Identification of the binding sites for potential regulatory proteins in the upstream enhancer element of the *Drosophila fushi tarazu* gene

Stephen D. Harrison and Andrew A. Travers

MRC Laboratory of Molecular Biology, Hills Road, Cambridge, CB2 2QH, UK

Received October 26, 1988; Accepted November 18, 1988

ABSTRACT

With a view to identifying proteins that regulate the expression of the *Drosophila fzf* gene we have sequenced its enhancer-like upstream element (USE) and determined the binding sites for embryonic nuclear proteins within this region by *in vitro* DNAaseI footprinting. We find that greater than 50% of this element is bound by nuclear protein. By footprinting and gel-retardation studies in embryonic extracts from different developmental stages, we have characterised a number of USE/protein complexes whose nature alters in concert with changes in the *fzf* expression pattern, suggesting that these USE-binding proteins may be involved in the regulation of gene activity. In some cases this suggestion is substantiated by the observation that the protected DNA sequences show homology to the binding sites for *fzf* regulating DNA-binding proteins such as the pair-rule gene product *even-skipped*.

INTRODUCTION

Genes that control the process of segmentation in *Drosophila melanogaster* have been identified by screening for mutations that give rise to recognisable developmental phenotypes (1). In practice, however, such screens are unlikely to identify segmentation genes whose activity is also required for the expression of essential genes of more general function. An alternative approach -pioneered by Shore et al(2) in their analysis of the yeast mating-type silencer- is first to characterise biochemically the proteins which interact with the regulatory region of a particular gene and then identify the genes which encode these proteins. This strategy has the added advantage that it identifies genes whose products directly target DNA sequences involved in the regulation of gene expression. In the study of *Drosophila* segmentation an analysis of this sort involves the identification of proteins which bind within the cis-acting DNA control regions of an important segmentation gene. Genetic screens have shown that segmentation in *Drosophila* is controlled by both maternally and zygotically expressed genes and that the zygotic genes fall into three classes based on the phenotypes of mutant embryos i.e the gap, pair-rule, and segment polarity genes (1). These genes operate in a hierarchy in which the gap genes, responding to maternally derived signals (3;4), are expressed in broad domains throughout the embryo and control the more localised expression of subsequent genes including those of the pair-rule class. The pair-rule genes show expression in alternating segment-width stripes along the anterior/posterior axis of the embryo at the blastoderm stage of development and, with the gap genes, are responsible for the localisation of expression of the segment polarity and homeotic classes of developmental genes, thereby acting to establish cell fates. (5;6;7 for reviews).

We have chosen to identify genes involved in controlling the expression pattern of the pair-rule gene *fushi-tarazu* (*fzf*), which encodes a homeo-domain containing protein (8).

Early in development *ftz* is expressed in even numbered parasegments (9) and later, after this initial expression has decayed, in all parasegments of the developing central nervous system.(10; 11). The control of the temporal and spatial localisation of *ftz* has been studied by the examination of its expression pattern in various mutant and transgenic embryos and it has been shown that the gene is regulated in a complex fashion by the maternal, the gap and four of the pair-rule gene products, although it is not clear at what level of expression this control acts (12–20).

If *ftz* expression is directly controlled at the transcriptional level then the gene should possess cis-acting DNA sequences that mediate the various aspects of its control. Three such control regions have been defined and partially characterised by Hiromi et al. (11,20) (Figure 1a.). The 'zebra element' is responsible for the striped pattern of expression in the mesodermal precursor cells and mediates the control by a number of known *ftz* trans-regulating genes. In addition there is the 'neurogenic element' which is required for expression in the central nervous system and the multi-functional 'upstream element' (USE). This latter element has enhancer-like properties; appears to be responsible for *ftz* expression in the ectodermal precursor cells; restricts the *ftz* domain within the embryo and may also act as the response element for *ftz* activation by the gap gene *Krüppel* and the *ftz* protein itself. There is a possibility that the *ftz* -regulating genes acting through the zebra element also act via the upstream or neurogenic elements. The genetic approach used does not address the question of whether the regulating gene products interact with the control sequences directly.

In this paper we describe the localisation of the binding sites for a number of proteins by DNAaseI footprinting of the 2.6 kb USE in staged embryo nuclear extracts. The USE was chosen because this is the only one of the three *ftz* control elements which has defined limits; it unambiguously regulates expression at the transcriptional level i.e. it does not encode any part of the transcript and it is thought to mediate the effect of at least one of the genes that regulate *ftz* expression *in trans*. We have assessed the regulatory significance of binding factors by examining the temporal profile of their footprints and further investigating the nature of the proteins and DNA sequences they protect. This analysis has allowed us to identify several potential trans-regulators of *ftz* expression.

RESULTS AND DISCUSSION

Sequence of the USE.

As a prerequisite to preparation of footprinting probes and the examination of protected DNA sites, the 2.6 kb KpnI/XbaI fragment defining the *ftz* USE was subcloned and sequenced. The sequence derived restriction map corresponds to that of the element (Figure 1a and 2; Y. Hiromi, personal communication)

The USE base composition is 33% T;19% C; 30% A and 18% G, but notably contains long stretches of A+T rich DNA. One such region of 149 bp, between nucleotides 466 and 614, contains only four GC base pairs. (Figure 1b and 2). This region is included in the 782 bp DNA fragment which has been implicated in nuclear scaffold attachment (21); a proposal which is consistent with its sequence composition and with the clustering of potential topoisomerase II cleavage sites that we have found within these 782 bp. (Figure 1b). As nuclear scaffold attached regions(SAR's) have been found associated with other *Drosophila* gene cis-regulatory elements (21), this evidence serves to confirm the supposition that the USE is involved in transcriptional control.

Footprinting using a nuclear extract from 0–10hr embryos.

DNAaseI digestion of end labelled DNA suggested that the maximum length of sequence

Fig. 1. a) Partial restriction map of the *ftz* transcription unit. Modified from reference 11. Exons are represented by solid boxes; the intron by an open box. Transcription direction is shown by a wavy line. Only relevant restriction sites are shown. K, KpnI; R, EcoRI; X, XbaI; Bal, Ball; H, HindIII. b) Partial restriction map of *ftz* USE showing footprinting probes. Only those sites used in subcloning are shown. K, KpnI; R, EcoRI; A, AluI; SPH, SphI; STU, SstI; S, SacI; B, BglII; SSP, SspI; STY, StyI; EV, EcoRV; X, XbaI. Arrows show the extent of the cloned fragments. The arrowhead marks the end of the fragment closest to the PstI site of the vector polylinker i.e. the 5' end of the labelled strand. The small black boxes represent sites that show 87% homology to the consensus topoisomerase II cleavage site and 100% homology to the 6 bp core. (21). The cross-hatched box delimits a region of 149bp which is 97.3% A+T rich. The large black box represents the 782 bp scaffold-attached region (SAR) of Gasser and Laemmli (21), while the clear box shows the region of uncertainty in its definition.

Initial experiments involved footprinting selected probes in a nuclear extract made from *Drosophila* embryos staged between zero and ten hours after egg deposition at 24°C (AED) (see Materials and Methods). This corresponds to stages 1 to 12 of Campos-Ortega and Hartenstein (23). The assays were done in the presence of excess poly [d(I-C)] DNA to block the non-specific binding of proteins to probe DNA.

The footprinted regions fall into two major classes i) small ($< 40\text{b.p.}$), well-defined footprints (e.g. nucleotides 1672 to 1702) and ii) larger ($> 40\text{b.p.}$) footprints covering DNA which is often A+T rich. (e.g. nucleotides 465–570). Of the small footprints a number protect DNA sequences containing direct or inverted repeats e.g. the footprint between nucleotides 770 and 795 covers a region including a 20b.p. imperfect palindrome (table

	SEQUENCE POSITION	HOMOLOGIES	FOOTPRINT	TIME	COMMENTS
A. HOMOLOGIES WITHIN USE					
	341-353	5'-GGAAATAGAAATA-3' *****	305-371	Both	
	1792-1780	5'-GGAAATAGAAATA-3' cs *****	1740-1856	Both	
	1180-1192	5'-GGAAATGGAAATG-3' *****	1172-1190	Both	
	1416-1430	5'-TTTCCTGTCCAGTTC-3' *****	1400-1430	Both	13/15 match between two
	1184-1170	5'-TTTCCTGTCCAGTTC-3' cs *****	1172-1190	Both	C+T/A+G rich footprints
	1748-1762	5'-TATATTATAGATTAA-3' *****	1740-1856	Both	Many other highly
	871-885	5'-TATATTATAAATTAA-3' cs *****	810-900	Both	homologous A+T rich footprints.
B. HOMOLOGIES TO ZEBRA ELEMENT.					
	1620-1630	5'-GGACGTTATCC-3' ***	1613-1644	F.D.	816 DNA (see text)
	-239- -229 in Zebra element	5'-GGAAGTTATCC-3' cs *****	----	---	(sequence: Laughon and Scott, 1984)
C. HOMOLOGIES TO DNA-BINDING PROTEIN RECOGNITION SEQUENCES.					
	412-434	5'-ATGTAATTTTAGATTTTAATTGA-3' *****	413-447	1-4hr	
	Ev* site	5'-ATTTAATTGA ATTTAATTGA-3' *****	----	---	A+T rich consensus of Hoey et al, 1988.

Footnotes to Table 1. Numbering refers to nucleotide positions in the USE, as shown in Fig. 2, unless otherwise stated. Stars show positions of identical nucleotides. The 'TIME' column refers to the type of footprint as shown in Fig. 4. Abbreviations: F.D., Form differences (Fig. 4); cs, complementary strand to that shown in Fig. 2 or, in the case of the zebra element to that shown in reference 8.

1); the footprint between 1672 and 1702 nucleotides covers DNA containing four 5'-ACAT-3' repeats and the DNA footprinted between nucleotides 1020 and 1060 contains two potential cruciform structures(Figure 2).

In addition to the better defined footprints there are also some areas in which there are differences in DNAaseI cleavage patterns which are not simple footprints, but combinations of enhancements and protections which may be due to structural changes in the DNA associated with protein binding. (e.g. nucleotides 231 to 245)

Differential footprinting of USE during development.

The approach outlined above localises the binding sites for proteins within an important cis-acting element, although it does not give any information about the relative contribution that each makes to the developmental regulation of *ftz*. While some binding proteins may have a regulatory role, others may be chromatin associated factors of no developmental significance or may not even bind to the USE *in vivo*. In the absence of *in vitro* transcription systems that measure the effect of factors associated with far upstream elements and due to the time consuming nature of assaying the *in vivo* enhancer activity of all the individual binding sites (e.g. by P-element transposition), we chose initially to make a relatively rapid assessment of the regulatory importance of the binding proteins by observing whether their binding affinity and specificity alters in concert with changes in *ftz* expression. There are two phases of *ftz* expression. Initially *ftz* transcription is switched on at about one and a half hours AED (Stage 4 of Campos-Ortega and Hartenstein, (23))(26; 27). From this time a pattern of seven *ftz* RNA and protein stripes is seen to develop in the central region of the *Drosophila* embryo. Subsequently, *ftz* expression is switched off and no RNA or protein is observed after about four hours AED (Stage 8)(10;26) Later, bet-

ween approximately four and a half hours and ten hours AED (Stages 9–12), the *ftz* gene is switched on again and expressed in all parasegments of the developing central nervous system(10; 11). We thus prepared nuclear extracts from early (1–4hr) and late (5–10hr) phase embryos and compared the USE footprinting patterns in each.

Using standardised extracts (Materials and Methods) we have shown that four DNA regions are better protected in 1–4hr extracts (e.g. Figure 3c) and nine regions are better protected in 5–10hr extracts (e.g. Figure 3a). There are also eighteen regions which are equally footprinted in both (e.g. Figure 3a and 3b). In addition there are two representatives of a fourth class of footprint in which essentially the same region of DNA is footprinted but the form of protection differs between extracts (e.g. Figure 3b). The pattern of footprinting in 5–10 hr extracts is essentially the same as that seen in 0–10 hr extracts. A summary of the data for the USE, undertaken on both strands of the DNA, is shown in Figure 4.

Footprinting of the USE was also carried out in nuclear extracts from Schneider 1, embryonic *Drosophila* tissue culture cells and the protection pattern has features of both 1–4hr and 5–10hr embryonic extracts but there are some differences from both. For instance, there is a cell extract footprint between nucleotides 1200 and 1219 that is not seen in embryonic extracts (Data not shown).

Footprints which show differences between 1–4 hr and 5–10 hr extracts, by the criterion outlined above, define the binding sites for potential trans-regulators of *ftz* expression. Indeed, such a correlation between the temporal regulation of factor binding and developmental control of transcription has recently been shown in the case of the *Drosophila* ADH gene distal promoter. (28).

Assessment of binding-protein specificity.

DNA binding trans-regulatory proteins are involved in the control of specific sets of genes and so might be expected to show binding site specificity. The footprinting proteins we have observed show specificity in that they bind preferentially to specific probe DNA sequences rather than to excess poly[d(I-C)] DNA. However, poly [d(I-C)] DNA may not compete all the low specificity DNA-binding proteins in extracts (28.), particularly those which bind to A+T rich regions. This possibility prompted us to look at the specificity of the factors binding to various sites in the USE with respect to other competitors and to assess the usefulness of this specificity as a measure of functional relevance of the binding proteins.

We find that some of the footprints initially identified in the presence of poly[d(I-C)] DNA are also observed in the presence of excess calf thymus DNA e.g. nucleotides 2204 to 2231 which are only protected in 5–10hr extracts. Other footprints are however competed by calf-thymus DNA. In addition we have examined specifically A+T rich footprints in the USE and, in some cases, still observe protection, in the presence of calf thymus (e.g. nucleotides 810 to 900, figure 4) and poly[d(A-T)] DNAs (e.g. nucleotides 506–530, figure 5d), however there are also examples of binding proteins which are competed by both(e.g. nucleotides 530–570 , figure 3d).

Fig. 3. Selected DNaseI footprint analyses of nuclear extract factors on end-labelled USE probes (Fig. 2). Assays and functional standardisation were performed as described.(Materials and Methods) a) 2–3' probe. Lanes: 1, no extract; 2, 1–4 hr extract; 3, 5–10 hr extract; 4, Maxam and Gilbert G-track. b) as a) but 4–5' probe. c) as a) but 7–5' probe and no G-track. d) DNA competition on 7–5' probe. Lanes 2 and 3 both contain 5–10 hr extract. Lanes: 1, no extract; 2, 2 μ g poly(dI-dC) DNA; 3, 2 μ g poly(dA-dT) DNA. a), b) and c) were repeated at least three times. Footprints are demarked by a solid line. Dotted lines indicate regions that are shown to be clearly footprinted at higher protein concentration or are protected on the complementary DNA strand.

Fig. 4. Summary of major footprints for both 1–4 hr and 5–10 hr nuclear extracts on *ftz* USE. Nucleotides numbered as in Fig. 1b. Boxes represent maximum extent of footprints on either strand of the DNA. Footprint shading key: 'Both.'; equally protected in both 1–4 hr and 5–10 hr extracts: '1–4 hr.'; preferentially protected in 1–4 hr extracts: '5–10 hr.'; preferentially protected in 5–10 hr extracts: 'Form Differences.'; as for 'Both.' but form of footprint differs between extracts. A dotted line around a footprint represents weak protection. Footprint boxes with nucleotide symbols underneath cover sequences rich in those nucleotides on the strand shown in Fig. 2. In the case of 'A+T,' this is greater than 92% A+T; 'C+T', 85% C+T; 'G+A', 84% G+A. Only defined blocks of protection are shown.(see text) Each protected region was footprinted at least three times on both strands in both extracts.

Thus there is sequence specific binding of proteins even within A+T rich regions, although, consistent with the findings of Heberlein and Tjian (28), there are also less specific interactions. Even some footprints which show differential protection between the different phases of *ftz* expression (e.g. nucleotides 770–795) are among those competed by calf thymus DNA. What then is the biological significance of this lack of precise binding specificity? It is possible that proteins interacting with low specificity have low selectivity of binding or are binding with low affinity to sites that are not their *in vivo* targets. On the other hand, studies on developmentally important homeodomain-containing proteins have shown that even these have limited selectivity *in vitro*, and might well be competed from their true targets by fortuitous sites in competitor DNA (29). Hence, it is possible

Fig. 5. Gel-retardation experiments with #16 and #17 DNAs. **a)** Sequences of the oligonucleotides used to make gel-retardation probes. These cover the USE sequences (Fig.2) denoted in the 'EXTENT OF OLIGO' section of the table flanked by SalI sticky ends. Small arrows mark 5'-ACAT-3' repeats and dotted lines show an extended region of homology between 5'-ACAT-3' repeat regions on both #16 and #17 DNAs, as noted in the 'HOMOLOGIES' section of the table. The 'EXTENT OF FOOTPRINT' section details the nucleotide positions in the USE sequence (Fig. 2) protected in the various extracts marked in bold type. 'CELL' refers to Schneider 1 cell nuclear extract. **b)** Gel-retardation assay of #16 probe. Times above the lanes indicate the embryonic stages from which the extract used in those lanes was made and the values directly above individual lanes represent volume of the indicated extract added to the assay in μ l. Probe/protein complexes referred to in the text are labelled with arrows. **c)** As for **b** but using #17 probe. **d)** As for **b** but values above lanes represent mass of total nuclear protein added in μ g, because, due to the lack of an appropriate reference (see 'Functional standardisation section of Materials and Methods.), extracts were balanced with respect to total protein levels.

that even footprints which are competed by competitors other than poly[d(I-C)] DNA represent the binding sites for potential *ftz* regulatory proteins and because of these considerations we have chosen to use only poly[d(I-C)] competitor in our initial characterisation of the binding factors.

Different footprinting patterns correspond to formation of distinct protein/DNA complexes. To further assess the likelihood that footprints differentially protected in different extracts represent the binding sites for trans-acting regulators of *ftz* expression, we made a closer study of the associated protein/DNA complexes using the gel retardation assay (30). Oligonucleotide probes were made that corresponded to a region that is differentially protected between extracts (nucleotides 1613 to 1644) and one for which the binding pattern does not change (nucleotides 1672 to 1702) (Figure 5a). These probes, numbered #16 and #17 respectively, were shown to bind proteins by footprinting in 5–10hr extracts (Data not shown).

The #17 probe apparently becomes associated with the same pair of specifically competent complexes in both 1–4hr and 5–10hr extracts (Figure 5c)

Specifically competent complexes are also formed with the #16 probe, but they vary in mobility between 1–4hr and 5–10hr extracts (Figure 5b and d). In a 1–4hr extract there are three complexes (*i-iii* in figures 5b and d). In competition experiments complexes *i,ii* and *iii* all appear to be the result of the specific binding of proteins to the #16 probe. As the extract concentration increases (Figure 5b), the probe becomes progressively chased from higher to lower mobility forms (*i* to *iii*), perhaps reflecting multimerisation of the binding proteins or association with other extract factors. In a 5–10hr extract there are two low mobility, low specificity bands (Figure 5b and d). No trace of a higher mobility, specific complex is seen in 5–10hr extracts even at low protein concentration (Figure 5b). These results are consistent with the footprint data, which show differential footprinting between extracts for #16 DNA, but no difference for the #17 DNA, suggesting that the same specific protein/DNA interactions are being studied in both assays. The difference in footprinting specificity and in mobility of the probe/protein complexes on #16 DNA may reflect the binding of distinct proteins in different extracts. On the other hand, the proteins which bind in the 1–4hr extracts may be modified in 5–10hr extracts e.g. by altering their association with other proteins or by chemical modifications. Such a modification could be part of the mechanism by which #16 DNA-binding proteins regulate expression. In one model the modified proteins have altered DNA-binding characteristics and give rise to the shift in protection pattern of #16 DNA with time (i.e. nucleotides 1622–1644 are protected in 1–4hr and 1613–1633 in 5–10hr). This could result in alteration of a functional association between #16 DNA-binding proteins and proteins associated with nearby sites e.g. #17 DNA and as #16 and #17 DNAs show homology (Figure 5a), it is possible that such a functional interaction might involve homodimerisation of a factor binding within both regions. Disruption of functional homodimeric interactions by increased spacing of binding sites has been discussed by Dunn et al. (31) for the *Ara* BAD operon of *E.coli*.

Thus, the observed correlation between footprint pattern and protein/DNA complex changes is consistent with the notion that #16 DNA binding proteins may be acting as *ftz* trans-regulators.

Observation of protein/DNA complexes during early ftz expression.

We further investigated the role of #16 and #17 DNA-binding proteins in *ftz* regulation by looking more closely at the temporal changes in protein binding throughout development of the early *ftz* expression pattern. This approach required the use of short time-window extracts, from which it is difficult to obtain high protein yields and hence the sensitive gel-retardation assay was particularly suitable.

No differences in mobility are seen for the probe/protein complexes for either probe in 1–4hr, 2–3hr or 3–4hr extracts (Figure 5d) suggesting that factors associated with these sites bind to the same extent and with the same specificity throughout this period. Thus, although our previous data suggests that proteins binding #16 DNA are potentially involved in the differential expression of *ftz* between its early and late phases, the #16 and #17 DNA-binding factors are probably not involved in the control of the dynamics of early patterning. This suggestion is subject to the general limitation that we are using extracts from whole embryos and consequently may be masking temporal variations in binding at specific embryonic locations.

Footprints show homology to other cis-acting control elements and trans-acting factor binding sites.

The regulatory relevance of the USE binding-proteins was further assessed by searching for DNA sequence homologies between individual binding sites and the putative control sequences of *Drosophila* developmental genes. Comparisons were made with the zebra element of *ftz* and with the sites protected *in vitro* by various DNA binding proteins known to be involved in the developmental regulation of gene expression. Some of the findings are summarised in Table 1. Within the the USE the motif 5'-GGAAAT-3' occurs a number of times, often in direct and inverted repeats. The occurrence of a double repeat in three footprints implies that the same protein may be binding to this site in each case (Table 1A). The # 16 binding site shows homology to an 11 bp sequence in the *ftz* zebra element (Table 1B). The occurrence of this potential binding site in another of the *ftz* cis-acting elements is a further indication that the # 16 DNA-binding protein is involved in the regulation of the gene.

One footprint, which is preferentially protected in 1–4hr extracts, has homology to the binding site consensus for the homeodomain-containing pair-rule gene product *even-skipped* (32; 33; Table 1C). This is of interest because of genetic data which suggest that *ftz* expression is regulated by *even-skipped* in the early phase of the patterning process (14;34) and thus the footprint could be the result of *even-skipped* binding to the USE. Like the binding sites for other homeodomain-containing proteins the protected DNA is A + T rich (e.g. reference 33.) and it is possible that the other observed A + T rich footprint sites also bind homeodomain proteins. One candidate for this binding is the *ftz* protein, which enhances its own expression via the USE(20). Many of these sites, however, also reside within the proposed nuclear scaffold-attached region and thus may correspond to the large 200bp protein binding domains seen in similar regions of the *Drosophila* histone genes(22). This is a possibility for those footprints which are present throughout the time when the DNA remains scaffold associated i.e. in both 1–4hr and 5–10hr extracts(21).

Thus , on the basis of binding-site homologies, we have further evidence for the potential regulatory nature for at least two of the USE-binding factors.

Conclusions

In this work we have sequenced the USE of the *Drosophila ftz* gene and shown that embryonic nuclear proteins associate with large regions of this DNA *in vitro* . As much recent work has shown that control by cis-regulatory elements tends to be manifested by the binding of one or more effector proteins (reference 35 for review), the association of many DNA-binding proteins with the USE is consistent with its multi-functional nature and indeed certain sets of footprints may define distinct functional subelements.

To control the various phases of *ftz* expression it is likely that regulating DNA-binding factors are required to bind and vacate DNA regions within the USE at precise times, in order to initiate, maintain and repress transcription within particular embryonic domains. Our observation that the footprinting pattern is dynamic, both with respect to binding-site selectivity and the nature of the protein/DNA complexes, is consistent with this expectation. In addition it gives us reason to believe that those footprints which show time dependent protection represent the binding sites for proteins involved in the developmental regulation of *ftz* . For certain proteins this belief is substantiated by homology of the binding sites to those of important classes of proteins regulating development.

The characterisation of the several potentially interesting factors we have identified will

be facilitated by the observation that certain oligonucleotides interact with specific nuclear proteins *in vitro* i.e. the effect of these proteins on *in vitro* transcription in nuclear extracts and *in vivo* expression in transgenic embryos can be examined using constructs in which the oligonucleotide binding sites are included close to existing reporter gene promoter elements. Also, the genes for the binding-proteins may be isolated by screening cDNA expression libraries with labelled oligonucleotide binding sites.(36).

MATERIALS AND METHODS

Sequencing of USE.

The 2.6 kb KpnI/XbaI DNA fragment defining the *fz* USE was isolated from plasmid *pry FKGH*, provided by Y.Hiromi. The fragment was then subcloned into Bluescribe⁺ vector (Bsc⁺)(Vector Cloning Systems) to yield *pHPB13* and sequenced in M13 phage DNA by the random sequencing method described by Bankier and Barrell (37).

Preparation of subclones.

The probe fragments shown in figure 2 were cut from *pHPB13*, their ends were made blunt and they were then subcloned into the HincII site of Bsc⁺. Clones containing the fragment in either orientation were identified by restriction endonuclease mapping and double-stranded sequencing (38).

Collection of embryos.

Embryos were collected and aged at 24°C. Before any collection fresh well-yeasted plates were left in the cages for at least one hour to allow females time to lay eggs that have already started to develop in the abdomen. The distribution of ages was examined by light microscopy and shown to be consistent with the collection period (23). Embryos were not aged beyond 10 hours, as proteases from the developing gut can hamper extract preparation. (R.Jack, personal communication.)

Preparation of embryo nuclear extracts.

Crude extracts (no heparin agarose chromatography) were prepared from Oregon R wild-type *Drosophila* as described by Heberlein and Tjian (28) except that all solutions contained 1mM benzamidine, 0.12 TIU/ml aprotinin, 5µg/ml leupeptin, 2µg/ml pepstatin and HEMG (25mM Hepes pH 7.6, 0.1mM EDTA, 12.5mM MgCl₂, 10% Glycerol, 1mM DTT) solution was replaced with HEG containing no MgCl₂. Embryos were often stored at 4°C for up to 6 hours prior to homogenisation to allow multiple collections to be processed simultaneously. This treatment of embryos does not affect the ability of extracts to transcribe DNA *in vitro*. (M.Biggin personal communication.) As little as 2.5g of embryos were used in the preparation of some extracts. Extracts were frozen in liquid N₂ and stored at -70°C, however the quality of extracts decreases with repeated freezing and thawing.

Preparation of end labelled probes.

As probes were cloned in both orientations into the HincII site of the Bsc⁺ polylinker, fragments can be isolated with an AvaI site at either end by an AvaI/PstI digest of the appropriate clone. Restriction fragments were then labelled at the AvaI end by end-filling(39). Following the labelling, probes were purified by size fractionation on Sephadex G-50 and 5% polyacrylamide gels. Labelling efficiencies of greater than 1x 10⁷ cpm/pmol were routinely obtained.

Cloning of oligonucleotides.

Oligonucleotides corresponding to the single strands in figure 5 were synthesised on an Applied Biosystems 380B DNA synthesiser and then gel purified on a 10% polyacrylamide-urea gel. The strands were annealed and the resulting fragments were then ligated into

the Sall site of Bsc⁺. All clones were checked by double-stranded sequencing and end-labelled probes made as described previously.

DNA binding gel-retardation assay.

The DNA binding assay used was based on that of Riddihough and Pelham (40). 0 to 8 μ l of crude nuclear extract was incubated in a total of 20 μ l of 0.1M HEG, 1mM Spermidine, 0.1 μ g/ μ l poly[d(I-C)] DNA, 2mM EDTA, 0.1% NP40 for 15 min on ice. After addition of $\approx$ 2 fmol probe, the mix was incubated at 30°C for 20 min and then loading buffer added to a final concentration of 4mM Tris-HCl pH 7.5, 1.5mM NaCl, 0.01% bromophenol blue, 0.01% xylene cyanol and 4% glycerol. The samples were loaded onto a 4%, 30:1 acrylamide:bisacrylamide gel in 1/4 X TBE. Gels were run at 7V/cm until the bromophenol blue dye reached the gel bottom and then dried and autoradiographed. In most competition experiments 2 fmol to 2 pmol of the smaller PvuII fragment from either the oligonucleotide subclone or from vector Bsc⁺ DNA was added with the probe.

DNAseI footprint assays.

The binding reactions were as before, but up to 10 fmol probe and 0 to 16 μ l extract were used. No NP40 was included. After the 30°C incubation the mixes were made 6mM MgCl₂, 6mM CaCl₂ and DNAse I was added to 0.1 to 16 μ g/ml DNAse I concentration was chosen by titration to give a uniform DNA cleavage ladder, for each amount of extract. After 60 seconds at room temperature 50 μ l of 50mM EDTA, 0.2% SDS, 10 μ g/ml tRNA, 100 μ g/ml proteinase K was added and the mixture incubated at 42°C for 45 minutes. The DNA was then phenol extracted; ethanol precipitated; washed twice in 70% ethanol; dried; resuspended in formamide dyes; heated at 100°C for 30 seconds and run on a 6% polyacrylamide-urea gel.

In competition experiments the following amounts of competitor were used: 1–4 μ g poly(dI-dC) DNA; 1–4 μ g poly(dA-dT) DNA; 0.5 μ g calf-thymus DNA.

Functional standardisation of extracts.

Nuclear extract preparations from different embryo stages differ qualitatively and it is possible that the final protein composition may have varying proportions of nuclear protein and contaminants. Thus, rather than standardising extracts with respect to total protein, a functional standardisation was used. This involved footprinting the 4–5' probe in increasing amounts of extract and considering as equivalent those amounts of different extracts which gave the same degree of protection of the region spanning nucleotides 1672–1702 (reference footprint, figure 3b). Hence all differences in degree of protection at other sites are relative to a zero difference for the reference. Approximately the same degree of protection of the reference footprint was used on all occasions.

G-tracks of the probes were prepared by the method of Maxam and Gilbert (41).

Computer homology searches.

These were performed using the DIAGON program of Staden (42).

ACKNOWLEDGEMENTS

We would like to thank Sandra Satchwell and Mark Biggin for helpful technical advice; Terry Smith for synthesising oligonucleotides and Ruth Lehmann, Maria Leptin, Jonathan Hodgkin and Carlos V. Cabrera for useful comments on the manuscript. S.D.H. would also like to thank Helen Benjafield for help in entering many kilobases of DNA sequence into the computer. S.D.H. was supported by an MRC studentship.

REFERENCES

1. Nusslein-Volhard, C., and Wieschaus, E. (1980). Nature **287**, 795–801.
2. Shore, D., Stillman, D.J., Brand, A.H., and Nasmyth, K. (1987). EMBO J. **6**, 461–467.

3. Gaul, U., and Jackle, H. 1987. *Cell* **51**, 549–555.
4. Tautz, D. (1988). *Nature* **332**, 281–284.
5. Akam, M. (1987) *Development* **101**, 1–22.
6. Scott, M.P., and Carroll, S.B. (1987). *Cell* **51**, 689–698.
7. Ingham, P. W. (1988). *Nature* **335**, 25–33.
8. Laughon, A., and Scott, M.P. (1984). *Nature* **310**, 25–31.
9. Martinez-Arias, A., and Lawrence, P.A. (1985) *Nature* **313**, 639–642.
10. Carroll, S.B., and Scott, M.P. (1985). *Cell* **43**, 47–57.
11. Hiromi, Y., Kuroiwa, A., and Gehring, W.J. (1985). *Cell* **43**, 603–613.
12. Degelmann, A., Hardy, P.A., Perrimon, N., and Mahowald, A.P. (1986). *Develop.Biol.* **115**, 479–489.
13. Carroll, S.B., Winslow G.M., Schupbach, T., and Scott, M.P. (1986). *Nature* **323**, 278–280.
14. Carroll, S.B., and Scott, M.P. (1986). *Cell* **45**, 113–126.
15. Ingham, P.W., Ish-Horowicz, D., and Howard, K.R. (1986). *EMBO J* **5**, 1659–1665.
16. Mlodzik, M., DeMontrion, C.M; Hiromi, Y., Krause, H.M., and Gehring, W.J. (1987). *Genes and Develop.* **1**, 603–614.
17. Mahoney, P.A., and Lengyel, J.A. (1987). *Develop.Biol.* **122**, 464–470.
18. Howard, K., and Ingham, P.W. (1986). *Cell* **44**, 949–957.
19. Ish-Horowicz, D., and Pinchin, S.M. (1988) *Cell* **51**, 405–415.
20. Hiromi, Y., and Gehring, W.J. (1987). *Cell* **50**, 963–974.
21. Gasser, S.M., and Laemmli, U.K. (1986b). *Cell* **46**, 521–530.
22. Gasser, S.M., and Laemmli, U.K. (1986a). *EMBO J.* **5**, 511–518.
23. Campos-Ortega, J.A., and Hartenstein, V. (1985). *The Embryonic Development of Drosophila Melanogaster*, Springer-Verlag, Berlin + Heidelberg.
24. Davidson, I., Fromental, C., Augereau, P., Wildeman, A., Zenke, M., Chambon, P. (1986) *Nature* **323**, 544–548.
25. Soeller, W.C., Poole, S.J., and Kornberg, T. (1987). *Genes and Development* **2**, 68–81.
26. Hafen, E., Kuroiwa, A., and Gehring, W.J. (1984). *Cell* **37**, 833–841.
27. Weir, M.P. and Kornberg, T. (1985). *Nature* **318**, 433–439.
28. Heberlein, U., and Tjian, R. (1988). *Nature* **33**, 410–415.
29. Desplan, C., Theis, J., O'Farrell, P.H. (1985). *Nature* **318**, 630–635.
30. Fried, A., and Crothers, D.M. (1981). *Nuc. Acids Res.* **9**, 6505–6525.
31. Dunn, T.M., Hahn, S., Ogden, S., and Schleif, R.F. (1984) *Proc. Natl. Acad. Sci.* **81**, 5017–5020.
32. MacDonald, P.M., Ingham, P., and Struhl, G. (1986). *Cell* **47**, 721–734.
33. Hoey, T., and Levine, M. (1988). *Nature* **332**, 858–861.
34. Frasch, M., and Levine, M. (1987). *Genes and Develop* **1**, 981–995.
35. Maniatis, T., and Goodbourne, S., and Fischer, J.A. (1987). *Science* **236**, 1237–1245.
36. Singh, H., Le Bowitz, J.H., Baldwin A.S. Jr., Sharp .P.A. (1988). *Cell* **52**, 415–423.
37. Bankier, A.T., and Barrell, B.G. (1983). Shotgun DNA sequencing. In R.A.Flavell (ed) *Techniques in the Life Sciences B5, Nucleic Acid Biochemistry B508*, Elsevier Scientific Publishers, Ireland Ltd. , pp 1–34.
38. Hattori, M., and Sakaki, Y. (1986). *Analytical Biochemistry* **152**, 232–238.
39. Travers, A.A., Lamond, A.I., Mace, H.A.F., Berman, M.L. (1983). *Cell* **35**, 265–273.
40. Riddihough, G., and Pelham, H. (1987). *EMBO J.* **6**, 3729–3734.
41. Maxam, H., and Gilbert, W. (1980) *Meth. Enzymol.* **65**, 499–560.
42. Staden, R. 1982. *Nuc. Acids Res.* **10**, 2951–2961.