The EMBO Journal vol.13 no.10 pp.2362–2377, 1994

Deformed protein binding sites and cofactor binding sites are required for the function of a small segment-specific regulatory element in *Drosophila* embryos

Chaoyang Zeng[1], Julia Pinsonneault[1], Gabriel Gellon[2], Nadine McGinnis[1] and William McGinnis[1,2]

Departments of [1]Molecular Biophysics and Biochemistry and [2]Biology, Yale University, New Haven, CT 06520-8114, USA

Communicated by M.Noll

How each of the homeotic selector proteins can regulate distinct sets of DNA target elements in embryos is not understood. Here we describe a detailed functional dissection of a small element that is specifically regulated by the Deformed homeotic protein. This 120 bp element (module E) is part of a larger 2.7 kb autoregulatory enhancer that maintains *Deformed* (*Dfd*) transcription in the epidermis of the maxillary and mandibular segments of *Drosophila* embryos. *In vitro* binding assays show that module E contains only one Dfd protein binding site. Mutations in the Dfd binding site that increase or decrease its *in vitro* affinity for Dfd protein generate parallel changes in the regulatory activity of module E in transgenic embryos, strong evidence that the *in vitro*-defined binding site is a direct target of Dfd protein in embryos. However, a monomer or multimer of the Dfd binding region alone is not sufficient to supply *Dfd*-dependent, segment-specific reporter gene expression. An analysis of a systematic series of clustered point mutations in module E revealed that an additional region containing an imperfect inverted repeat sequence is also required for the function of this homeotic protein response element. The Dfd binding site and the putative cofactor binding site(s) in the region of the inverted repeat are both necessary and in combination sufficient for the function of module E.

Key words: autoregulation/*Deformed*/*Drosophila*/enhancer/homeotic

Introduction

Homeotic selector (HOM) genes of *Drosophila melanogaster* are responsible for determining regional identities along the anterior–posterior axis of the fly body. In developing embryos, HOM genes are expressed in specific segments or parasegments and usually instruct these regions to adopt unique morphology [Kaufman *et al.*, 1980; Bender *et al.*, 1983; reviewed in García-Bellido (1977), Lewis (1978), Kaufman *et al.* (1990) and McGinnis and Krumlauf (1992)]. All the known HOM proteins, *labial* (*lab*), *proboscipedia* (*pb*), *Deformed* (*Dfd*), *Sex combs reduced* (*Scr*), *Antennapedia* (*Antp*), *Ultrabithorax* (*Ubx*), *abdominal-A* (*abd-A*) and *Abdominal-B* (*Abd-B*), contain similar 61 amino acid homeodomains which are known or suspected to bind similar DNA sequences and mediate transcriptional regulation [reviewed in Scott *et al.* (1989), Affolter *et al.* (1990), Hayashi and Scott (1990), Laughon (1991), Gehring (1992) and Dessain and McGinnis (1993)]. The ability of these proteins to activate or repress the expression of distinct sets of downstream 'realizator' genes is believed to explain their remarkable capacity to direct cells into one of several alternative developmental pathways which eventually results in different structures in different segments [García-Bellido, 1977; reviewed in Morata (1993)]. HOM-type gene clusters appear to be structurally conserved in many or all animal species, with evidence for functional conservation in flies, beetles, mice and nematodes [Duboule and Dollé, 1989; Graham *et al.*, 1989; Wang *et al.*, 1993; reviewed in McGinnis and Krumlauf (1992) and Krumlauf (1993)], indicating an ancient and conserved role in specifying positional identities on the anterior–posterior axis of developing metazoans.

Since the amino acid sequences of HOM-type homeodomains are 60–90% identical in pairwise comparisons and all except Abd-B have identical 'recognition helices' (helix III), it is not surprising that HOM proteins have similar DNA binding specificities *in vitro* (Desplan *et al.*, 1988; Hoey and Levine, 1988; Laughon, 1991; Dessain *et al.*, 1992; Ekker *et al.*, 1992; Vachon *et al.*, 1992; Appel and Sakonju, 1993). There are indeed different preferred optimal binding sites for some HOM proteins (Dessain *et al.*, 1992; Ekker *et al.*, 1992; Appel and Sakonju, 1993), but it seems extremely unlikely that the modest differences in affinity for these sites are sufficient to account for specific target selection of HOM-type or other types of homeodomain proteins in embryos. For example, an artificial enhancer containing a multimerized *in vitro* binding site for the homeodomain protein engrailed (en) is not activated by en protein in embryos, though it is possible that en is binding and exerting a repressive effect through the multimerized sites (Vincent *et al.*, 1990). Nor is a small element with multiple fushi tarazu (ftz) protein binding sites sufficient to confer *ftz*-dependent expression in embryos, even though ftz binding sites can mediate an activation response in the context of larger composite elements (Nelson and Laughon, 1993; Schier and Gehring, 1993b). It seems that protein–protein interactions with as yet undefined protein cofactors, or other influences, are also required for the *in vivo* target specificity of most homeodomain proteins. We use the term target specificity simply to indicate that a homeodomain protein can activate or repress a specific transcription unit, or act through a specific enhancer in embryos. Target specificity is not meant to imply a specific mechanism such as cooperative DNA binding. To elucidate the molecular mechanisms underlining the targeting specificity of the HOM proteins, it is crucially important to obtain a detailed understanding of their normal *cis*-response elements in the context of developing embryos.

In this paper we report the detailed dissection of a HOM

protein response element — the *Dfd* epidermal autoregulatory element (*Dfd* EAE). *Dfd* is a homeotic selector gene required for establishing and maintaining the identities of maxillary and mandibular segments. Early patterning genes initiate *Dfd* expression in the posterior head region of the anterior–posterior axis. Once established, the Dfd protein is capable of activating its own transcription (Kuziora and McGinnis, 1988; Bergson and McGinnis, 1990; McGinnis *et al.*, 1990; Regulski *et al.*, 1991). A 2.7 kb DNA sequence upstream of the *Dfd* transcription unit has been identified as an epidermal autoregulatory element (Bergson and McGinnis, 1990). At least part of this EAE requires Dfd protein binding sites for function (Regulski *et al.*, 1991).

We chose to dissect the *Dfd* EAE because very few other target elements of Dfd or other HOM proteins are known at present. *Distal-less (Dll)* and *1.28* are two genes reported recently to be regulated by *Dfd* (Mahaffey *et al.*, 1993; O'Hara *et al.*, 1993), but whether they are directly regulated by Dfd protein is not yet known. By inserting *Dfd* responsive elements from the *Dfd* EAE in front of a reporter gene and reintroducing the constructs into random locations of the fly genome by P element transformation, we generated artificial 'downstream genes' for *Dfd* in those locations. There is another reason to focus on the *Dfd* EAE as an example of a HOM protein response element. In transgenic mouse embryos, the *Drosophila Dfd* EAE activates reporter gene expression within the anterior region of the expression domain of *mouse Dfd*-like genes, such as *Hoxb-4* (Awgulewitsch and Jacobs, 1992). It has also been discovered recently that Hoxb-4 protein function is required for persistent transcription of the *Hoxb-4* locus in the anterior regions of its CNS expression domain (Ramírez-Solis *et al.*, 1993). Thus, we believe that the details of the *Dfd* autoregulatory circuitry derived from the study of *Dfd* EAE may have general implications for the targeting specificity of Dfd-like proteins from many animal species.

We report here that the 2.7 kb *Dfd* EAE contains multiple modules that can function independently. Deletion analysis of a 0.47 kb fragment (module F) suggests that it is a composite of multiple binding sites that are all required for full activity. We focus most of our attention on a small but potent 120 bp autoregulatory sequence (module E). *In vitro* footprinting experiments of module E detect only a single Dfd binding site that we show is likely to be directly bound by Dfd protein in developing embryos. We also identify a block of nearby DNA sequence that apparently contains functionally important cofactor binding sites. Although neither the Dfd binding site alone nor the putative cofactor binding sites alone are capable of specifically responding to Dfd regulatory activity; a combination of both is sufficient to reconstitute nearly all the activity of module E.

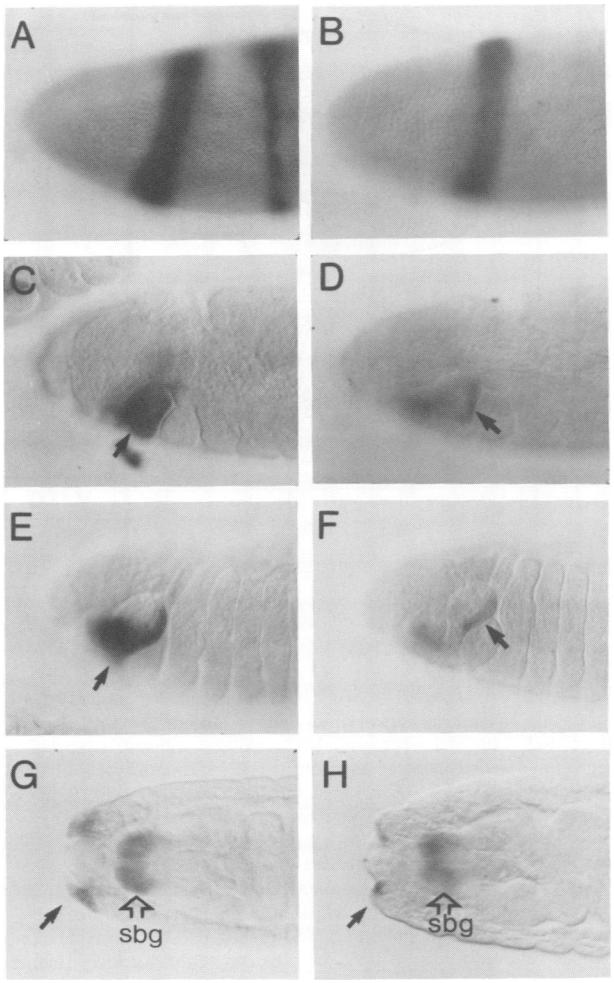

Fig. 1. Dfd protein is required for maintaining persistent *Dfd* transcription. Left panels show wild type embryos, right panels show Dfd^{w21} embryos (similar results were obtained with Dfd^{rR11} embryos). Transcripts were detected with digoxygenin-labeled probes complementary to *Dfd* and *lacZ*. Embryos homozygous for Dfd^{w21} were identified by the absence of a *Ubx*-like *lacZ* expression pattern produced by a *Ubx-lacZ* transgene inserted in the balancer chromosome. Anterior is to the left in all panels. Solid arrows point to the maxillary lobe. Open arrows indicate the subesophageal ganglions (sbg) of the central nervous system. (**A**) *Dfd* transcripts in wild type embryos at the cellular blastoderm stage (stage 5, lateral view). The more posterior stripe is due to the *Ubx-lacZ* transgene. (**B**) Dfd^{w21} embryos at the same stage as in (A). (**C** and **D**) Lateral view of embryos at the late germ band extended stage (stage 10). In the Dfd^{w21} embryos in (D), the transcripts are limited to a few lateral–posterior cells of the maxillary segment. (**E** and **F**) Lateral view of embryos at the germ band retracted stage (stage 12). (**G** and **H**) Ventral view of embryos near the completion of head involution (stage 16). The remaining signal in the sbg (open arrow) of the Dfd^{w21} embryos in (H) seems to be slightly stronger than that in the epidermis (solid arrow).

Results

Dfd protein is required for maintenance of Dfd transcription

A stripe of *Dfd* expression is established in the posterior head primordia of cellular blastoderm embryos (stage 5) under the control of coordinate, gap and segmentation genes such as *bicoid*, *hunchback*, *fushi tarazu* and *even-skipped* (Jack *et al.*, 1988; Jack and McGinnis, 1990). Because the protein products of these genes are no longer expressed shortly after the *Dfd* stripe is established, they cannot be responsible for the maintenance phase of *Dfd* transcriptional activation which continues for the rest of embryogenesis. Several experiments have indicated that *Dfd* is capable of activating its own transcription. When transient *Dfd* expression is induced ubiquitously in embryos transgenic for heat shock-inducible *Dfd* cDNAs (*hs-Dfd*), persistent transcription from the endogenous *Dfd* locus is ectopically induced in groups of cells in the ventral–posterior region of most segments (Kuziora and McGinnis, 1988). This was presumed to

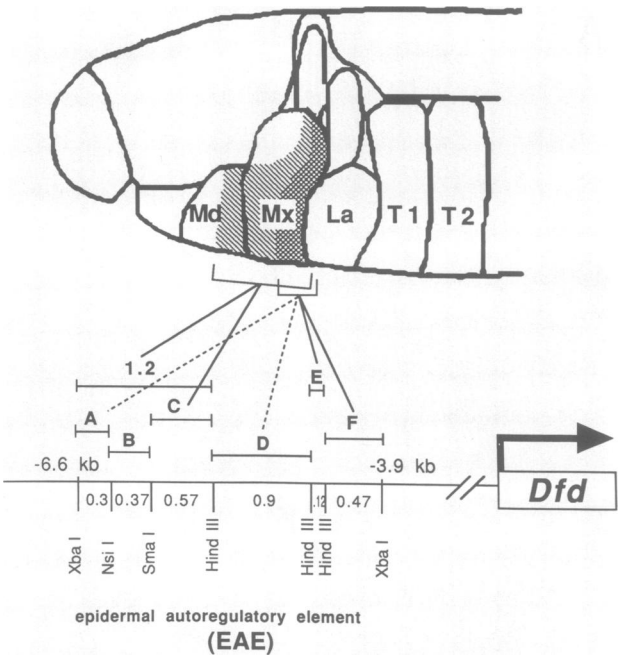

Fig. 2. The structure of the *Dfd* EAE and the subelements. (**Top**) Schematic of the head region of an embryo at germ band retracted stage: Md, mandibular segment; Mx, maxillary segment; La, labial segment; T1, first thoracic segment; and T2, second thoracic segment. (**Bottom**) The 2.7 kb *Dfd* epidermal autoregulatory element (EAE) is localized from 3.9 to 6.6 kb upstream of the *Dfd* transcription initiation site (Bergson and McGinnis, 1990). Using the shown restriction sites, the EAE is divided into modules A–F, whose sizes (in kb) are indicated under the line. Modules E and F each independently drive reporter expression in the posterior part of the maxillary segment. Modules A and D (dotted lines) drive expression in regions believed to derive from the posterior of the maxillary segment at late embryonic stages after head involution. Module C or a larger 1.2 kb fragment drives expression both in the posterior mandibular cells and in *Dfd*-expressing cells of the maxillary segment.

indicate that the maintenance of *Dfd* transcription in its normal domain, the maxillary and mandibular segments, also requires the function of Dfd protein.

However, Dfd autoregulation might be one of several redundant mechanisms for maintaining *Dfd* expression. To test whether Dfd protein function is actually required for persistent *Dfd* transcription in its normal expression domain, we assayed the transcriptional expression pattern of two mutant *Dfd* genes that have termination codons prior to the homeodomain coding sequence, and thus presumably produce truncated, inactive proteins. Figure 1 shows the pattern of *Dfd* transcripts at successive stages of embryogenesis obtained with one of these null mutations, Dfd^{w21}. In this mutant allele the 346th codon, which codes for Trp in wild type *Dfd*, is changed to a stop codon [see Regulski *et al.* (1987) for the Dfd coding sequence]. At the cellular blastoderm stage, embryos homozygous for Dfd^{w21} establish a normal amount of *Dfd* transcripts in a normal pattern (Figure 1A and B). However, shortly after this initial burst of expression, *Dfd* transcripts are eliminated from the epidermis except for a weak residual pattern consisting of a few cells near the posterior border of the maxillary segment (Figure 1). *Dfd* transcripts also persist in cells that will be incorporated into the subesophageal ganglions (sbg) of the central nervous system (CNS). However, the CNS transcripts also appear to be in fewer cells and at lower levels than in wild type embryos. Similar results were also obtained with the Dfd^{rR11} mutation, in which codon number 210 (Tyr) is mutated to a stop codon. We conclude that in most cells of the maxillary and mandibular epidermis the maintenance phase of the *Dfd* transcription pattern is completely dependent on the presence of Dfd protein. In the posterior maxillary segment and CNS, a low level of *Dfd* transcription is maintained by mechanism(s) that are independent of the *Dfd* autoregulatory circuit.

Multiple Dfd EAE modules with overlapping functions

Cis-regulatory sequences that recapitulate the maintenance phase of *Dfd* transcription in the embryonic epidermis map between 6.6 and 3.9 kb upstream of the transcription initiation site (Figure 2; Bergson and McGinnis, 1990). This 2.7 kb *Dfd* EAE activates reporter gene expression specifically in the maxillary and mandibular epidermis from extended germ band stage [stage 10 or 11 of Campos-Ortega and Hartenstein (1985)] of embryogenesis onwards. In *Dfd* mutant embryos, this element exhibits no activity until late stages of embryogenesis (stage 15 or 16) when some weak reporter expression is detected in the maxillary region of *Dfd* mutant embryos.

In an effort to define a small *Dfd*-dependent element that could be analyzed in detail, we tested the functions of fragments of the *Dfd* EAE. The 2.7 kb EAE was arbitrarily divided into six sequence modules, A–F (Figures 2 and 4). These modules were individually cloned into the reporter constructs pHZ50 (Hiromi and Gehring, 1987) or pCaSpeR hs43 lacZ (Thummel and Pirrotta, 1991), which are P-element transformation vectors containing the basal promoter of the *hsp70* gene upstream of coding sequences for *Escherichia coli* β-galactosidase (β-gal). Multiple independent lines of transgenic flies were established for each construct and β-gal expression patterns were detected by antibody staining of whole mount embryos.

Module A, which is 294 bp in size, directs weak β-gal expression in two groups of about five to 10 cells that are within the *Dfd* expression domain during the late stages of embryogenesis (from stage 16 onwards; data not shown). Module B, 370 bp, does not generate consistent expression. In contrast, expression driven by the 570 bp module C can be detected in early- to mid-stage embryos (stage 11) in a small group of cells located in the medial–ventral part of the maxillary segment (Figure 3A). By the time the germ band is retracted, the module C expression pattern has expanded from the initial spot to include much of the maxillary epidermis and the posterior mandibular epidermis, a pattern that closely mimics the epidermal expression pattern of the endogenous *Dfd* gene at this stage (compare Figures 3C and 1E). When isolated from the rest of the *Dfd* EAE, module C also provides reporter gene expression in a few posterior cells of the labial segment. These labial cells do not express Dfd protein at any stage of embryogenesis.

The expression pattern generated by module D (900 bp) has been described previously (Bergson and McGinnis, 1990; Regulski *et al.*, 1991) and consists of weak activity in the maxillary segment of stage 16 and older embryos. *In vitro*, this module contains four Dfd protein binding sites which are required for its embryonic regulatory function. Among them, sites A and D have the highest affinity for Dfd protein and share an ATCATTA consensus sequence (Regulski *et al.*, 1991).

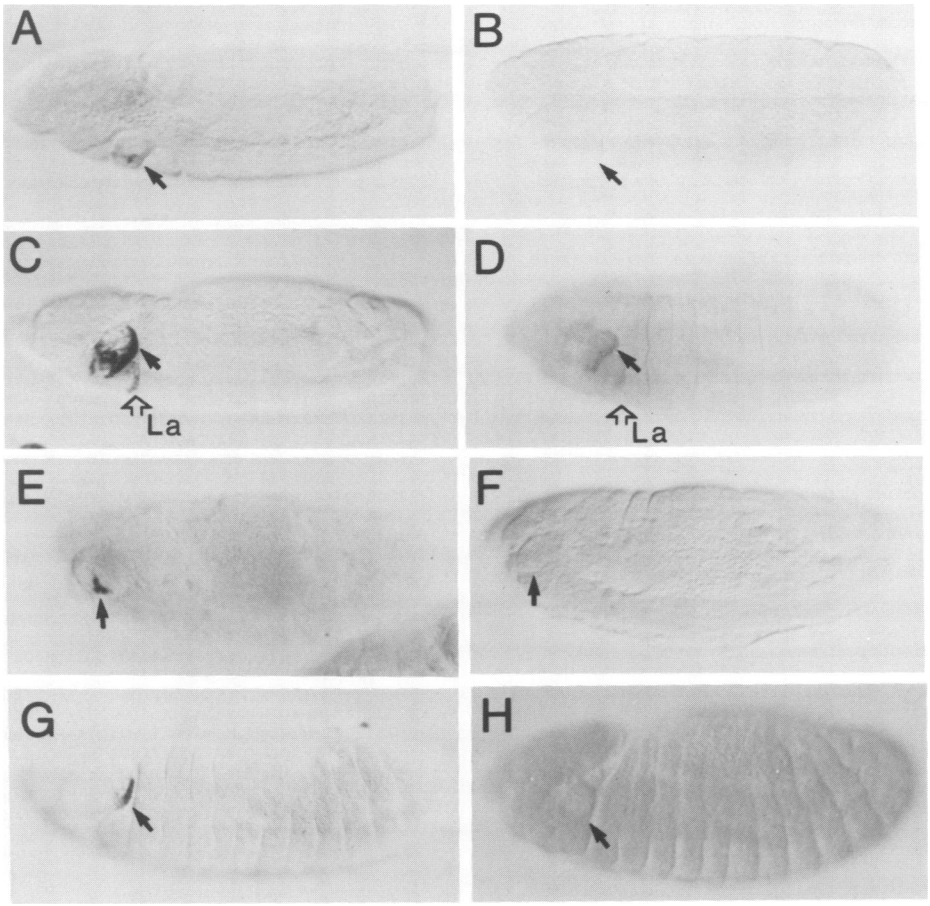

Fig. 3. The 2.7 kb *Dfd* EAE contains multiple Dfd-dependent modules. The β-gal expression patterns were detected by antibody staining of whole-mount embryos as described in the Materials and methods. The left panels show the expression patterns from different modules in wild type embryos. The right panels show them in *Dfd* mutant embryos. All embryos are oriented with anterior to the left and dorsal up. Solid arrows indicate the posterior border of the maxillary lobe. (**A**–**D**) Expression of β-gal under module C control in wild type embryos in (A) and (C), and in Dfd^- embryos in (B) and (D). At the germ band extended stage (stage 11) in (A), β-gal is expressed in a few cells in the maxillary segment. At the germ band retracted stage (stage 12), expression was detected in the posterior mandibular cells, the anterior maxillary cells and the posterior maxillary cells (C), and is almost entirely Dfd-dependent (D). Module C also contains a *Dfd*-independent regulatory activity, as evidenced by the weak expression pattern in the labial (La and the open arrows) in (C) and (D), as well as residual expression in the maxillary segments in Dfd^- embryos in (D). (**E** and **F**) Expression of β-gal under module E control was detected during the middle of head involution (stage 14) in (E) and is absent in *Dfd* mutant embryos at the same stage in (F). (**G** and **H**) Expression of β-gal under module F control was detected in the posterior of the maxillary segment at the germ band retracted stage (stage 12) in (G) and is absent in *Dfd* mutant embryos at the same stage in (H).

Module E (120 bp) activates β-gal expression in cells belonging to the maxillary segment from stage 14 to the end of embryogenesis (Figure 3E). During stage 14 the maxillary segment rotates so that cells at the original posterior border assume a more ventral position relative to the embryonic A–P axis. At this stage, the maxillary segment, along with the rest of the head segments, also begins to move anteriorly during the process of head involution. This process will conclude with many of the maxillary epidermal cells residing on the extreme anterior end of the fully developed embryo, contributing to the differentiated structures of the pseudocephalon and larval mouth such as cirri and mouth hooks (McGinnis *et al.*, 1990).

Module F, 471 bp and the most 3′ fragment of the 2.7 kb *Dfd* EAE, activates β-gal expression in the posterior cells of the maxillary segment. Expression first appears at stage 12 in six to 10 cells at the posterior border of the maxillary segment (Figure 3G). Later, the expression pattern expands to include more cells in the posterior maxillary segment.

Reporter gene expression driven by modules C, E or F requires the function of Dfd protein. When modules E and F were tested in *Dfd* mutant embryos, reporter gene expression was eliminated (Figure 3F and H). Module C still activates weak expression in the posterior maxillary epidermis in a *Dfd* mutant background and, as expected, retains the weak labial expression outside the *Dfd* expression domain (Figure 3B and D).

Dfd EAE module F is a highly composite element

Among the three relatively strong EAE modules C, E and F, module C was not analyzed further because it contains a *Dfd*-independent activity which was reflected in expression outside the *Dfd* expression domain as well as in residual maxillary activity in *Dfd* mutant embryos (Figure 3D). The dissection of modules F and E is reported in the following sections. We observed with modules F, E and their mutant derivatives that an expression pattern that appeared at an earlier stage includes relatively more cells at later stages and produces more abundant expression within those cells as judged by staining intensity. In the following, we therefore use the stage at which expression is first detected from a given construct as a rough measure of its strength.

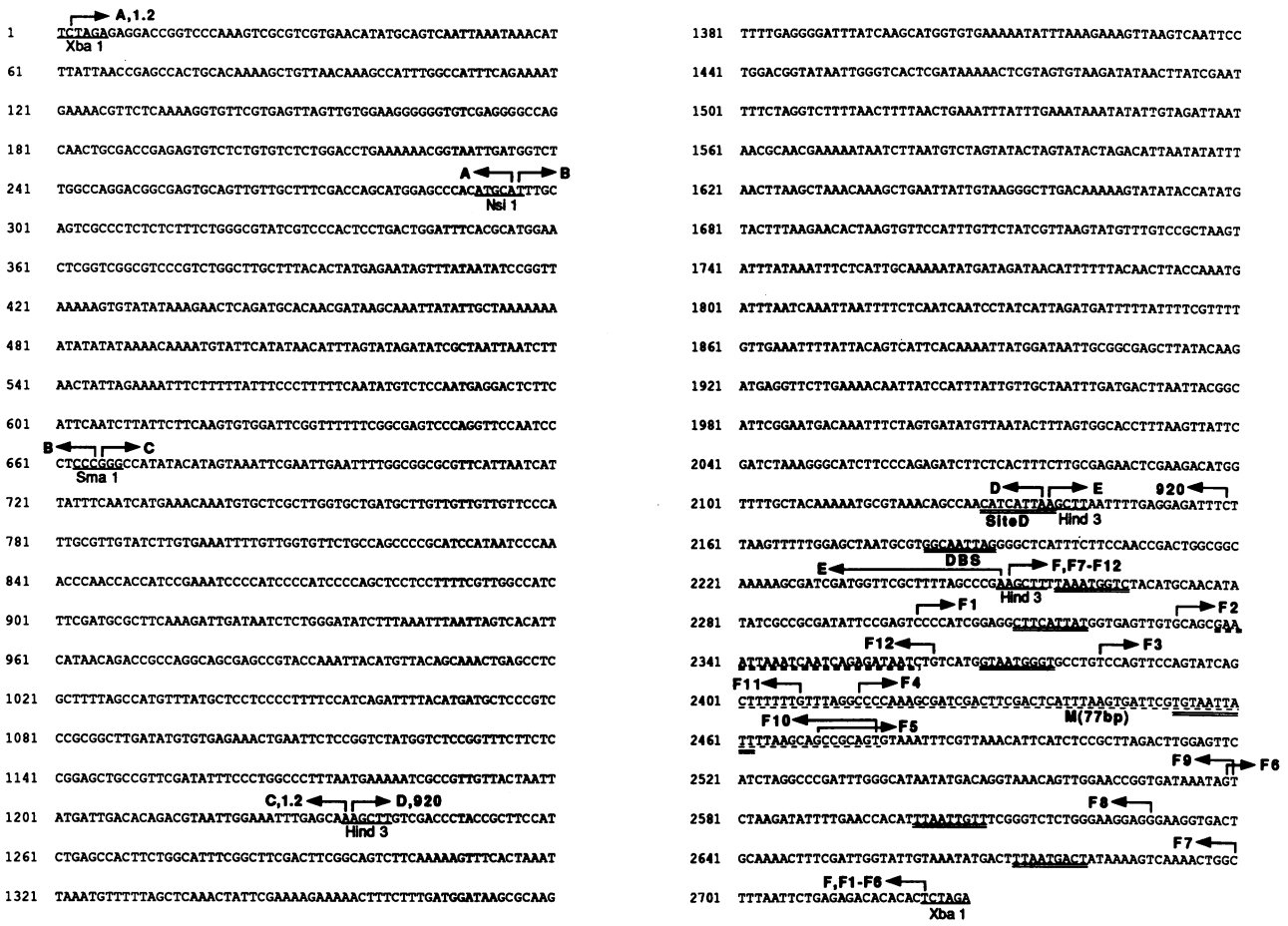

Fig. 4. DNA sequence of the 2.7 kb *Dfd* EAE. The DNA sequence begins at a position that is ~6.6 kb upstream of the *Dfd* transcription start. Restriction sites used to subdivide the EAE are indicated. The 5′ and 3′ ends of the modules, as well as those of F1–F12, the deletion derivatives of module F, are indicated by arrows. The sequence of the 77 bp module M is underlined by dotted lines. The end points of the 920 bp element and the Dfd binding site D of Regulski *et al.* (1991) are included for cross reference. The Dfd footprinting sites are double underlined. A weakly protected region is indicated by double dashed lines.

We sought to locate the essential sequences within the 471 bp module F by a systematic deletion analysis in reporter construct HZ50. As seen in Figure 4, stepwise deletions from either direction progressively weaken the strength of the element, indicating that important sequences are broadly distributed in module F. Constructs F1–F6 have deletions of 49, 80, 131, 162, 217 and 326 bp respectively from the 5′ end of the 471 bp module F (Figure 4). F1 and F2 are indistinguishable from the full-length module F in driving β-gal expression (data not shown). F3 is activated in stage 14 embryos, ~3 h later than module F (Figure 5B). F4 is activated during stage 15, ~4.5 h after module F (Figure 5C). Though rotated and translocated by the movements of head involution, the cells expressing β-gal under F3 or F4 control both derive from the same posterior maxillary subregion that expresses the parental module F construct. F5 confers weak β-gal expression at late stage 16, shortly before head involution is completed and ~7 h after the initial expression of module F (Figure 5D). This expression persists in a small part of the *Dfd* expression domain during stage 17. F6, which retains only 145 bp of the 3′ part of module F, generates no expression in embryos (data not shown).

Constructs F7–F12 have deletions of 51, 92, 144, 246, 315 and 359 bp respectively from the 3′ end of module F (Figure 4). The first three, F7–F9, are indistinguishable from the full-length module F in driving β-gal expression (data not shown). F10 confers reporter expression in the maxillary segment at the beginning of stage 14 (Figure 5E), a delay of ~2 h compared with the parental module. F11 generates β-gal expression in a small subset of *Dfd*-expressing cells at early stage 16 (Figure 5F), a delay of ~6 h compared with the full-length module F. F12, which retains only 112 bp of the 5′ part of module F, produces no β-gal expression at any embryonic stage (Figure 5G).

A relatively severe reduction in function was observed in the deletion step between F10 and F11, which differ by only 69 bp. To find out whether this region could function autonomously, we tested a 77 bp DNA fragment called M which includes the 69 bp difference between F10 and F11 (Figure 4). Module M activates β-gal expression at stage 17 in cells belonging to the maxillary segment (Figure 5H). This weak expression is similar in abundance and extent to that provided by the 900 bp module D which was analyzed by Regulski *et al.* (1991).

Cofactor binding sites are required for the function of a Dfd binding site in module E

The deletion analysis of module F suggests that its control may be highly combinatorial with multiple sequence motifs

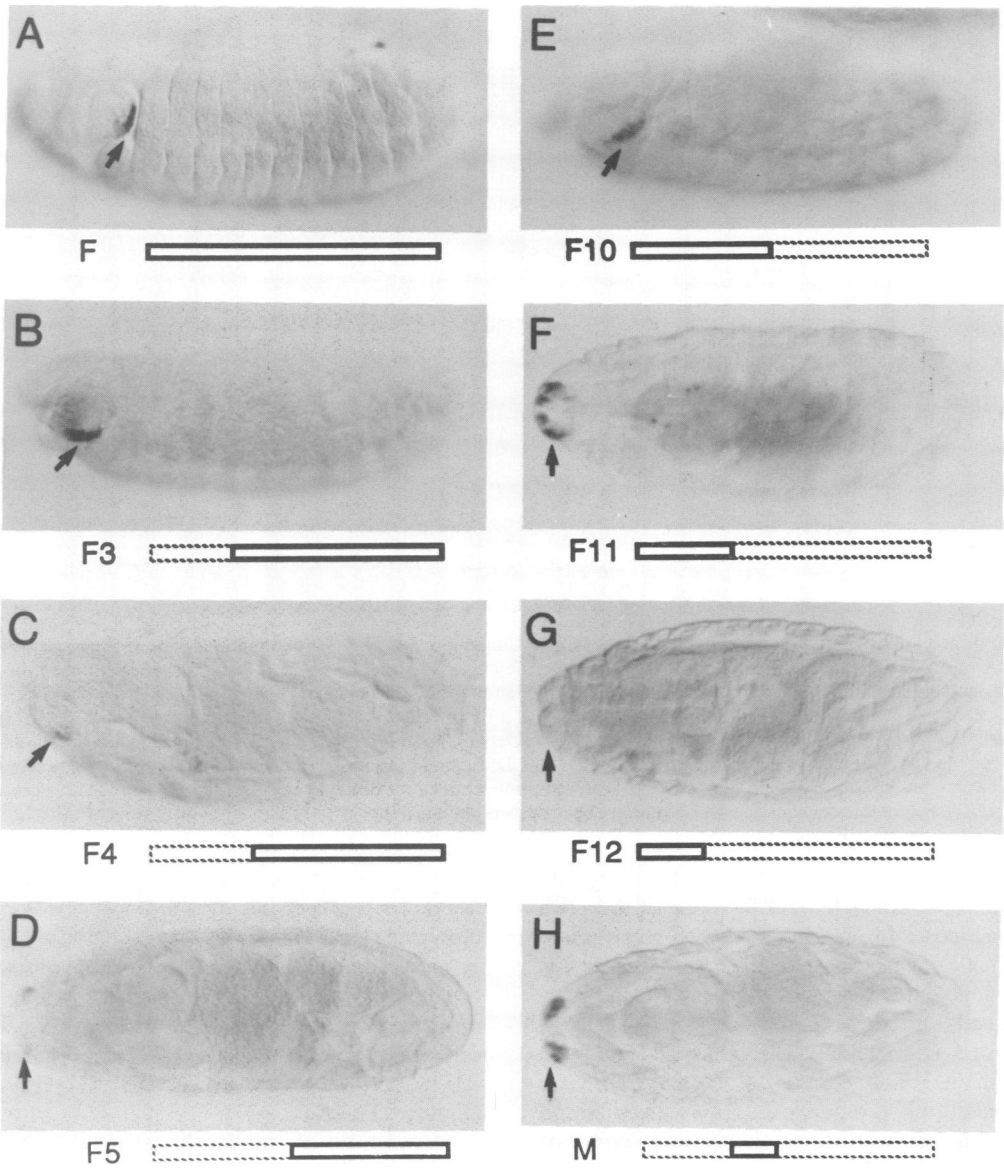

Fig. 5. Successive deletions of module F gradually reduce its regulatory activity. The β-gal expression patterns were detected by antibody staining of whole-mount embryos. Anterior is to the left in all panels. The solid boxes underneath each panel represent the sequences that are preserved in the constructs; the dotted boxes represent those deleted. Detailed DNA sequence end points are shown in Figure 4. Solid arrows indicate the maxillary segment. (**A**) Lateral view of embryos at the germ band retracted stage (stage 12) and transgenic for the intact module F. The expression pattern is detected in the posterior of the maxillary segment. (**B**) Lateral view of the β-gal expression pattern generated by construct F3 and detected shortly after stage 14, ~3 h after module F. (**C**) Lateral view of the β-gal expression pattern generated by construct F4 and detected at stage 15, ~4.5 h after module F. (**D**) Ventral view of the β-gal expression pattern generated by construct F5 and detected at stage 16, ~7 h after module F. (**E**) Lateral view of the β-gal expression pattern generated by construct F10 and detected at stage 14, ~2 h after module F. (**F**) Ventral view of the β-gal expression pattern generated by construct F11 and detected at stage 16, ~6 h after module F. (**G**) Ventral view of a stage 17 embryo transgenic for F12. No β-gal expression was detected. (**H**) Ventral view of the β-gal expression pattern generated by a 77 bp element M, which is an internal section of module F. The β-gal was first detected at stage 17, shortly before the end of embryogenesis.

distributed over a relatively large region adding to its overall strength. For detailed analysis, it would be better to identify a small but robust regulatory element with a minimal number of Dfd protein binding sites, if such an element can actually function. As we describe below, module E (Figures 2, 4 and 11) contains such an element.

To identify potential Dfd binding sites in the 120 bp of module E, full-length Dfd protein was produced in *E.coli* and used for DNase footprinting analysis. A single protected region including the sequence GGCAATTAG was detected (Figure 6). Even at high Dfd protein concentrations, this was the only site footprinted. To test whether this sequence is functionally important, we mutated the Deformed binding site (DBS) to a random sequence, ACTGTAGGA, in an otherwise wild type module E. The resulting construct EN (Figure 11) was no longer detectably bound by Dfd protein in footprinting assays (Figure 6), and forms much less stable complexes with Dfd protein in mobility shift binding assays (Figure 7). *In vivo*, construct EN was incapable of activating reporter expression in stage 14–16 embryos (Figure 8B). A residual weak expression pattern generated by EN was visible in a few maxillary cells during stage 17 (data not shown). In a complementary experiment, we changed the sequence of the module E Dfd binding site (GGCAATTAG)

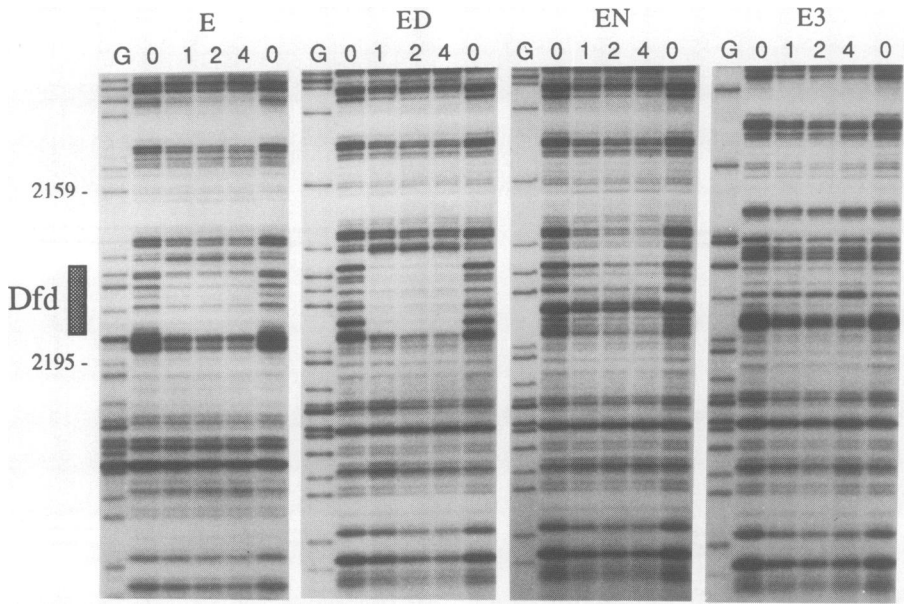

Fig. 6. Dfd protein protects one site in modules E and ED from DNase I digestion. DNase I footprinting assays were performed on the 5′ end-labeled duplex DNA of elements E, ED, EN and E3 using the protocol described in Heberlein *et al.* (1985). Sequences from modules ED, EN and E3 were labeled on the bottom strand (relative to the DNA sequence in Figure 4) at *SalI* polylinker sites in pBluescript (Stratagene) clones pB-ED, pB-EN and pB-E3 (see Materials and methods). Module E was labeled at a nearby *XhoI* site in the same polylinker. Lanes labeled 0 contain no protein, lanes labeled 1, 2 and 4 contain 1, 2 or 4 μl, respectively, of bacterially produced Dfd protein extract (Dessain *et al.*, 1992). 1 μl of extract in a 50 μl reaction provide an approximate concentration of 3×10^{-7} M Dfd protein. Lanes labeled G contain the same restriction fragment cleaved at guanine bases (Maxam and Gilbert, 1977). The G lane of the E footprint contains some non-G cleavages in addition to the normal cleavages. The labeled box on the left indicates the extent of the Dfd footprint. The numbers indicated are the positions of two G residues in the 2.7 kb *Dfd* epidermal autoregulatory sequence (Figure 4).

to CATCATTAA. This site, derived from site D of Regulski *et al.* (1991), preserves only the ATTA core of the DBS and more stably binds Dfd protein *in vitro* (Figures 6 and 7). We found that the resulting construct ED (Figure 11) drives reporter gene expression to a higher level and in 2- to 3-fold more cells in the maxillary segment of stage 14 embryos compared with wild type module E (Figure 8, compare C with A). The correlation between *in vitro* mutations that enhance Dfd binding stability with enhanced activity of the mutant regulatory elements in embryos indicates that the Dfd binding site of module E is likely to directly interact with Dfd protein.

To localize the other module E sequences that might contribute to its function, we generated and analyzed a series of clustered substitutional mutations. We used sequence homology between the *D.melanogaster* module E and an evolutionary variant of module E that resides in upstream sequences of the *D.hydei Dfd* gene to guide these mutations. Functionally important *cis*-regulatory sequences are often evolutionarily conserved (Mitsialis *et al.*, 1987; Treier *et al.*, 1989; Kassis, 1990; Maier *et al.*, 1990; Wu *et al.*, 1990; Moses and Rubin, 1991; Schier and Gehring, 1993b). A 168 bp DNA sequence (fragment H) from ~5 kb upstream of the transcription site of the *Dfd* gene from *D.hydei* has numerous blocks of sequence identity to the 120 bp module E. There are apparently two insertions in the *D.hydei* fragment H relative to module E, but otherwise the two sequences are 65% identical (Figure 11). Fragment H was tested for a regulatory function in a manner similar to that of module E in transgenic *D.melanogaster* embryos. We found that it activates β-gal expression during stage 16 in cells of the maxillary segment (data not shown), indicating that fragment H is a weaker regulatory element in comparison with module E in *D.melanogaster* embryos. Despite the difference in strength, we surmised that the sequence similarities might still be useful in guiding our clustered point mutations. Accordingly, we divided module E into regions 1–6, each corresponding to a relatively continuous stretch of conserved sequence between *melanogaster* and *hydei*. Each of these six non-overlapping regions were individually substituted by random sequences in mutant reporter constructs E1–E6 respectively (Figure 11).

Many previous studies have noted that reporter gene expression can often be increased significantly by multimerizing regulatory sequences without having much influence on functional specificity (Driever *et al.*, 1989; Jiang *et al.*, 1991; Topol *et al.*, 1991; Hoch *et al.*, 1992; Jiang and Levine, 1993). This also applies to reporter construct 4XE, which contains a tandem repeat of four copies of module E in the same orientation. 4XE activates β-gal in cells of the maxillary segment from stage 13 until the end of embryogenesis (Figures 9A and 10C and F). In a *Dfd* mutant background, 4XE function is mostly abolished (Figure 9B). In a *hs-Dfd* genetic background after heat shock, ectopic 4XE activity was detected in the posterior part of the labial segment, in a procephalon segment and weakly in the T1 and T2 segments (Figure 9C). We concluded that the multimerization of module E has not altered its nature as a *Dfd* responsive element. Therefore, the clustered point mutations of module E were all tested as 4X multimers.

4XE1 activates β-gal expression at the beginning of stage 13 in the posterior compartment of the maxillary segment, similar to parental 4XE. However, a small group of cells localized near the center of the maxillary segment no longer activate reporter gene expression in 4XE1 when compared with 4XE (Figure 9D). 4XE2 activity is similar to wild type

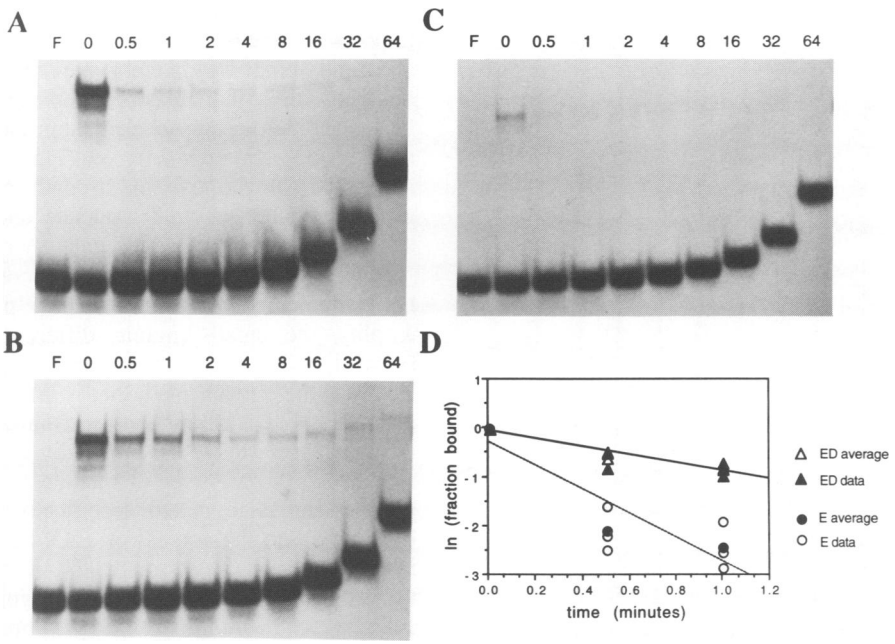

Fig. 7. Comparative dissociation rates of Dfd protein complexed with E, ED and EN binding sites. Gel mobility shift assays to measure dissociation rates were carried out as outlined in Materials and methods. (**A**) Mobility shift complex between the Dfd binding region of module E and Dfd protein, and its dissociation with time. The mobility shift probe (5′ end-labeled with T4 polynucleotide kinase) was a 34 bp fragment extending from base 2173 to 2206 (Figure 4), which encompasses the Dfd binding site. 4 μl of a 1:10 dilution of bacterially produced Dfd protein extract (Dessain *et al.*, 1992) were added to a binding reaction (see Materials and methods) and equilibrated on ice for 1 h. The approximate concentration of Dfd protein in the 300 μl reaction was 2×10^{-8} M. At time 0, a 10 000-fold molar excess amount of a specific competitor DNA [a cold 24 bp oligonucleotide that includes the module E Dfd binding region (2171−2194 in Figure 4)] was added to the reaction and aliquots of the reaction were loaded onto a running gel at the times indicated. (**B**) Mobility shift complex between the Dfd binding region of module ED (Figure 11) and Dfd protein, and its dissociation with time. An analogous 34 bp oligomer to that used in (A), but with ED sequence (Figure 10), was used as a mobility shift probe, and all other conditions were as in (A). (**C**) Mobility shift complex between a region of module EN and Dfd protein, and its dissociation with time. An analogous 34 bp oligomer to that used in (A), but with EN sequence (Figure 11), was used as a mobility shift probe. Because module EN lacks a detectable Dfd binding site by footprinting, 10 times as much Dfd protein extract was added. Despite this, little of the labeled EN probe was detectable in a mobility shift complex, and the EN−Dfd protein kinetic dissociation rate was too rapid to measure under our conditions. (**D**) A graph of data gathered from three separate experiments testing site E contrasted with three separate experiments on site ED. Data points represent the natural log of the fraction of probe bound with Dfd protein, where the fraction bound at time 0 is 1. Averages of the data points were calculated and a best fit line drawn to determine the kinetic dissociation constant (the negative of the slope of line of best fit) of each site complexed with Dfd protein. The kinetic dissociation constant (K_d) for the site ED−Dfd protein complex is 0.8 ± 0.1 min^{-1} and its half-life ~0.8 min. The K_d for the site E−Dfd protein complex is 2.4 ± 0.7 min^{-1} and the half-life ~0.3 min.

4XE (Figure 9E). 4XE3 has an 18 bp substitution mutation in region 3 which contains the DBS (Figure 11). 4XE3 is inactive until stage 15 (Figures 9F and 10D), and its expression is limited to a few cells within the *Dfd* expression domain at reduced levels (Figure 10, compare C with D). The residual maxillary-specific expression of this multimerized construct in the absence of a Dfd binding site may be due to fortuitous Dfd binding sites in this specific construct, perhaps in the linkers or in the rest of the reporter vector. Alternatively, it is possible that in multimerized form, the 4XE3 mutant version of module E has inherent maxillary-specific activity that is not dependent on *Dfd* input. 4XE4 activates β-gal in a pattern indistinguishable from that of parental 4XE (Figure 9G).

The 4XE5 construct has a 28 bp mutant substitution (Figure 11) and is inactive until stage 15 (Figures 9H and 10E). The maxillary expression under 4XE5 control is weak and is limited to only a few cells within the *Dfd* expression domain. In addition, a novel pattern is detected in extreme dorsal anterior cells, where *Dfd* is not expressed (Figure 10, compare C with D). The most 3′ region of module E is mutated in 4XE6 (Figure 11). The 4XE6 construct is completely inactive until the latest stages of embryogenesis, when very weak reporter expression can be detected occasionally in the maxillary region of transgenic embryos (Figures 9I and 10H). This weak, occasional expression seen in the maxillary segment at stage 17 (Figure 10H) is not much stronger than the weak, occasional maxillary staining seen in some embryonic strains transgenic for the host reporter vector lacking any regulatory DNA insert (data not shown). 4XE6 also activates β-gal expression on the dorsal/anterior aspect of embryos in an ectopic, *Dfd*-independent pattern that appears very similar to the ectopic pattern detected in 4XE5.

The results indicate that the DNA sequences in regions 1, 2 or 4 are not functionally important, and that regions 3, 5 and 6 might be sufficient for module E function. To test this, we combined the E1, E2 and E4 substitutional mutations in a single construct ES (Figure 11). We found that 4XES activates β-gal in a pattern similar to that of 4XE1 (compare Figures 10A and B and 9D). We conclude that the 83 bp of unchanged DNA sequences in regions 3, 5 and 6 of ES are sufficient for achieving most of the regulatory activity of the entire module E. To test whether a precise spacing between region 3 and regions 5/6 was required for the activity of module E, a mutation deleting one helical turn

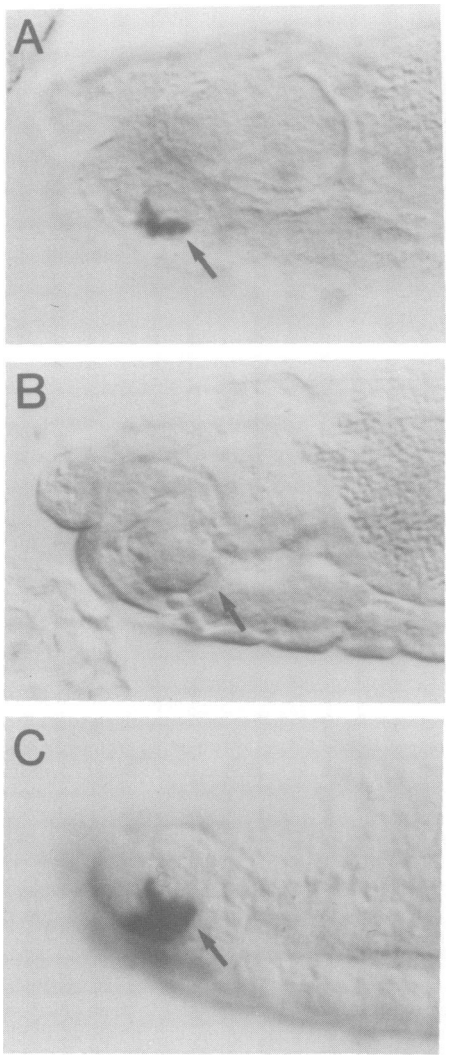

Fig. 8. Dfd binding site (DBS) mutations in module E. Panels show the lateral view of the head region of embryos during head involution (stage 14). Anterior is to the left. Solid arrows indicate the maxillary segment. The β-gal expression patterns were detected by antibody staining of whole-mount embryos. (**A**) The reporter expression conferred by wild type element E. (**B**) Expression is eliminated when the DBS is mutated to a random sequence in construct EN. The *in vitro* stability of Dfd protein binding to EN is dramatically decreased. (**C**) The expression pattern conferred by ED is in more maxillary cells and at a higher level compared with that of E. ED contains mutations in the Dfd binding site, changing it to the sequence of site D (Regulski *et al.*, 1991) and consequently increasing the stability of Dfd protein binding to ED by at least 2-fold (Figures 6 and 7) when compared with the wild type E sequence.

within region 4 was constructed and tested. This internal deletion mutant has a regulatory activity that is indistinguishable from wild type module E (data not shown).

Discussion

A Dfd response element consists of more than Dfd binding sites

By analyzing the reporter gene expression driven by wild type and mutant versions of a *Dfd* response element (module E), we have attempted to define the essential DNA sequences that determine its specific activation by Dfd versus other embryonically expressed homeodomain proteins that could potentially activate it. In module E, one functionally important sequence binds Dfd protein *in vitro* and also apparently binds Dfd in embryonic maxillary cells, because reducing (EN) or increasing (ED) the *in vitro* affinity for Dfd in an otherwise intact element produces parallel changes in regulatory activity in embryos (Figures 8 and 11). Though important for the overall strength of the element, the interaction between this sequence and Dfd alone is not sufficient for a *Dfd*-specific regulatory response in embryos.

It has been proposed that the cooperative binding of homeodomain proteins to multiple 'preferred' sites may amplify the small binding differences among various homeodomains and result in specificity (Ekker *et al.*, 1992). This may be a mechanism that is important in many target elements, but it is certainly not a requirement for Dfd protein targeting specificity, as shown by these experiments on module E. A monomer of module E containing one Dfd binding site is regulated specifically, and an intact Dfd binding site in the context of a 4X multimer of module E in which region 6 was mutated (4XE6) was incapable of activating embryonic expression (Figures 9I and 10H). In addition, an 8X multimer of only 24 bp of module E, which excluded regions 5 and 6, provided an ectopic, *Dfd*-independent expression pattern that is abundant in the labial and at lower levels in other segments (data not shown).

Though mediating an exquisitely specific regulatory function, the functional Dfd binding site that we identified in element E, GGCAATTAG, does not have a particularly good match with a high-affinity consensus Dfd site, resembling a Ubx optimal site more than the Dfd optimal site (Ekker *et al.*, 1991, 1992). The best match between the module E Dfd binding site is actually with consensus sequences (i.e. GCAATTA) for high-affinity ftz protein binding sites (Pick *et al.*, 1990; Florence *et al.*, 1991). In the context of module E, this site may or may not be bound in embryonic cells by Ubx, ftz or other homeodomain proteins, but is apparently incapable of being activated by any of them.

Regions 5 and 6, together encompassing 57 bp, appear to be at least as important to the function of the module E *Dfd* response element as the binding sites for Dfd itself. These essential sequences have no detectable specific binding affinity for Dfd protein *in vitro* and are likely to contain one or more binding sites for proteins other than Dfd, which contribute to the overall strength of the element as well as its property of being specifically activated by Dfd. Mutating either region 5 or 6 has similar (4XE5) or even more severe (4XE6) effects on the function of the module than the elimination of the Dfd binding site (4XE3) (Figure 9). Only a combination of a Dfd binding site and the region 5/6, as in constructs E or ES, constitutes a meaningful activation element for Dfd. Examination of the DNA sequences from regions 5 and 6 reveals an imperfect inverted repeat structure (13/16 base match) centered near the junction between regions 5 and 6 (Figure 11), which may represent a homo- or heterodimer binding site. No good matches were found between the region 5/6 sequence and the conserved sequence elements and known nuclear factor binding sites in the *ftz* autoregulatory enhancer (Han *et al.*, 1993; Schier and Gehring, 1993b). The isolation and characterization of *Drosophila* nuclear proteins that bind regions 5 and 6 is in progress.

Although module F has not been subjected to as detailed

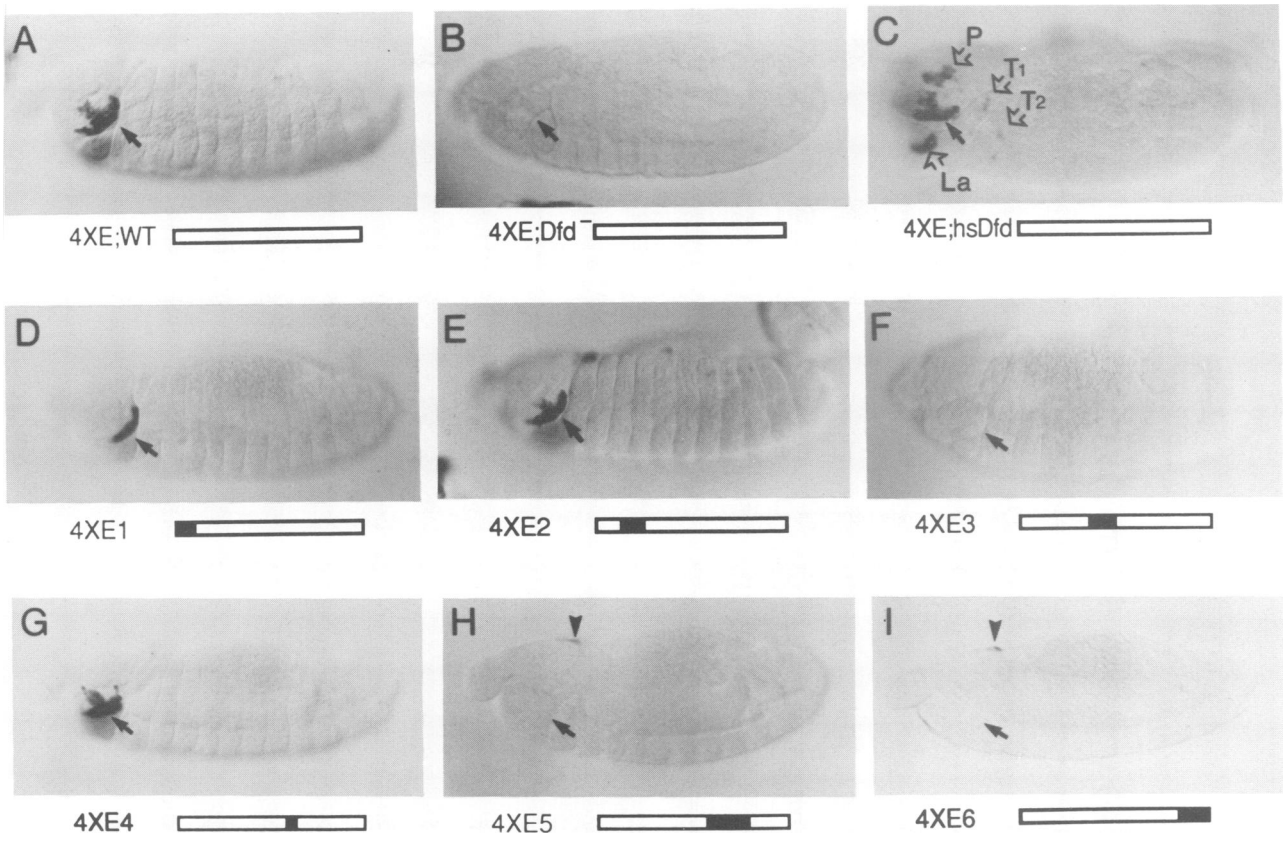

Fig. 9. Clustered point mutations identify other functional sequences in module E. Panels show lateral views of the β-gal expression patterns detected by antibody staining of whole-mount embryos at stage 13. Anterior is to the left. Solid arrows indicate the posterior border of the maxillary segment. Under each embryonic expression pattern panel, the corresponding mutation is indicated by a shadowed portion on a open bar representing the wild type module E. (**A**) Expression provided by wild type module E tandemly repeated to four copies (4XE). (**B**) The same construct 4XE as in (A) but in *Dfd*⁻ background. The expression pattern is absent. Embryos homozygous for *Dfd* mutations are identified by the lack of global staining associated with the balancer chromosomes (see Materials and methods). (**C**) A stage 14 embryo transgenic for both 4XE and *hs-Dfd*. Ubiquitous *Dfd* expression was induced transiently by heat shock at cellular blastoderm stage, which induces persistent *Dfd* expression in the posterior of most segments (Kuziora and McGinnis, 1988). 4XE is also ectopically activated, as indicated by open arrows, in the posterior of the labial segment (La), a procephalon segment (P) and weakly in posterior of the first and second thoracic segments (T1 and T2). The pattern of ectopic β-gal expression is presumably due in large part to this module being activated by the localized, persistent ectopic Dfd protein expression that occurs in *hs-Dfd* embryos (Kuziora and McGinnis, 1988). (**D**) The expression pattern generated by 4XE1 is similar to that of 4XE as in (A), except that the more anterior cells lose β-gal expression in 4XE1. (**E**) The expression pattern generated by 4XE2 is very similar to that of 4XE as in (A). (**F**) The β-gal expression is gone at stage 13 when region 3, which contains the Dfd binding site (DBS), is mutated in 4XE3. (**G**) The expression pattern generated by 4XE4 is very similar to that of 4XE, as in (A). (**H**) The β-gal expression under the control of 4XE5 is absent in the maxillary segment at stage 13. (**I**) The β-gal expression under the control of 4XE6 is absent in the maxillary segment at stage 13. A novel expression pattern (solid arrow heads) in anterior dorsal cells of both E5 and E6 embryos is shown in (H) and (I).

a mutagenic dissection as module E, the regulatory sequences that make contributions to its activity appear to be distributed broadly throughout its length. The 471 bp module F was progressively deleted from both ends (Figure 4). Except for a few initial deletion steps (F1, F2 and F7−F9), all other deletions from either direction gradually weaken the element (Figure 5). For example, from F2 to F5, every ~60 bp deletion from the 5′ end delays reporter expression in the maxillary segment by an additional ~2 h (Figure 5). It seems that the full activity of module F depends on the summation of several DNA sequences located at various regions of the element. For example, the 254 bp F5, 156 bp F11 and 77 bp M are three mostly non-overlapping sections of the parental module F (Figure 4). Each region alone generates similar weak and late expression patterns (stages 16 or 17) (Figure 5D, F and H). In fact, module F contains no more DNA sequences than a combination of F11, M and F5 (Figure 4), but provides much more abundant expression and is activated at a much earlier stage than F11, M or F5 (stage 12, ~6−7 h earlier). These regions seem to act additively or synergistically in module F.

Footprinting of module F (data not shown) detects six Dfd protein binding sites and a region that is moderately protected (Figure 4). Two of the strong footprinted sequences, GTAATGGGT and GTAATTATT, coincide with regions contributing to the strength of the constructs *in vivo*, as revealed by comparing F2 with F3 (or F11 with F12) and F4 with F5 (or F10 with F11) respectively (Figures 4 and 5). But some deletion steps that severely diminish activity do not remove Dfd binding sites; for example, F4 is significantly weaker than F3, F5 weaker than F6, and F9 weaker than F10 in activating β-gal expression (Figures 4 and 5). Conversely, the inclusion of Dfd binding sites is not sufficient for derivatives of module F to be activated in embryos. Constructs F6 and F12 each contain two strong Dfd footprinting sites, but both are inactive in embryos (Figures 4 and 5).

We also note that each of the successive deletion mutations

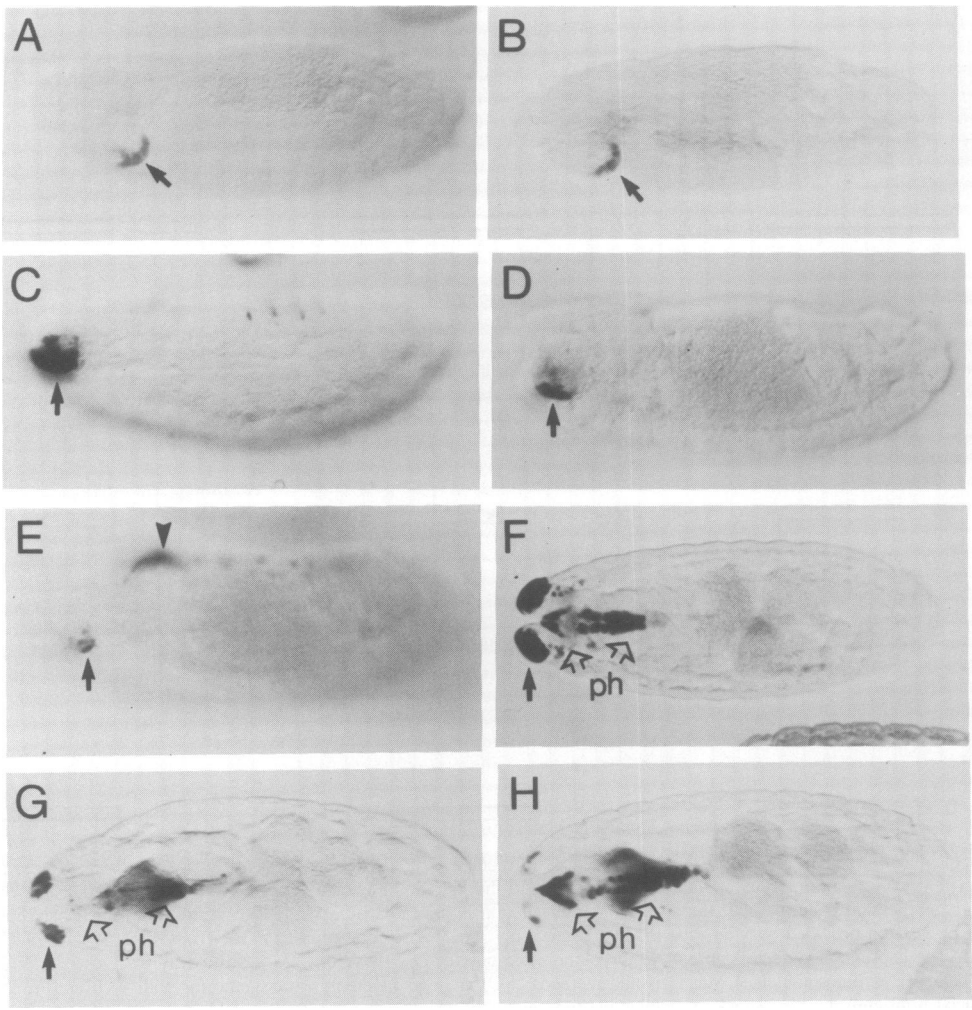

Fig. 10. The Dfd binding site (DBS)-containing region and the putative cofactor binding sites are both necessary and in combination sufficient for much of the regulatory activity of module E. Panels show the β-gal expression patterns detected by antibody staining of whole-mount embryos. Anterior is to the left. Solid arrows indicate the maxillary segment. Open arrows indicate a *Dfd*-independent expression pattern in the pharynx (ph) that is produced by these regulatory elements at very late embryonic stages (stage 17). (**A**, **C** and **F**) The expression patterns generated by the parental construct 4XE. Lateral view at stage 12 (A) and 14 (C). Ventral view at stage 17 (F). (**B**) Lateral view of the expression pattern generated by construct 4XE5, which contains mutations in regions 1, 2 and 4, at the same stage as in (A). (**D**) Lateral view of the expression pattern generated by 4XE3, which contains mutations in region 3 including the Dfd binding site, at the same stage as in (C). (**E** and **G**) The expression pattern generated by 4XE5. (E) Lateral view at the same stage as in (C). The novel dorsal expression pattern is indicated by a solid arrow head (E). (G) Ventral view at the same stage as in (F). The pharynx (ph) expression pattern is indicated by open arrows. (**H**) Ventral view of the expression pattern generated by 4XE6 at the same stage as in (F). The maxillary expression pattern is almost totally gone, but the *Dfd*-independent pharynx (ph) expression pattern (open arrows) is unaffected.

that have an effect on module F function results in its activity being restricted to successively fewer cells within the Dfd expression domain. We conclude that this module is composed mainly of multiple binding sites of activators and their facilitating cofactors, or that repressor sites are so closely linked to activator sites that they have not been independently identified.

From a wider perspective, the entire 2.7 kb *Dfd* EAE is composed of several smaller *Dfd*-dependent elements, analogous to the composite nature of module F though on a larger scale. Although these smaller elements can function independently of each other, they generate expression patterns that have substantial overlap between each other and probably quantitatively assist each other in their natural context. For example, both elements C and F can drive high-level expression in the posterior maxillary segment (Figure 3C and G). This kind of functional overlap seems to be a common feature of eukaryotic enhancers [reviewed in Dynan (1989), Courey and Tjian (1993) and Yamamoto *et al.* (1993)], including those controlling important *Drosophila* patterning genes (Dearolf *et al.*, 1989; Pick *et al.*, 1990; Hoch *et al.*, 1991; Pan *et al.*, 1991; Laney and Biggin, 1992; Qian *et al.*, 1993; Schier and Gehring, 1993b), and may be used to increase the stability of the transcriptionally active state of developmentally important genes and/or for fine-tuning expression patterns.

Our conclusions about the nature of *Dfd* response elements have many similarities with those drawn from the analyses of *ftz* and *eve* autoregulatory elements (Hiromi and Gehring, 1987; Harding *et al.*, 1989; Pick *et al.*, 1990; Jiang *et al.*, 1991; Schier and Gehring, 1992, 1993a,b). For both *ftz* and *eve* elements, deletions which do not remove either ftz or eve protein binding sites (defined *in vitro*) nevertheless severely attenuate the embryonic function of the elements

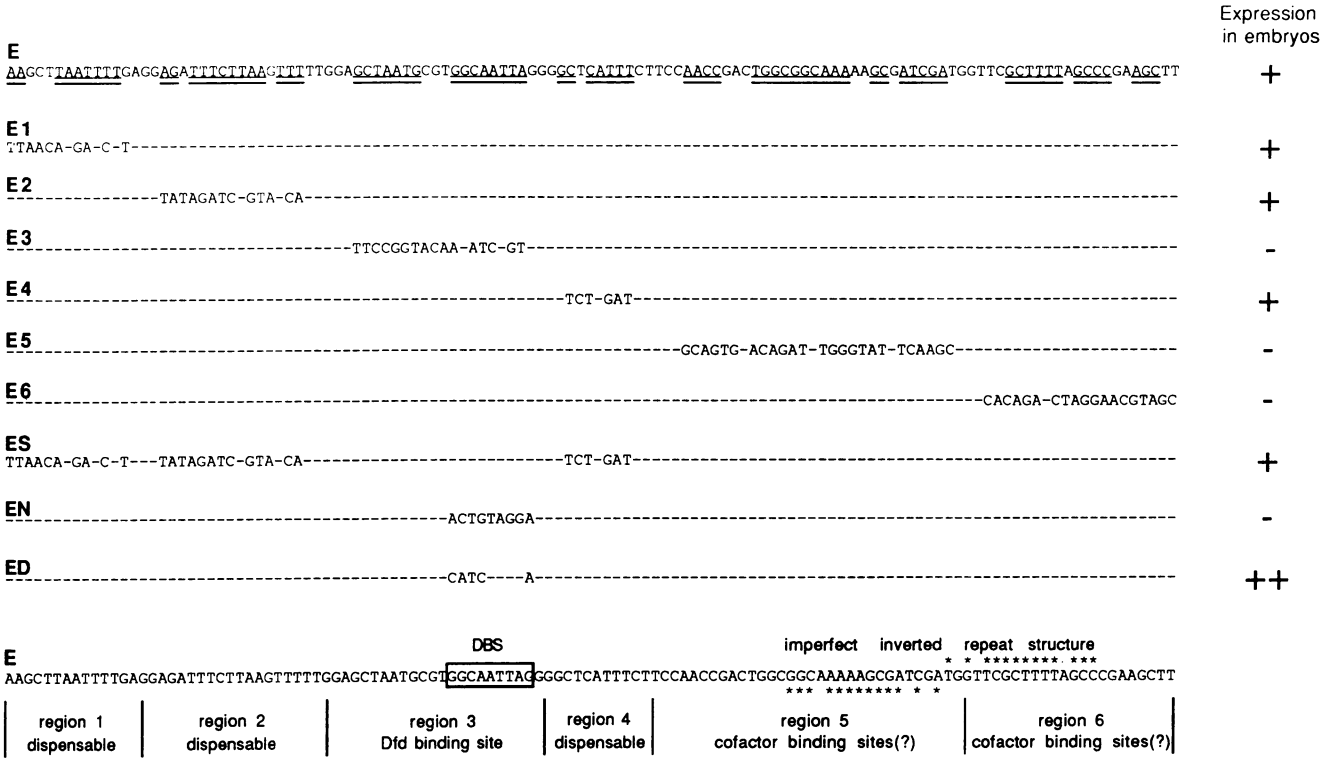

Fig. 11. The sequence and functional organization of module E. (**Top**) The wild type module E sequence with double underlines beneath the nucleotides that are conserved in fragment H, an apparent *D.hydei* version of module E of *D.melanogaster* (the entire sequence of fragment H is available on request). (**Middle**) The clustered point mutations and the mutations of the DBS are shown. Mutations are identified by letters denoting the substituted DNA. Dashes indicate unchanged nucleotides. The presence or absence of strong expression in the maxillary segment of mid-stage embryos is indicated by + or − to the right of the DNA sequences. ++ indicates an expression pattern that is stronger than that generated by the wild type element. (**Bottom**) The wild type DNA sequence of module E is shown again with the DBS enclosed in an open box and with the matched nucleotides of the imperfect inverted repeated structure highlighted by asterisks. Underneath the sequence is our current interpretation of the data.

(Jiang *et al.*, 1991; Schier and Gehring, 1992, 1993b). Although lacking a HOM-type function, the ftz protein contains a homeodomain very similar to HOM proteins [reviewed in Laughon (1991) and Dessain and McGinnis (1993)]. The 2.6 kb *ftz* autoregulatory element (the 'upstream' region) contains subelements with partially overlapping functions (the 'distal' and the 'proximal' elements) (Pick *et al.*, 1990). A systematic deletion analysis of a 429 bp subelement called AE revealed the existence of multiple sequences that are combined to generate enhancer activity. The effect exerted by sequentially deleting AE is the successive reduction of reporter expression (Han *et al.*, 1993; Schier and Gehring, 1993b). Evidence that the integration of multiple sequence elements is required to achieve a functional regulatory effect has also been observed in enhancers known or suspected to be under the control of other proteins in the homeodomain superfamily (Struhl *et al.*, 1989; Xue *et al.*, 1992, 1993; Appel and Sakonju, 1993; Lipkin *et al.*, 1993; Thüringer and Bienz, 1993).

Cofactors and the functional specificity of HOM proteins

One of the best understood transcriptional regulators is yeast protein GAL4. GAL4 binds specifically to a 17mer UAS_G, which serves as a chromosomal 'zip code' for GAL4 to tether the transcriptional activation domain of GAL4 to its target regulatory elements (Ptashne, 1988). GAL4 expressed in *Drosophila* from transgenic constructs activates UAS_G reporter constructs in patterns that closely resemble the GAL4 protein distribution (Fischer *et al.*, 1988; Brand and Perrimon, 1993), indicating that its target specificity is largely attributable to its high binding affinity for UAS_G sites. Additional proteins, such as yeast SWI1, SWI2 and SWI3 (Peterson and Herskowitz, 1992), probably participate in the GAL4-dependent activation to generally increase the level of transcription without affecting its targeting specificity. The putative cofactor proteins that act through the module E region 5/6 sequences along with Dfd are probably not such general activators. As evidence for this, the 4XE regulatory construct is capable of activating reporter expression in a *Dfd*-independent fashion in cells lining the pharynx of late stage embryos (Figure 10F). This ectopic expression pattern is still intact in 4XE6 embryos. In contrast, the maxillary expression in 4XE6 embryos is reduced to background levels (Figure 10H). Our interpretation of these observations is that the region 5/6 cofactors are able to exhibit some specificity in their combinatorial interaction, whether direct or indirect, with the Dfd protein.

Our results provide evidence that is consistent with the idea that HOM proteins require specific cofactors for their targeting specificity. Previous evidence supporting this notion has come principally from homeodomain swap experiments on HOM proteins, which has indicated that the target specificity of the homeotic selectors is encoded mainly in the homeodomain and nearby amino acid sequence [Kuziora and McGinnis, 1989, 1991; Chan and Mann, 1993; Furukubo-Tokunaga *et al.*, 1993; Zeng *et al.*, 1993; reviewed in Gross and McGinnis (1994)]. Since many of

the important discriminatory residues identified in these studies are known or suspected to have little influence on the bimolecular interaction between a HOM protein and a single DNA binding site, the supposition has been that the discriminatory residues are interacting with other protein cofactors. The target specificity of other transcriptional regulatory elements is dependent on the input of multiple *trans*-acting regulatory proteins [for example, reviewed in Courey and Tjian (1993), Johnson (1993) and Yamamoto *et al.* (1993)]. And in some of these cases the important amino acid residues are within or near the DNA binding domain of the respective *trans*-regulatory proteins (Johnson, 1993; Yamamoto *et al.*, 1993).

There are many possible mechanisms through which cofactors might contribute to the target specificity of HOM proteins. For Dfd, one possibility is that Dfd protein and its cognate cofactors cooperatively bind close to each other in the autoregulatory element or in *Dfd* responsive elements of downstream genes. The juxtaposition of the Dfd binding sites (in region 3, for example) and the cofactor binding sites (region 5/6, for example) might stabilize otherwise weak (but specific) interactions between the cofactors and Dfd protein, which would not occur off the DNA. Dfd and its cognate cofactors might also be able to selectively recruit each other to their respective binding sites. Dfd proteins, the cofactors and the specific DNA sequences of a regulatory element could each contribute a part of the interaction specificity, only a mutually compatible combination of the three resulting in a functional transcriptional regulatory complex.

The second mechanism postulates that cofactors or modifiers which do not bind DNA sequences themselves can modulate the HOM proteins to bind a DNA sequence that is not, or is only weakly, bound by HOM proteins alone. In this view, it is possible that region 5/6 does actually bind functional variants of Dfd that are not produced in *E.coli*. Although we cannot exclude this possibility, the lack of any detectable *in vitro* interactions between Dfd protein and region 5/6 sequences causes us to prefer the first mechanism as the most likely possibility.

Our model is most similar to the cell type-specific transcriptional regulation of a- and α-specific gene expression in the yeast *Saccharomyces cerevisiae* involving two homeodomain proteins a1 and α2, and the DNA binding protein MCM1. For example, the α2 protein binds to DNA sites with modest specificity and affinity *in vitro* but selectively regulates a small number of target genes (asg operators) *in vivo* (Johnson, 1993). This is in part due to the ability of α2 and MCM1 proteins to bind asg operators cooperatively. The mutation of either the MCM1 binding sites or the α2 binding sites results in abnormal operator function (Johnson and Herskowitz, 1985; Smith and Johnson, 1992), a result that is superficially similar to our results with element E. What differs in module E is that a precise spacing between the homeodomain and putative cofactor binding sequences is not required; this could be due to the possibility that Dfd protein and putative cofactors binding region 5/6 do not contact each other directly, but instead a third kind of relatively flexible 'bridging' proteins mediates the contact. Another similar example of combinatorial control is seen in the POU homeodomain proteins Oct-1 and Oct-2, which can bind the same octanucleotide site *in vitro*. However, by specifically interacting with VP16, Oct-1 (but not Oct-2) gains the ability to regulate a specific target element (the OCTA$^-$ version of TAATGARAT motif). When the Oct-2 homeodomain is replaced by the homeodomain of Oct-1, the chimeric Oct-2/Oct-1 gains the ability to complex with VP16, indicating that VP16 can discriminate between the two proteins based on the identity of their homeodomains (Stern *et al.*, 1989; Stern and Herr, 1991; Pomerantz *et al.*, 1992; Cleary *et al.*, 1993).

In comparison with GAL4, combinatorial regulation involving cofactors may be seen as a relatively inelegant way to achieve HOM protein targeting specificity. It is theoretically possible that an amplification of different *in vitro* binding preferences among HOM proteins could achieve target specificity without cofactors (Ekker *et al.*, 1991, 1992). However, combinatorial control has the advantage of flexibility. One HOM protein could have several cofactors, some of which might be expressed or modified in certain spatial/temporal patterns. It would allow the cell to generate a large number of expression patterns using combinations of a much smaller number of proteins (Dynan, 1989). By recruiting a particular cofactor, a regulatory element could channel additional morphogenetic cues to the HOM target genes. For example, *Dll*, a target gene of *Dfd*, is only expressed in ~20 ventral−lateral cells within the maxillary segment, although Dfd protein is expressed in many more maxillary epidermal cells (O'Hara *et al.*, 1993). It could be that cofactor activity that is required for the function of the *Dfd* response element of the *Dll* gene is present only in those cells within the maxillary segment. The different expression patterns generated by several *Dfd* EAE modules also seem likely to be attributable to differential contributions from cofactors, since the expression levels of Dfd protein do not dramatically differ in the posterior head expression domain.

In the combinatorial control of yeast cell type-specific genes, homeodomain protein α2 and cofactor MCM1 are both needed to regulate the asg operators, which contain both α2 and MCM1 binding sites. By invoking a different cofactor, namely a1, α2 is able to regulate a different set of target genes (hsg operators) which contain essentially the same α2 binding sites as those found in the asp operators (Goutte and Johnson, 1993; Mak and Johnson, 1993). Although MCM1 protein is also present, its interaction with α2 is not facilitated by the hsg operators which lack the MCM1 binding sites. Similarly, in the regulatory elements of the HOM genes, a constellation of HOM protein binding sites and specific cofactor binding sites could be used to subdivide a body segment into finer structures by a network of transcription factors.

Currently, no cofactors are known for the *Drosophila* HOM proteins. The highly diverged homeodomain protein encoded by the *extradenticle* (*exd*) gene is the best current candidate (Flegel *et al.*, 1993; Rauskolb *et al.*, 1993), but it seems likely to be only one of many. In *exd* mutants homeotic transformations are observed, although many HOM genes are expressed normally (Peifer and Wieschaus, 1990). It is hypothesized that some HOM proteins activate different target elements dependent on *exd* function (Peifer and Wieschaus, 1990). Module E contains no matches with high-affinity binding sites defined for a human exd protein homolog (*pbx1*) that is particularly highly conserved relative to exd protein in the homeodomain and surrounding region (Van Dijk *et al.*, 1993). In addition, module E is not detectably footprinted *in vitro* by human pbx1 protein.

Interestingly, the binding affinity of Dfd protein for the module E Dfd binding site is significantly enhanced by this human *exd*-like protein, but the biological significance of this finding awaits more experiments (J.Pinsonneault and W.McGinnis, unpublished results).

Autoregulation is the principal mechanism for maintaining embryonic Dfd expression

Temperature shift assay on conditional mutations and mosaic analysis indicated that expression of the homeotic selector genes is required throughout most of development for proper morphogenesis (Lewis, 1964; Struhl, 1982; Merrill *et al.*, 1987). Because the activities of early acting genes involved in initiation are transient, different mechanisms must be responsible for maintaining the expression patterns. The *Polycomb* group genes have been proposed to 'lock' the homeotic selector genes in repressed states outside their expression domains (Duncan and Lewis, 1982; Wedeen *et al.*, 1986; Zink *et al.*, 1991). On the other hand, the *trithorax* group genes are believed to be capable of positively maintaining the transcribed states of many HOM genes in their normal domains (Breen and Harte, 1993; Kennison, 1993). However, the epidermal expression of *Dfd* is only slightly affected in *trithorax*$^-$ embryos (Breen and Harte, 1993) and the effect of mutations in other *trithorax* group genes is not known at present. Previous research has shown that *Dfd* is capable of activating its own expression in ectopic locations and in its own expression domain (Kuziora and McGinnis, 1988; Bergson and McGinnis, 1990; Regulski *et al.*, 1991).

The results here (Figure 1) indicate that positive autoregulation of *Dfd* is the major mechanism for maintaining its own expression in epidermal cells. It has long been speculated that the homeotic genes can use autoregulation for transducing short-lived signals into stable expression patterns during development (García-Bellido and Capdevila, 1978; Serfling, 1989, Jiang *et al.*, 1991). In fact, a total of five HOM genes, *lab*, *Dfd*, *Antp*, *Ubx* and *Abd-B*, are now known to autoregulate their own transcription in one or more cell types (Bienz and Tremml, 1988; Kuziora and McGinnis, 1991; Choinard and Kaufman, 1992; Christen and Bienz, 1992; Lamka *et al.*, 1992; Tremml and Bienz, 1992; Appel and Sakonju, 1993). In *Dfd*^{w21} embryos, some residual transcription still occurs in a few maxillary epidermal cells and in most of the subesophageal ganglions (sbg) (Figure 1). The remaining *Dfd*^{w21} expression may be due to the action of general homeotic regulators such as *trithorax* group genes.

Materials and methods

DNase I footprinting

DNase I footprinting experiments were carried out as described in Heberlein *et al.* (1985), except that 3 μl of a 1/4000 dilution of stock DNase (10 units/ml) was used. Dfd protein was produced in *E.coli* and purified according to Dessain *et al.* (1992).

Dissociation rate measurements

Reactions were carried out under the following conditions: 25 mM HEPES pH 7.6, 50 mM KCl, 0.1 mM EDTA, 1 mM DTT, 10% glycerol, 5 mM $MgCl_2$ and 10 mg/ml carrier DNA. A 300 μl reaction (enough for 12 samples) was set up in one tube. One 25 μl sample (free probe) was removed before protein was added. Reaction was incubated with Dfd protein on ice for 1 h. At time zero, free probe and another 25 μl sample were loaded on a native 4% polyacrylamide gel running at 200 V at 4°C. Unlabeled specific competitor DNA was added and time points were taken at 0.5, 1, 2, 4, 8, 16, 32 and 64 min. The gel was run for an additional 2 h. Radioactivity in the gels was quantified either by a Molecular Dynamics PhosphorImager or by cutting out the bands and using a scintillation counter. Kinetic dissociation constants were determined by plotting ln(fraction of bound DNA) as a function of time where ln(probe bound at time t/probe bound at time 0) = $-K_d t$. Half-lives were calculated using the formula $t_{1/2} = \ln 2/K_d$.

DNA sequencing

The DNA sequences of the 2.7 kb *Dfd* EAE, the *D.hydei* version of module E (fragment H) and the module F and E mutant regulatory constructs were determined by sequencing both strands by the dideoxy-termination method using the Sequenase kit (US Biochemical).

Regulatory constructs

Module A. The 294 bp *Xba*I−*Nsi*I fragment was subcloned into the *Xba*I and *Pst*I sites of pBluescript (Stratagene) to create pB-A. The smaller *Xba*I−*Xho*I fragment of pB-A was then cloned into the *Xba*I and *Xho*I sites of the reporter vector pCaspeR hs43 LacZ III, a modified version of pCaspeR hs43 LacZ (Thummel and Pirrotta, 1991).

Module B. The 370 bp *Nsi*I−*Sma*I fragment was subcloned into the *Pst*I and *Hinc*II sites of pBluescript to create pB-B. The smaller *Xba*I−*Xho*I fragment of pB-B was then cloned into the *Xba*I and *Xho*I sites of pCaspeR hs43 LacZ III.

Module C. The 570 bp *Sma*I−*Hin*dIII fragment was subcloned into the *Sma*I and *Hin*dIII sites of pBluescript to create pB-C. The smaller *Xba*I−*Xho*I fragment of pB-B was then cloned into the *Xba*I and *Xho*I sites of pCaspeR hs43 LacZ III.

Module D. Module D was reported in Bergson and McGinnis (1990).

Module E. The 120 bp *Hin*dIII fragment was subcloned into the *Hin*dIII site of pBluescript to create pB-E. The smaller *Xba*I−*Xho*I fragment of pB-B was then cloned into the *Xba*I and *Xho*I sites of pCaspeR hs43 LacZ. Element E is in the same orientation relative to the basal promoter in the construct as in the *Dfd* locus.

Module F. The 471 bp *Hin*dIII−*Xba*I fragment was first blunted at the *Xba*I end only and ligated to a *Kpn*I linker, and then subcloned into the *Hin*dIII and *Kpn*I sites of pBluescript to create pB-F. The smaller *Xba*I fragment of pB-F was then cloned into the *Xba*I site of a reporter vector HZ50PL (Hiromi and Gehring, 1987).

Constructs F1−F12. The 471 bp *Hin*dIII−*Xba*I fragment was first blunted and then subcloned into the *Hin*dIII and *Hinc*II sites of pBluescript to create pB-F0. The Exonuclease III double-stranded nested deletion kit (Pharmacia LKB Biotechnology) was used to generate deletions following the manufacturer's instruction. pB-F0 was linearized by *Pst*I (the Exo III-resistant end) and *Hin*dIII (the Exo III-sensitive end), or by *Apa*I (the Exo III-resistant end) and *Xho*I (the Exo III-sensitive end), digested by Exonuclease III for various times, followed by Nuclease S1 digestion and religation, to produce pB-F1 to pB-F12. The end points of the deletions were confirmed by DNA sequencing. The smaller *Xba*I−*Kpn*I fragments of pB-F1 to pB-F12 were then cloned into the *Xba*I and *Kpn*I sites of HZ50PL.

Module M. The *Alu*I−*Kpn*I fragment of pB-F10 containing the desirable sequence was cloned into the *Hinc*II and *Kpn*I sites of pBluescript to create pB-M. The smaller *Xba*I−*Kpn*I fragment of pB-M was then cloned into the *Xba*I and *Kpn*I sites of HZ50PL.

Constructs ED and EN. Six pairs of oligonucleotides representing either the wild type or mutated regions 1−6 were ligated to create the 120 bp ED or EN fragment, which was then cloned into the *Pst*I and *Hin*dIII sites of pBluescript to create pB-ED and pB-EN via the addition of *Pst*I and *Hin*dIII linkers. The smaller *Sal*I−*Bam*HI fragment of pB-ED or pB-EN was then cloned into the *Xho*I and *Bam*HI sites of pCaspeR hs43 LacZ.

Constructs 4XE1 to 4XE6 and 4XES. Single-copied pB-E1 to pB-E6 and pB-ES were created in a similar way as above. The larger *Kpn*I−*Hinc*II fragments were then ligated with the smaller *Sma*I−*Kpn*I fragments of the same plasmid (generated in separate digestions) to create pB-2XE1 to pB-2XE6 and pB-2XES. The two-copied plasmids were then digested by *Bam*HI and *Hinc*II, as well as by *Sma*I and *Sal*I, in two separate reactions. The respective smaller fragments of the two digestions were ligated to the *Xho*I- and *Bam*HI-digested pCaspeR hs43 LacZ in three-way ligation to create the final constructs.

Germline transformations

The reporter constructs (at 500 mg/ml) described above were co-injected with the P-element helper plasmid pπ25.7Δ2-3 (at 100 mg/ml) into ry^{506} or Df(1)w, y w^{67c23} embryos essentially as described in Rubin and Spradling (1982). Numbers of independent inserts for each transgene are: A, 10; B, 11; C, 9; E, 10; F, 6; F1, 8; F2, 8; F3, 5; F4, 9; F5, 12; F6, 6; F7, 7; F8, 9; F9, 4; F10, 5; F11, 13; F12, 10; M, 3; EN, 10; ED, 10; 4XE, 10; 4XE1, 9; 4XE2, 10; 4XE3, 8; 4XE4, 10; 4XE5, 10; 4XE6, 10; and 4XES, 10.

Expression pattern detection

Dfd transcripts were detected by *in situ* hybridization using digoxygenin-labeled RNA probes, essentially as described in Tautz and Pfeifle (1989). The *lacZ* reporter gene expression patterns were detected by immunohistochemical staining of whole-mount embryos, essentially as described in Bergson and McGinnis (1990). Mouse monoclonal anti-β-gal antibody (Promega), goat anti-mouse antibody conjugated to biotin (Jackson Immunoresearch) and the ABC-HRP kit (Vector Laboratory) were used.

Drosophila strains

Fly culture and crosses followed standard procedures. The *hs-Dfd* transgene and the heat shock conditions were as described in Kuziora and McGinnis (1988). Three *Dfd* null mutations, Dfd^{Rx1}, Dfd^{w21} and Dfd^{rR11}, were used to generate *Dfd* mutant backgrounds. Flies homozygous for Dfd^{-} were identified by the absence of a *Ubx*-like *lacZ* pattern produced by the *Ubx-lacZ* transgene [known as 35UZ in Irvine *et al.* (1991)], or by the absence of global β-gal staining produced by the *Actin5C-lacZ* transgene inserted in the balancer chromosomes.

Mutant genomic DNA extraction and sequencing

Embryos for each of Dfd^{w21} *red e/TM6 Tb Hu ca* and Dfd^{rR11} p^{p}*/TM6B Tb Hu e ca* were collected for 3 h and aged 24 h at 25°C. Embryos homozygous for *Dfd* mutations were hand selected based on their phenotypes (lack of mouth hooks; Merrill *et al.*, 1987). The embryos were pulverized and the DNA was purified. Parental chromosomal DNA (*red e* for *w21* and *Ki roe* p^{p} for *rR11*) was purified from third instar larvae.

Synthetic oligos (20–24mer) were made to match both upstream and downstream sequences ~30 bp distant from each the five exons of the *Dfd* transcription unit. Using the oligos as primers, each exon was amplified by PCR (Saiki *et al.*, 1988) using an automated DNA thermal cycler (Perkin-Elmer Cetus).

PCR-amplified products were purified by agarose gel electrophoresis. Asymmetric PCR was used on 20 ng of purified double-stranded PCR products to produce single-stranded templates for sequencing by using a 1:50 molar ratio of primers and 40 cycles of amplification (Gyllensten and Erlich, 1988). Amplification products were passed through a Centricon-30 filter unit (Amicon Inc.) before sequencing. The concentrated single-strand templates (10–30 μl derived empirically for each set of reactions) were sequenced using 10 pmol of the appropriate oligonucleotide present, in limiting concentration in the asymmetric PCR, as the primer. Both strands of each exon for each *Dfd* allele and each parental chromosome were sequenced. The only change in Dfd^{w21} from the parental chromosome coding sequence was in codon 346 (Regulski *et al.*, 1987), which was changed to TGA from TGG. The only change in the Dfd^{rR11} from the parental chromosome coding sequence was in codon 210, which was changed from TAT to TAA.

Acknowledgements

We are very grateful to Cornelius Gross for the *D.hydei* sequence, to Vincenzo Pirrotta for the reporter vector pCaspeR hs43 LacZ, to Walter Gehring for the reporter vector HZ50PL, to Corey Goodman for the balancer chromosomes containing *Actin5C-lacZ* transgene, to Richard Mann for the balancer chromosomes containing 35UZ transgene, to Michael Kuziora for *hs-Dfd* flies and to Claude Desplan for pBl-lacZ. We are also grateful for Alan Garen's generous sharing of incubators and dissecting microscopes and for Joan Steitz's generous sharing of the PhosphorImager. We thank Mark Biggin and the members of the McGinnis laboratory, especially Cornelius Gross, Kate Harding and Hugh Taylor, for comments on the manuscript. This work was supported by a grant from the NIH (HD28315).

References

Affolter,M., Schier,A. and Gehring,W.J. (1990) *Curr. Opin. Cell Biol.*, **2**, 485–495.
Appel,B. and Sakonju,S. (1993) *EMBO J.*, **12**, 1099–1109.
Awgulewitsch,A. and Jacobs,D. (1992) *Nature*, **358**, 341–344.
Bender,W., Akam,M.E., Karch,F., Beachy,P.A., Pfeifer,M., Spierer,P., Lewis,E.B. and Hogness,D.S. (1983) *Science*, **221**, 23–29.
Bergson,C. and McGinnis,W. (1990) *EMBO J.*, **9**, 4287–4297.
Bienz,M. and Tremml,G. (1988) *Nature*, **333**, 576–578.
Brand,A.H. and Perrimon,N. (1993) *Development*, **118**, 401–415.
Breen,T.R. and Harte,P.J. (1993) *Development*, **117**, 119–134.
Campos-Ortega,J.A. and Hartenstein,V. (1985) *The Embryonic Development of Drosophila melanogaster*. Springer-Verlag, Berlin.
Chan,S. and Mann,R.S. (1993) *Genes Dev.*, **7**, 796–811.
Choinard,S. and Kaufman,T.C. (1992) *Development*, **113**, 1267–1280.
Christen,B. and Bienz,M. (1992) *Mech. Dev.*, **39**, 73–80.
Cleary,M.A., Stern,S., Tanaka,M. and Herr,W. (1993) *Genes Dev.*, **7**, 72–83.
Courey,A.J. and Tjian,R. (1993) In McKnight,S.L. and Yamamoto,K.R. (eds), *Mechanisms of Transcriptional Control as Revealed by Studies of Human Transcription Factor Sp1. Transcriptional Regulation.* Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, Vol. 2, pp. 743–770.
Dearolf,C.R., Topol,J. and Parker,C.S. (1989) *Genes Dev.*, **3**, 384–398.
Desplan,C., Theis,J. and O'Farrell,P.H. (1988) *Cell*, **54**, 1081–1090.
Dessain,S. and McGinnis,W. (1993) In Wassarman,P. (ed.), *Drosophila Homeobox Genes. Advances in Developmental Biochemistry*. JAI Press Inc., Greenwich, CT, pp. 1–55.
Dessain,S., Gross,C.T., Kuziora,M. and McGinnis,W. (1992) *EMBO J.*, **11**, 991–1002.
Driever,W., Thoma,G. and Nüsslein-Volhard,C. (1989) *Nature*, **340**, 363–367.
Duboule,D. and Dollé,P. (1989) *EMBO J.*, **8**, 1497–1505.
Duncan,I. and Lewis,E.B. (1982) In Subtelny,S. and Green,P.B. (eds), *Genetic Control of Body Segment Differentiation in Drosophila. Developmental Order: Its Origin and Regulation*. Alan R.Liss, New York, pp. 533–554.
Dynan,W.S. (1989) *Cell*, **58**, 1–4.
Ekker,S.C., Young,K.E., von Kessler,D.P. and Beachy,P.A. (1991) *EMBO J.*, **10**, 1179–1186.
Ekker,S.C., von Kessler,D.P. and Beachy,P.A. (1992) *EMBO J.*, **11**, 4059–4072.
Fischer,J.A., Giniger,E., Maniatis,T. and Ptashne,M. (1988) *Nature*, **332**, 853–856.
Flegel,W.A., Singson,A.W., Margolis,J.S., Bang,A.G., Posakony,J.W. and Murre,C. (1993) *Mech. Dev.*, **41**, 155–161.
Florence,B., Handrow,R. and Laughon,A. (1991) *Mol. Cell. Biol.*, **11**, 3613–3623.
Furukubo-Tokunaga,K., Flister,S. and Gehring,W.J. (1993) *Proc. Natl Acad. Sci. USA*, **90**, 6360–6364.
García-Bellido,A. (1977) *Am. Zool.*, **17**, 613–629.
García-Bellido,A. and Capdevila,M.P. (1978) In Subtelny,S. and Sussex,I.M. (eds), *The Clonal Basis of Development. 36th Symposium of the Society for Developmental Biology*. Academic Press, New York, pp. 3–21.
Gehring,W.J. (1992) *Trends Biochem. Sci.*, **17**, 277–280.
Goutte,C. and Johnson,A.D. (1993) *J. Mol. Biol.*, **233**, 359–371.
Graham,A., Papalopulu,N. and Krumlauf,R. (1989) *Cell*, **57**, 367–378.
Gross,C.T. and McGinnis,W. (1994) In Hames,B.D. and Glover,D.M. (eds), *The Function of Homeodomain Proteins in Drosophila Development. Eukaryotic Gene Transcription-Frontiers in Molecular Biology*. IRL Press, Oxford, UK, in press.
Gyllensten,U.B. and Erlich,H.A. (1988) *Proc. Natl Acad. Sci. USA*, **85**, 7652–7656.
Han,W., Altan,N. and Pick,L. (1993) *Mol. Cell. Biol.*, **13**, 5549–5559.
Harding,K., Hoey,T., Warrior,R. and Levine,M. (1989) *EMBO J.*, **8**, 1205–1212.
Hayashi,S. and Scott,M.P. (1990) *Cell*, **63**, 883–894.
Heberlein,U., England,B. and Tjian,R. (1985) *Cell*, **41**, 965–977.
Hiromi,Y. and Gehring,W.J. (1987) *Cell*, **50**, 963–974.
Hoch,M., Seifert,E. and Jäckle,H. (1991) *EMBO J.*, **8**, 2267–2278.
Hoch,M., Gerwin,N., Taubert,H. and Jäckle,H. (1992) *Science*, **256**, 94–97.
Hoey,T. and Levine,M. (1988) *Nature*, **332**, 858–861.
Irvine,K.D., Helfand,S.L. and Hogness,D.S. (1991) *Development*, **111**, 407–424.
Jack,T. and McGinnis,W. (1990) *EMBO J.*, **9**, 1187–1198.
Jack,T., Regulski,M. and McGinnis,W. (1988) *Genes Dev.*, **2**, 635–651.
Jiang,J. and Levine,M. (1993) *Cell*, **72**, 741–752.
Jiang,J., Hoey,T. and Levine,M. (1991) *Genes Dev.*, **5**, 265–277.
Johnson,A.D. (1993) In McKnight,S.L. and Yamamoto,K.R. (eds), *A Combinational Regulatory Circuit in Budding Yeast. Transcriptional Regulation.* Cold Spring Harbor Laboratory Press, Cold Spring Harbor,

NY, Vol. 2, pp. 975–1006.
Johnson,A. and Herskowitz,I. (1985) *Cell*, **42**, 237–243.
Kassis,J.A. (1990) *Genes Dev.*, **4**, 433–443.
Kaufman,T.C., Lewis,R. and Wakimoto,B. (1980) *Genetics*, **94**, 115,133.
Kaufman,T.C., Seeger,M.A. and Olsen,G. (1990) *Adv. Genet.*, **27**, 309–362.
Kennison,J.A. (1993) *Trends Genet.*, **9**, 75–79.
Krumlauf,R. (1993) *Curr. Opin. Genet. Dev.*, **3**, 621–625.
Kuziora,M.A. and McGinnis,W. (1988) *Cell*, **55**, 477–485.
Kuziora,M.A. and McGinnis,W. (1989) *Cell*, **59**, 563–571.
Kuziora,M.A. and McGinnis,W. (1991) *Mech. Dev.*, **33**, 83–94.
Lamka,M.L., Boulet,A.M. and Sakonju,S. (1992) *Development*, **116**, 841–854.
Laney,J.D. and Biggin,M.D. (1992) *Genes Dev.*, **6**, 1531–1541.
Laughon,A. (1991) *Biochemistry*, **30**, 11357–11367.
Lewis,E.B. (1964) In Locke,M. (ed.), *Genetic Control and Regulation of Developmental Pathways. The Role of Chromosomes in Development.* Academic Press, New York, pp. 231–252.
Lewis,E.B. (1978) *Nature*, **276**, 565–570.
Lipkin,S.M., Naar,A.M., Kalla,K.A., Sack,R.A. and Rosenfeld,M.G. (1993) *Genes Dev.*, **7**, 1674–1687.
Mahaffey,J.W., Jones,D.F., Hickel,J.A. and Griswold,C.M. (1993) *Development*, **118**, 203–214.
Maier,D., Preiss,A. and Powell,J.R. (1990) *EMBO J.*, **9**, 3957–3966.
Mak,A. and Johnson,A.D. (1993) *Genes Dev.*, **7**, 1862–1870.
Maxam,A.M. and Gilbert,W. (1977) *Proc. Natl Acad. Sci. USA*, **74**, 560–564.
McGinnis,W. and Krumlauf,R. (1992) *Cell*, **68**, 283–302.
McGinnis,W., Jack,T., Chadwick,R., Regulski,M., Bergson,C., McGinnis,N. and Kuziora,M.A. (1990) *Adv. Genet.*, **27**, 363–402.
Merrill,V.K.L., Turner,F.R. and Kaufman,T.C. (1987) *Dev. Biol.*, **122**, 379–395.
Mitsialis,S.A., Spoerel,N., Leviten,M. and Kafatos,F.C. (1987) *Proc. Natl Acad. Sci. USA*, **84**, 7987–7991.
Morata,G. (1993) *Curr. Opin. Genet. Dev.*, **3**, 606–614.
Moses,K. and Rubin,G.M. (1991) *Genes Dev.*, **5**, 583–593.
Nelson,H.B. and Laughon,A. (1993) *Roux's Arch. Dev. Biol.*, **202**, 341–354.
O'Hara,E., Cohen,B., Cohen,S.M. and McGinnis,W. (1993) *Development*, **117**, 847–856.
Pan,D., Huang,J.-D. and Courey,A.J. (1991) *Genes Dev.*, **5**, 1892–1901.
Peifer,M. and Wieschaus,E. (1990) *Genes Dev.*, **4**, 1209–1223.
Peterson,C.L. and Herskowitz,I. (1992) *Cell*, **68**, 573–584.
Pick,L., Schier,A., Affolter,M., Schmidt-Glenewinkel,T. and Gehring,W.J. (1990) *Genes Dev.*, **4**, 1224–1239.
Pomerantz,J.L., Kristie,T.M. and Sharp,P.A. (1992) *Genes Dev.*, **6**, 2047–2057.
Ptashne,M. (1988) *Nature*, **335**, 683–689.
Qian,S., Capovilla,M. and Pirrotta,V. (1993) *EMBO J.*, **12**, 3865–3877.
Ramírez-Solis,R., Zheng,H., Whiting,J., Krumlauf,R. and Bradley,A. (1993) *Cell*, **73**, 279–294.
Rauskolb,C., Peifer,M. and Wieschaus,E. (1993) *Cell*, **74**, 1101–1112.
Regulski,M., McGinnis,N., Chadwick,R. and McGinnis,W. (1987) *EMBO J.*, **6**, 767–777.
Regulski,M., Dessain,S., McGinnis,N. and McGinnis,W. (1991) *Genes Dev.*, **5**, 278–286.
Rubin,G.M. and Spradling,A.C. (1982) *Science*, **218**, 348–353.
Saiki,R.K., Gelfand,D.H., Stoffel,S., Scharf,J., Higuchi,R., Horn,G.T., Mullis,K.B. and Erlich,H.A. (1988) *Science*, **239**, 487–491.
Schier,A.F. and Gehring,W.J. (1992) *Nature*, **356**, 804–807.
Schier,A.F. and Gehring,W.J. (1993a) *Proc. Natl Acad. Sci. USA*, **90**, 1450–1454.
Schier,A.F. and Gehring,W.J. (1993b) *EMBO J.*, **12**, 1111–1119.
Scott,M.P., Tamkun,J.W. and Hartzell,I.G.W. (1989) *Biochim. Biophys. Acta*, **989**, 25–48.
Serfling,E. (1989) *Trends Genet.*, **5**, 131–133.
Smith,D.L. and Johnson,A.D. (1992) *Cell*, **68**, 133–142.
Stern,S. and Herr,W. (1991) *Genes Dev.*, **5**, 2555–2566.
Stern,S., Tanaka,M. and Herr,W. (1989) *Nature*, **341**, 624–630.
Struhl,G. (1982) *Proc. Natl Acad. Sci. USA*, **79**, 7380–7384.
Struhl,G., Struhl,K. and MacDonald,P.M. (1989) *Cell*, **57**, 1259–1273.
Tautz,D. and Pfeifle,C. (1989) *Chromosoma*, **98**, 81–85.
Thummel,C.S. and Pirrotta,V. (1991) *Dros. Inf. Newsletter*, **2**.
Thüringer,F. and Bienz,M. (1993) *Proc. Natl Acad. Sci. USA*, **90**, 3899–3903.
Topol,J., Dearolf,C.R., Prakash,K. and Parker,C.S. (1991) *Genes Dev.*, **5**, 855–867.
Treier,M., Pfeifle,C. and Tautz,D. (1989) *EMBO J.*, **8**, 1517–1525.
Tremml,G. and Bienz,M. (1992) *Development*, **116**, 447–456.
Vachon,G., Cohen,B., Pfeifle,C., McGuffin,M.E., Botas,J. and Cohen,S.M. (1992) *Cell*, **71**, 437–450.
Van Dijk,M.A., Voorhoeve,P.M. and Murre,C. (1993) *Proc. Natl Acad. Sci. USA*, **90**, 6061–6065.
Vincent,J.P., Kassis,J.A. and O'Farrell,P.H. (1990) *EMBO J.*, **9**, 2573–2578.
Wang,B.B., Muller-Immergluck,M.M., Austin,J., Robinson,N.T., Chisholm,A. and Kenyon,C. (1993) *Cell*, **74**, 29–42.
Wedeen,C., Harding,K. and Levine,M. (1986) *Cell*, **44**, 739–748.
Wu,C.Y., Mote,J.J. and Brennan,M.D. (1990) *Genetics*, **125**, 599–610.
Xue,D., Finney,M., Ruvkun,G. and Chalfie,M. (1992) *EMBO J.*, **11**, 4969–4979.
Xue,D., Tu,Y. and Chalfie,M. (1993) *Science*, **261**, 1324–1328.
Yamamoto,K.R., Pearce,D., Thomas,J. and Miner,J.N. (1993) In McKnight,S.L. and Yamamoto,K.R. (eds), *Combinatorial Regulation at a Mammalian Composite Response Element. Transcriptional Regulation.* Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, Vol. 2, pp. 1169–1192.
Zeng,W., Andrew,D.J., Mathies,L.D., Horner,M.A. and Scott,M.P. (1993) *Development*, **118**, 339–352.
Zink,B., Engstrom,Y., Gehring,W.J. and Paro,R. (1991) *EMBO J.*, **10**, 153–162.

Received on February 17, 1994