to generation of the pattern. The ultimate integration of the spatial information conveyed by the intercellular signals thus occurs in the *cis*-regulatory modules directing the expression of these transcription factors, which are thereby expressed in newly defined spatial subdivisions of the embryo. These events occur in a context of growth and cellular expansion. Cell differentiation programs are called into play only later in the process, but how many linkages downstream is hard to guess. Currently the best known examples are in appendicular development, i.e., imaginal discs in *Drosophila* and limb buds in amniotes (for some references see the reviews cited above). Expression of *Hox*/*HOM-C* genes in axial specification during body plan formation is clearly an early step in a process that operates by a similar mechanism. The same can be said for genes such as *eve* and *ftz* (Fig. 1), which set up transient metameric patterns of transcription factor expression that are necessary for, and that foreshadow, parasegmental morphogenesis. But again, the number of linkages that separate these initial stages of the pattern formation process from parasegmental morphogenesis is obscure, except that it is very unlikely to be a small number.

One circumstance where there are at least some indications is at the very outset of development. The transcription factors that activate the initial cohorts of zygotic genes expressed in an embryo are all maternal. The two *Drosophila* examples in Fig. 4A are interesting in that they show that the maternal *bcd* and *dl* factors are each immediately utilized to activate genes encoding both positively and negatively acting transcriptional regulators, but also other genes at a level further removed in the network. Furthermore, the same kinds of relationships obtain in the two cases shown in Fig. 4A, though *eve* is an internal and *rho* a peripheral gene in the network. Over 30 genes encoding transcription factors are expressed in spatial patterns soon after cellularization in *Drosophila* (see reviews of Harding and Levine, 1989; Jäckle, 1992). Given an average of five different factors per *cis*-regulatory module, one can count on the existence of hundreds of network linkages even at this very early stage, without even considering the terra incognita that separates this stage from expression of the peripheral genes that effect morphogenesis.

In type 1 embryos, the network sectors that relate maternal transcription factors to at least the initial expression of peripheral genes are likely to be far more shallow. These embryos characteristically begin to express peripheral genes encoding differentiation proteins early in embryonic development, sometimes by the end of cleavage (Davidson, 1990, 1991). The *S. purpuratus CyIIIa* and *Endo16* genes considered in Fig. 1 are examples (Shott et al., 1984; Lee et al., 1986; Godin et al., 1996). Strikingly, virtually all the factors that regulate the *CyIIIa* gene are present as maternal activities, though those that have been studied are soon transcribed zygotically as well (Calzone et al., 1997; Zeller et al., 1995b; Wang et al., 1995; Coffman et al., 1996; see also Gan et al., 1995). This leads to the educated guess shown in Fig. 4B, a relationship diagram the import of which is that the initial cohort of spatially

confined peripheral gene expressions does not require the participation of many levels of genes in the internal portion of the network. The complex *cis*-regulatory systems of genes such as *Endo16* and *CyIIIa*, and their extensive spatial information processing capabilities, may follow from the shallow nature of the regulatory network sectors required for early development in these embryos. We cannot resist adding that, given the accessibility of such embryos to extensive *cis*-regulatory analysis, the relative simplicity of this portion of the network greatly enhances their attractiveness for regulatory analysis: complete network sectors for peripheral gene expression in the early embryo may here actually be within reach.

CONCLUSION

The most general conclusion that one might draw from these arguments concerns the central importance of *cis*-regulatory analysis. Therein lies the experimental path to understanding the organization of the genomic program for development. Isolated causal relationships between individual internal genes of the network have been enormously revealing over the last decade, and will continue to be so, as the identity of important internal genes and the evolutionary conservation of their linkages is established. But we think the time has come for workers in our field to bite the bullet and approach directly the organization of developmental regulatory networks in the genome. This can only be done by direct analysis of *cis*-regulatory systems situated at all levels of the network. Knowledge of the structure of these networks will lead to solutions of fundamental mysteries of both development and evolution.

It is safe to predict that in this task primary genomic sequence data will become an invaluable source of information. For the regulatory developmental biologist this indeed may be the most important ultimate outcome of the genome project.

The authors are grateful to reviewers of drafts of this manuscript, for their extremely helpful criticisms and comments, and above all, their perspicacity. These reviewers were Professor Barbara Wold, Professor Ellen Rothenberg and Dr R. Andrew Cameron of this Institute; Professor Lee Hood of the University of Washington; Professor Nadia Rosenthal of Harvard Medical School; Professor Eric Olson of the University of Texas Southwestern Medical Center; Dr Carmen Kirchhamer of Harvard University and Professor James W. Posakony University of California, San Diego. This work was supported by the Stowers Institute for Medical Research and by the National Institute of Child Health and Human Development.

REFERENCES

Andersen, B. and Rosenfeld, M. G. (1994). Pit-1 determines cell types during development of the anterior pituitary gland. *J. Biol. Chem*. **269**, 29335-29338.

Arnosti, D. N., Barolo, S., Levine, M. and Small, S. (1996). The *eve* stripe 2 enhancer employs multiple modes of transcriptional synergy. *Development* **122**, 205-214.

Baumgartner, S. and Noll, M. (1991). Network of interactions among pair-rule genes regulating *paired* expression during primordial segmentation of *Drosophila*. *Mech. Dev*. **33**, 1-18.

Bender, W., Akam, M., Karch, F., Beachy, P. A., Peifer, M., Spierer, P., Lewis, E. B. and Hogness, D. S. (1983). Molecular genetics of the bithorax complex in *Drosophila melanogaster*. *Science* **221**, 23-29.

Biggin, M. D. and Tjian, R. (1988). Transcription factors that activate the *Ultrabithorax* promoter in developmentally staged extracts. *Cell* **53**, 699-711.

Bodmer, R. (1995). Heart development in *Drosophila* and its relationship to vertebrates. *Trends Cardiovasc. Med.* **5**, 21-28.

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

Britten, R. J. and Davidson, E. H. (1971). Repetitive and non-repetitive DNA sequences and a speculation on the origins of evolutionary novelty. *Quart. Rev. Biol*. **46**, 111-138.

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

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

Calzone, F. C., Grainger, J., Coffman, J. A. and Davidson, E. H. (1997) Extensive maternal representation of DNA-binding proteins that interact with regulatory target sites of the *Strongylocentrotus purpuratus CyIIIa* gene**.** *Mol. Marine Biol. Biotech.*, in press.

Carmena, A., Bate, M. and Jiménez, F. (1995). *lethal of scute*, a proneural gene, participates in the specification of muscle progenitors during *Drosophila* embryogenesis. *Genes Dev*. **9**, 2373-2383.

Christy, R. J., Kaestner, K. H., Geiman, D. E. and Lane, M. D. (1991). CCAAT/enhancer binding protein gene promoter: binding of nuclear factors during differentiation of 3T3-L1 preadipocytes. *Proc. Natl. Acad. Sci. USA* **88**, 2593-2597.

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

Coffman, J. A., Kirchhamer, C. V., Harrington, M. G. and Davidson, E. H. (1996). SpRunt-1, a new member of the Runt-domain family of transcription factors, is a positive regulator of the aboral ectoderm-specific *CyIIIa* gene in sea urchin embryos. *Dev. Biol*. **174**, 43-54.

Coffman, J. A., Kirchhamer, C. V., Harrington, M. G. and Davidson, D. H. (1997). A Myb target site in the regulatory domain of the *Strongylocentrotus purpuratus CyIIIa* gene is required for spatial restriction of SpRunt-1 mediated transcriptional activation. (In press)

Damen, W. G. M. and van Loon, A. E. (1996). Multiple *cis*-acting elements act cooperatively in directing trochoblast-specific expression of the α-tubulin-4 gene in *Patella* embryos. *Dev. Biol*. **176**, 313-324.

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

Davidson, E. H. (1991). Spatial mechanisms of gene regulation in metazoan embryos. *Development* **113**, 1-26.

Davidson, E. H. (1993). Later embryogenesis: regulatory circuitry in morphogenetic fields. *Development* **118**, 665-690.

Davidson, E. H. (1994). Molecular biology of embryonic development: How far have we come in the last ten years? *BioEssays* **16**, 603-615.

Davidson, E. H., Peterson, K. J. and Cameron, R. A. (1995). Origin of bilaterian body plans: Evolution of developmental regulatory mechanisms. *Science* **270**, 1319-1325.

Dearolf, C. R., Topol, J. and Parker, C. S. (1989). Transcriptional control of *Drosophila fushi tarazu* zebra stripe expression. *Genes Dev*. **3**, 384-398.

deBoer, E., Antoniou, M., Mignotte, V., Wall, L. and Grosveld, F. (1988). The human β-globin promoter; nucler protein factors and erythroid specific induction of transcription. *EMBO J*. **7**, 4203-4212.

Donoviel, D. B., Shield, M. A., Buskin, J. N., Haugen, H. S., Clegg, C. H. and Hauschka, S. D. (1996). Analysis of muscle creatine kinase gene regulatory elements in skeletal and cardiac muscles of transgenic mice. *Mol. Cell. Biol.* **16**, 1649-1658.

Egan, C. R., Chung, M. A., Allen, F. L., Heschl, M. F. P., Van Buskirk, C. L. and McGhee, J. D. (1995). A gut-to-pharynx/tail switch in embryonic expression of the *Caenorhabditis elegans ges-1* gene centers on two GATA sequences. *Dev. Biol*. **170**, 397-419.

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

Frasch, M. (1995). Induction of visceral and cardiac mesoderm by ectodermal Dpp in the early *Drosophila* embryo. *Nature* **374**, 464-467.

Frudakis, T. N. and Wilt, F. (1995). Two cis elements collaborate to spatially repress transcription from a sea urchin promoter. *Dev. Biol.* **172**, 230-241.

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

Georgopoulos, K., Galson, D. and Terhorst, C. (1990). Tissue-specific nuclear factors mediate expression of the CD3δ gene during T-cell development. *EMBO J*. **9**, 109-115.

Georgopoulos, K., Moore, D. D. and Derfler, B. (1992). Ikaros, an early lymphoid-specific transcription factor and a putative mediator for T-cell commitment. *Science* **258**, 808-812

Gérard, M., Duboule, D. and Zákány, J. (1993). Structure and activity of regulatory elements involved in the activation of the *Hoxd-11* gene during late gastrulation. *EMBO J*. **12**, 3539-3550.

Giese, K., Kingsley, C., Kirshner, R. and Grosschedl, R. (1995). Assembly and function of a TCRα enhancer complex is dependent on LEF-1-induced DNA bending and multiple protein-protein interactions. *Genes Dev.* **9**, 995-1008.

Godin, R. E., Urry, L. A. and Ernst, S. G. (1996). Alternative splicing of the *Endo16* transcript produces differentially expressed mRNAs during sea urchin gastrulation. *Dev. Biol*. **179**, 148-159.

Gómez-Skarmeta, J. L., Rodríguez, I., Martínez, C., Culí, J., Ferrés-Marcó, D., Beamonte, D. and Modolell, J. (1995). *Cis*-regulation of *achaete* and *scute*: shared enhancer-like elements drive their coexpression in proneural clusters of the imaginal discs. *Genes Dev*. **9**, 1869-1882.

Gray, S. and Levine, M. (1996a). Transcriptional repression in development. *Curr. Opin. Cell Biol.* **8**, 358-364.

Gray, S. and Levine, M. (1996b). Short-range transcriptional repressors mediate both quenching and direct repression within complex loci in *Drosophila*. *Genes Dev.* **10**, 700-710.

Gray, S., Szymanski, P. and Levine, M. (1994). Short-range repression permits multiple enhancers to function autonomously within a complex promoter. *Genes Dev*. **8**, 1829-1838.

Gutjahr, T., Vanario-Alonso, C. E., Pick, L. and Noll, M. (1994). Multiple regulatory elements direct the complex expression pattern of the *Drosophila* segmentation gene paired. *Mech. Dev*. **48**, 119-128.

Harding, K., Hoey, T., Warrior, R. and Levine, M. (1989). Autoregulatory and gap gene response elements of the *even-skipped* promoter of *Drosophila*. *EMBO J*. **8**, 1205-1212.

Hiromi, Y., Kuroiwa, A. and Gehring, W. J. (1985). Control elements of the *Drosophila* segmentation gene *fushi tarazu*. *Cell* **43**, 603-613.

Hoch, M., Schröder, C., Seifert, E. and Jäckle, H. (1990). *Cis*-acting control elements for *Krüppel* expression in the *Drosophila* embryo. *EMBO J*. **9**, 2587-2595.

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

Howard, K. R. and Struhl, G. (1990). Decoding positional information: regulation of the pair-rule gene *hairy*. *Development* **110**, 1223-1231.

Huang, J.-D., Schwyter, D. H., Shirokawa, J. M. and Courey, A. J. (1993). The interplay between multiple enhancer and silencer elements defines the pattern of *decapentaplegic* expression. *Genes Dev*. **7**, 694-704.

Huang, J.-D., Dubnicoff, T., Liaw, G.-J., Bai, Y., Valentine, S. A., Shirokawa, J. M., Lengyel, J. A. and Courey, A. J. (1995). Binding sites for transcription factor NTF-1/Elf-1 contribute to the ventral repression of *decapentaplegic*. *Genes Dev*. **9**, 3177-3189.

Ip, Y. T., Park, R. E., Kosman, D., Bier, E. and Levine, M. (1992a). The *dorsal* gradient morphogen regulates stripes of *rhomboid* expression in the presumptive neuroectoderm of the *Drosophila* embryo. *Genes Dev*. **6**, 1728-1739.

Ip, Y. T., Park, R. E., Kosman, D., Yazdanbakhsh, K. and Levine, M. (1992b). *dorsal-twist* interactions establish *snail* expression in the presumptive mesoderm of the *Drosophila* embryo. *Genes Dev.* **6**, 1518-1530.

Jäckle, H., Hoch, M., Pankratz, M. J., Gerwin, N., Sauer, F. and Bronner, G. (1992). Transcriptional control by *Drosophila* gap genes. *J. Cell Sci. Suppl*. **16**, 39-51.

Jackson, P. D. and Hoffman, F. M. (1994). Embryonic expression patterns of the *Drosophila decapentaplegic* gene: separate regulatory elements control blastoderm expression and lateral ectodermal expression. *Dev. Dyn.* **199**, 28-44.

Jackson, P. D., Evans, T., Nickol, J. M. and Felsenfeld, G. (1989). Developmental modulation of protein binding to β-globin gene regulatory sites within chicken erythrocyte nuclei. *Genes Dev*. **3**, 1860-1873.

Kirchhamer, C. V. and Davidson, E. H. (1996). Spatial and temporal information processing in the sea urchin embryo: Modular and intramodular organization of the *CyIIIa* gene *cis*-regulatory system. *Development* **122**, 333-348.

Kirchhamer, C. V., Yuh, C.-H. and Davidson, E. H. (1996a). Modular *cis*-regulatory organization of developmentally expressed genes: Two genes transcribed territorially in the sea urchin embryo, and additional examples. *Proc. Natl. Acad. Sci. USA* **93**, 9322-9328.

Kirchhamer, C. V., Bogarad, L. D. and Davidson, E. H. (1996b). Developmental expression of synthetic *cis*-regulatory systems composed of spatial control elements from two different genes. *Proc. Natl. Acad. Sci. USA* **93**, 13849-13854.

Kozlowski, D. J., Gagnon, M. L., Marchant, J. K., Reynolds, S. D., Angerer, L. M. and Angerer, R. C. (1996). Characterization of a *SpAN* promoter sufficient to mediate correct spatial regulation along the animal-vegetal axis of the sea urchin embryo. *Dev. Biol*. **176**, 95-107.

Larkin, S, Farrance, I. K. G. and Ordahl, C. P. (1996). Flanking sequences modulate the cell specificity of M-CAT elements. *Mol. Cell. Biol.* **16**, 3742-3755.

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

Leiden, J. M. (1993). Transcriptional regulation of T-cell receptor genes. *Annu. Rev. Immunol.* **11**, 539-570.

Levine, M. S. and Harding, K. W. (1989). *Drosophila*: the zygotic contribution. In *Genes and Embryos* (ed. D. M. Glover and B. D. Hames), pp. 39-94. New York: IRL Press.

Liaw, G.-J., Rudolph, K. M., Huang, J.-D., Dubnicoff, T., Courey, A. J. and Lengyel, J. A. (1995). The *torso* response element binds GAGA and NTF-1/Elf-1, and regulates *tailless* by relief of repression. *Genes Dev.* **9**, 3163-3176.

Lilly, B., Galewsky, S., Firulli, A. B., Schultz, R. A. and Olson, E. N. (1994). D-MEF2: A MADS box transcription factor expressed in differentiating mesoderm and muscle cell lineages during *Drosophila* embryogenesis. *Proc. Natl. Acad. Sci. USA* **91**, 5662-5666.

MacLellan, W. R., Lee, T.-C., Schwartz, R. J. and Schneider, M. D. (1994). Transforming growth factor-β response elements of the skeletal α-actin gene. *J. Biol. Chem*. **269**, 16754-16760.

Maggert, K., Levine, M. and Frasch, M. (1995). The somatic-visceral subdivision of the embryonic mesoderm is initiated by dorsal gradient thresholds in *Drosophila*. *Development* **121**, 2107-2116.

Makabe, K. W., Kirchhamer, C. V., Britten, R. J. and Davidson, E. H. (1995). *Cis*-regulatory control of the *SM50* gene, an early marker of skeletogenic lineage specification in the sea urchin embryo. *Development* **121**, 1957-1970.

Mar, J. H. and Ordahl, C. P. (1988). A conserved CATTCCT motif is required for skeletal muscle-specific activity of the cardiac troponin T gene promoter. *Proc. Natl. Acad. Sci. USA* **85**, 6404-6408.

Martin, C. H., Mayeda, C. A., Davis, C. A., Ericsson, E. L., Knafels, J. D., Mathog, D. R., Celniker, S. E., Lewis, E. B. and Palazzolo, M. J. (1995). Complete sequence of the bithorax complex of *Drosophila*. *Proc. Natl. Acad. Sci. USA* **92**, 8398-8402.

McClay, D. R. and Logan, C. Y. (1996). Regulative capacity of the archenteron during gastrulation in the sea urchin. *Development* **122**, 607-616.

McGinnis, W. and Krumlauf, R. (1992). Homeobox genes and axial patterning. *Cell* **68**, 283-302.

McGrew, M. J., Bogdanova, N., Hasegawa, K., Hughes, S. H., Kitsis, R. N. and Rosenthal, N. (1996). Distinct gene expression patterns in skeletal and cardiac muscle are dependent on common regulatory sequences in the *MLC1/3* locus. *Mol. Cell. Biol*. **16**, 4524-4534.

Molnár, A and Georgopoulos, K. (1994). The Ikaros gene encoses a family of functionally diverse zinc finger DNA-binding proteins. *Mol. Cell. Biol.* **14**, 82992-8303.

Morgan, T. H. (1934). *Embryology and Genetics*. New York: Columbia University Press.

Morrison, A., Moroni, M. C., Ariza-McNaughton, L., Krumlauf, R. and Mavilio, F. (1996). In vitro and transgenic analysis of a human HOX4 retinoid-responsive enhancer. *Development* **122**, 1895-1907.

Moss, J. B., McQuinn, T. C. and Schwartz, R. J. (1994). The avian cardiac α-actin promoter is regulated through a pair of complex elements composed of E boxes and serum response elements that bind both positive- and negative-activing factors. *J. Biol. Chem*. **269**, 12731-12740.

Mueller, P. R. and Wold, B. (1989). In vivo footprinting of a muscle-specific enhancer by ligation-mediated PCR. *Science* **246**, 780-786.

Müller, J. and Bienz, M. (1992). Sharp anterior boundary of homeotic gene expression conferred by the *fushi tarazu* protein. *EMBO J*. **11**, 3653-3661.

Nemer, M., Stuebing, E. W., Bai, G. and Parker, H. R. (1995). Spatial regulation of *SpMTA* metallothionein gene expression in sea urchin embryos by a regulatory cassette in intron 1. *Mech. Dev*. **50**, 131-137.

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

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

Perkins, K. K., Dailey, G. M. and Tjian, R. (1988). In vitro analysis of the *Antennapedia* P2 promoter: identification of a new *Drosophila* transcription factor. *Genes Dev*. **2**, 1615-1626.

Philipsen, S., Talbot, D., Fraser, P. and Grosveld, F. (1990). The β-globin dominant control region: hypersensitive site 2. *EMBO J*. **9**, 2159-2167.

Pöpperl, H., Bienz, M., Studer, M., Chan, S.-K., Aparicio, S., Brenner, S., Mann, R. S. and Krumlauf, R. (1995). Segmental expression of *Hoxb-1* is controlled by a highly conserved autoregulatory loop dependent upon *exd/pbx*. *Cell* **81**, 1031-1042.

Pruzina, S., Hanscombe, O., Whyatt, D., Grosveld, F. and Philipsen, S. (1991). Hypersensitive site 4 of the human β-globin locus control region. *Nucl. Acids Res*. **19**, 1413-1419.

Ranganayakulu, G., Schulz, R. A. and Olson, E. N. (1996). Wingless signaling induces *nautilus* expression in the ventral mesoderm of the *Drosophila* embryo. *Dev. Biol*. **176**, 143-148.

Rao, M. V., Donoghue, M. J., Merlie, J. P. and Sanes, J. R. (1996). Distinct regulatory elements control muscle-specific, fiber-type-selective, and axially graded expression of a myosin light-chain gene in transgenic mice. *Mol. Cell. Biol*. **16**, 3909-3922.

Rhodes, S. J., Chen, R., DiMattia, G. E., Scully, K. M., Kalla, K. A., Lin, S.-C., Yu, V. C. and Rosenfeld, M. G. (1993). A tissue-specific enhancer confers Pit-1-dependent morphogen inducibility and autoregulation on the *pit-1* gene. *Genes Dev*. **7**, 913-932.

Riddihough, G. and Ish-Horowicz, D. (1991). Individual stripe regulatory elements in the *Drosophila hairy* promoter respond to maternal, gap, and pair-rule genes. *Genes Dev*. **5**, 840-854.

Rivera-Pomar, R., Lu, X., Perrimon, N., Taubert, H. and Jäckle, H. (1995). Activation of posterior gap gene expression in the *Drosophila* blastoderm. *Nature* **376**, 253-256.

Rosenthal, N., Wentworth, B., Engert, J., Grieshammer, U., Berglund, E. and Gong, X. (1992). The myosin light-chain 1/3 locus. A model for developmental control of skeletal muscle differentiation. In *Neuromuscular Development and Disease* (ed. A. M. Kelly and H. M. Blau), pp. 131-144. New York: Raven Press.

Rothenberg, E. V. and Ward, S. B. (1996). A dynamic assembly of diverse transcription factors integrates activation and cell-type information for interleukin 2 gene regulation. *Proc. Natl. Acad. Sci. USA* **93**, 9358-9365.

Schoenherr, C. J. and Anderson, D. J. (1995a). Silencing is golden - Negative regulation in the control of neuronal gene transcription. *Curr. Opin. Neurobiol*. **5**, 566-571.

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

Shashikant, C.S. and Ruddle, F. H. (1996). Combinations of closely situated *cis*-acting elements determine tissue-specific patterns and anterior extent of early *Hoxc8* expression. *Proc. Natl. Acad. Sci. USA* **93**, 12364-12369.

Shashikant, C.S., Bieberich, C.J., Belting, H.-G., Wang, J. C. H., Borbély, M. A. and Ruddle, F. H. (1995). Regulation of *Hoxc-8* during mouse embryonic development: Identification and characterization of critical elements involved in early neural tube expression. *Development* **121**, 4339-4347.

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

Singson, A., Leviten, M. W., Bang, A. G., Hua, X. H. and Posakony, J. W. (1994). Direct downstream targets of proneural activators in he imaginal disc include genes involved in lateral inhibitory signaling. *Genes Dev.* **8**, 2058-2071.

Sladek, F. M., Zhong, W., Lai, E. and Darnell, J. E., Jr. (1990). Liver-enriched transcription factor HNF-4 is a novel member of the steroid hormone receptor superfamily. *Genes Dev*. **4**, 2353-2365.

Small, S., Blair, A. and Levine, M. (1992). Regulation of *even-skipped* stripe 2 in the *Drosophila* embryo. *EMBO J*. **11**, 4047-4057.

Small, S., Arnosti, D. N. and Levine, M. (1993). Spacing ensures autonomous expression of different stripe enhancers in the *even-skipped* promoter. *Development* **119**, 767-772.

Small, S., Blair, A. and Levine, M. (1996). Regulation of two pair-rule stripes by a single enhancer in the *Drosophila* embryo. *Dev. Biol*. **175**, 314-324.

Sornson, M. W., Wu, W., Dasen, J. S., Flynn, S. E., Norman, D. J., O'Connell, S. M., Gukovsky, I., Carrière, Ryan, A. K., Miller, A. P., Zuo, L., Gleiberman, A. S., Andersen, B., Beamer, W. G. and Rosenfeld, M. G. (1996). Pituitary lineage determination by the *Prophet of Pit-1* homeodomain factor defective in Ames dwarfism. *Nature* **384**, 327-333.

Staehling-Hampton, K., Hoffmann, F. M., Baylies, M. K., Rushton, E. and Bate, M. (1994). *dpp* induces mesodermal gene expression in *Drosophila*. *Nature* **372**, 783-786.

Strauss, E. C. and Orkin, S. H. (1992). In vivo protein-DNA interactions at hypersensitive site 3 of the human β-globin locus control region. *Proc. Natl. Acad. Sci. USA* **89**, 5809-5813.

Studer, M., Pöpperl, H., Marshall, H., Kuroiwa, A. and Krumlauf, R. (1994). Role of a conserved retinoic acid response element in rhombomere restriction of *Hoxb-1*. *Science* **265**, 1728-1732.

Taylor, M. V., Beatty, K. E., Hunter, H. K. and Baylies, M. K. (1995). *Drosophila* MEF2 is regulated by *twist* and is expressed in both the primordia and differentiated cells of the embryonic somatic, visceral and heart musculature. *Mech. Dev*. **50**, 29-41.

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

Topol, J., Dearolf, C. R., Prakash, K. and Parker, C. S. (1991). Synthetic oligonucleotides recreate *Drosophila fushi tarazu* zebra-stripe expression. *Genes Dev*. **5**, 855-867.

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

Wei, Z., Angerer, L. M., Gagnon, M. L. and Angerer, R. C. (1995). Characterization of the SpHE promoter that is spatially regulated along the animal-vegetal axis of the sea urchin embryo. *Dev. Biol*. **171**, 195-211.

Wijgerde, M., Grosveld, F. and Fraser, P. (1995). Transcription complex stability and chromatin dynamics in vivo. *Nature* **377**, 209-213.

Wimmer, E. A., Simpson-Brose, M., Cohen, S. M., Desplan, C. and Jäckle, H. (1995). *Trans*- and *cis*-acting requirements for blastodermal expression of the head gap gene *buttonhead*. *Mech. Dev*. **53**, 235-245.

Wotton, D., Lake, R. A., Farr, C.J. and Owen, M. J. (1995). The high mobility group transcription factor, SOX4, transactivates the human *CD2* enhancer. *J. Biol. Chem*. **270**, 7515-7522.

Yeh, W.-C., Cao, Z., Classon, M. and McKnight, S. L. (1995). Cascade regulation of terminal adipocyte differentiation by three members of the C/EBP family of leucine zipper proteins. *Genes Dev*. **9**, 168-181.

Yuh, C.-H. and Davidson, E. H. (1996). Modular *cis*-regulatory organization of *Endo16*, a gut-specific gene of the sea urchin embryo. *Development* **122**, 1069-1082.

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

Yuh, C.-H., Moore, J. G. and Davidson, E. H. (1996). Quantitative functional interrelations within the *cis*-regulatory system of the *S. purpuratus Endo16* gene. *Development* **122**, 4045-4056.

Zeller, R. W., Griffith, J. D., Moore, J. G., Kirchhamer, C. V., Britten, R. J. and Davidson, E. H. (1995a). A multimerizing transcription factor of sea urchin embryos capable of looping DNA. *Proc. Natl. Acad. Sci. USA* **92**, 2989-2993.

Zeller, R. W., Britten, R. J. and Davidson, E. H. (1995b). Developmental utilization of SpP3A1 and SpP3A2: two proteins which recognize the same DNA target site in several sea urchin gene regulatory regions. *Dev. Biol*. **170**, 75-82.

Zhang, C.-C., Müller, J., Hoch, M., Jäckle, H. and Bienz, M. (1991). Target sequences for *hunchback* in a control region conferring *Ultrabithorax* expression boundaries. *Development* **113**, 1171-1179.

(Accepted 14 March 1997)