The interaction of proteins encoded by *Drosophila* homeotic and segmentation genes with specific DNA sequences

ALLEN LAUGHON[1], WILLIAM HOWELL[1] and MATTHEW P. SCOTT[2]

[1]*Laboratory of Genetics, University of Wisconsin, 445 Henry Mall, Madison, Wisconsin 53706, USA*
[2]*Department of Molecular, Cellular and Developmental Biology, University of Colorado, Campus Box 347, Boulder, Colorado 80309, USA*

Summary

The ANT-C gene cluster is part of a network of genes that govern pattern formation in the development of *Drosophila*. The ANT-C genes encode proteins that contain a conserved 60 amino acid sequence, the homeodomain. Here we show that the homeodomains encoded by two of the ANT-C loci confer sequence-specific DNA-binding activity. The DNA sequence specificities of the *Dfd* and *ftz* homeodomains appear to overlap completely *in vitro*, indicating that differences in regulatory specificity among ANT-C and BX-C proteins (assuming that differences exist) must be a consequence of the nonconserved protein sequences found outside of the homeodomains. Deletions that remove sequences from either end of the *ftz* homeodomain abolish DNA-binding activity, consistent with the commonly held assumption that the homeodomain is a structural domain. The relevance of *in vitro* DNA-

binding experiments to the regulatory function of *ftz* is supported by our finding that a temperature-sensitive *ftz* mutation that causes a pairwise fusion of embryonic segments also reduces the affinity of the *ftz* homeodomain for DNA. Restriction fragments containing *ftz* homeodomain binding sites were identified within a 90 kb stretch of DNA extending the *Antp* P1 and P2 promoters. Binding sites appear to be clustered near the P1 promoter but also occur near P2 and in the region between the two. The task remains of determining which of these sequences mediate regulation of *Antp* by *ftz* or by other genes that encode closely related homeodomains.

Key words: homeobox, *fushi tarazu*, *Drosophila*, gene regulation.

Introduction

The problem of pattern formation has been approached from a genetic standpoint in studying the segmentation of *Drosophila* (Nüsslein-Volhard & Wieschaus, 1980). The genetic program for the formation of segments involves a series of regulatory interactions that are able to generate spatial patterns of gene expression early during embryogenesis (Akam, 1987; Scott & Carroll, 1987). This regulation appears to occur primarily (but not exclusively) at the transcriptional level and therefore many of the segmentation genes can be expected to encode proteins that control transcription.

Current evidence suggests that the two homeotic gene complexes of *Drosophila*, the Antennapedia Complex (ANT-C) and the Bithorax Complex

(BX-C), constitute a family of transcription factor genes. The homeotic genes within this group specify the different identities of segments along the anterior–posterior axis (Lewis, 1978; Kaufman, 1983) and share a short conserved sequence, the homeobox (McGinnis *et al.* 1984; Scott & Weiner, 1984). Homeoboxes each encode a 60 amino acid protein sequence that appears to contain a helix-turn-helix DNA-binding motif utilized by a number of bacterial and yeast regulatory proteins (Laughon & Scott, 1984; Shepherd *et al.* 1984). The significance of this similarity is supported by experiments that have shown that the homeodomain of the *engrailed* gene will bind to specific DNA sequences *in vitro* (Desplan *et al.* 1985).

The homeodomain proteins are not exclusively devoted to the specification of segmental identity.

Three ANT-C homeobox-containing genes, *fushi tarazu* (*ftz*), *zerknüllt* and *bicoid*, regulate segment number, dorsal–ventral pattern and anterior–posterior polarity, respectively (Wakimoto *et al.* 1984; Doyle *et al.* 1986; Frohnhöfer & Nüsslein-Volhard, 1986; Frigerio *et al.* 1986). The common thread among homeobox-containing genes within the ANT-C and BX-C, as well those located elsewhere, seems to be their function of regulating the transcription of other genes.

The pair-rule gene, *ftz*, whose transcripts and protein are expressed at the blastoderm stage in even-numbered parasegments (Hafen *et al.* 1984*a*; Carroll & Scott, 1985), is required for the transcription of *en* in even-numbered parasegments (Howard & Ingham, 1986; DiNardo & O'Farrell, 1987) as well for regulating its own levels of expression (Hiromi & Gehring, 1987). In addition to its role in the formation of segments, several observations link *ftz* to the expression of homeotic genes at the molecular level. Akam (1985) observed that as *Ubx* transcripts begin to appear shortly after formation of the blastoderm, they exhibit higher levels of expression in even-numbered parasegments (in register with *ftz*), suggesting that early expression of *Ubx* is affected by a pair-rule gene, a phenomenon he termed pair-rule modulation. Ingham & Martinez-Arias (1986) found that the level of *Ubx* expression in parasegment 6 was greatly reduced in *ftz* mutant embryos. In addition, the early expression of transcripts from two other homeotic genes, those from the P2 promoter of *Antennapedia* (*Antp*) and from *Sex combs reduced* (*Scr*), were eliminated in *ftz* mutant embryos. Together, these results indicate that *ftz* is required for proper early expression of ANT-C and BX-C homeotic loci along the length of the gastrulating embryo. A connection between *ftz* and the expression of BX-C homeotic genes is also indicated by dominant *ftz* mutations that cause homeotic transformations similar in phenotype to particular BX-C mutations (Duncan, 1986).

To investigate the possibility that *ftz* directly regulates transcription, we wanted to know whether the *ftz* protein and, in particular, the *ftz* homeodomain would bind to specific DNA sequences, as has been demonstrated for the *en* homeodomain. An important question is whether different ANT-C homeodomains are able to recognize the same DNA sequences, as was suggested by the conservation of amino acids in the part of the homeodomain that appears to correspond to the base pair recognition helix of the bacterial helix-turn-helix DNA-binding fold. For these experiments, we compared the binding activities of the homeodomains encoded by *ftz* and *Deformed* (*Dfd*), an ANT-C homeotic gene involved in the formation of the head (Regulski *et al.*

1987). In addition, we have identified *ftz* homeo-domain binding sites in *Antp*, a gene that is subject to regulation by *ftz* and by BX-C genes with *ftz*-related homeodomains.

Results

Expression of ftz and Dfd protein sequences in E. coli

Previous work has shown that *ftz* encodes a 413 amino acid protein with the homeobox beginning 765 bp from the start of translation (Laughon & Scott, 1984). To express *ftz* in *E. coli*, we ligated the 5′ end of the *ftz* coding region in-frame to the 3′ end of the *lacZ* coding region in the expression vector pUR290 (Rüther & Müller-Hill, 1983) (Fig. 1). The resulting plasmid, pFTZ3, expresses a fusion protein (FTZ3) with an apparent relative molecular mass of 175×10^3 on SDS gels (Fig. 2). The predicted size of FTZ3 is 161×10^3 but the anomalously slow mobility is in line with the appearance of the 45×10^3 *ftz* protein as a 66×10^3 species on SDS gels (A. Boulet and A. L., unpublished observations). The identity of the 175×10^3 species was also confirmed by probing Western blots with affinity-purified antibodies against *ftz* protein (data not shown).

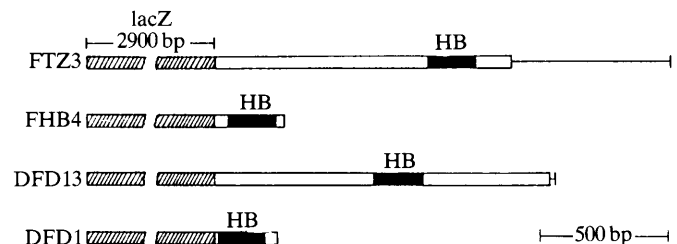Although large amounts of FTZ3 protein are

Fig. 1. The arrangements of pUR290, 291 or 292-derived *lacZ–ftz* and *lacZ–Dfd* gene fusions used to generate β-galactosidase fusion proteins. Hatched bars are *lacZ* sequences, solid bars are homeoboxes, open bars are protein-coding sequences flanking the homeoboxes in *ftz* or *Dfd*, and lines are nontranslated *ftz* or *Dfd* sequences. The procedures for construction and characterization of *lacZ* fusions using pUR vectors (Rüther & Müller-Hill, 1983), including the preparation of native protein extracts containing β-galactosidase fusion proteins and Western blotting are described in Carroll & Laughon (1987). FTZ3 was constructed from a *ftz* clone containing an artificial NdeI site, introduced at the *ftz* initiator AUG by site-directed oligonucleotide mutagenesis. FHB4 was constructed by insertion of an *Nae*I fragment from the *ftz* cDNA clone, G20 (Laughon & Scott, 1984) into the *Bam*HI site of pUR291. DFD13 contains all of the G23 cDNA coding sequences while DFD1 contains the *Hpa*II fragment that encompasses the homeobox (Laughon *et al.* 1985). Plasmids were transformed into the *lacI^Q* strain, 71-18 (Rüther & Müller-Hill, 1983).

apparent in detergent lysates, little of the intact protein is recovered when cells are lysed under nondenaturing conditions (Fig. 3). Intact FTZ3 is sequestered in inclusion bodies that can only be dissolved with detergent or high concentrations of chaotropes such as urea or guanidine (data not shown). Native extracts are enriched with degraded fusion protein that is apparently more soluble than the intact fusion. In contrast, unfused β-galactosidase expressed from pUR plasmid is intact and soluble at very high concentrations. The instability of *ftz* protein and its deposition in inclusion bodies is typical of the behaviour of many eukaryotic proteins when expressed in *E. coli* (Marston, 1987). Preparations of FHB4, a fusion of the *ftz* homeodomain to β-galactosidase, were less degraded and more soluble than FTZ3 (Fig. 2), although the bulk of FHB4 is also contained in inclusion bodies. Two *lacZ–Dfd* fusions (Fig. 1) encode proteins with solubility properties analogous to their *ftz* counterparts. (The native preparation of DFD1 shown in Fig. 2 is at least fivefold lower than in typical preparations.) For use in DNA-binding experiments, fusion proteins were partially purified from crude *E. coli* extracts by selective precipitation with ammonium sulphate, followed by

Fig. 2. Expression of β-galactosidase–*ftz* and –*Dfd* fusion proteins in *E. coli*. Proteins fractionated by electrophoresis in a 7·5 % SDS–polyacrylamide gel and stained with Coomassie brilliant blue are shown. The left lane of each pair was loaded with the lysate of cells, boiled in sample buffer containing 1 % SDS, from 50 μl of IPTG-induced log-phase culture. The right lane of each pair was loaded with 10 μl of native protein extract obtained from cells from 500 μl of the same cultures. (A) FTZ3; (B) FHB4; (C) SHB1; (D) FHB100; (E) FHB200; (F) DFD13; (G) DFD1. Arrows point out the positions of β-galactosidase (116×10^3) the FHB4 and DFD1 fusion proteins (129×10^3) and the FTZ3 and DFD13 fusions (175×10^3).

dialysis, and were stored frozen at $-80°$C or at $-20°$C in a buffer containing 50 % glycerol.

The *ftz* homeodomain has sequence-specific DNA-binding activity

FHB4 contains the *ftz* homeodomain plus 17 and 12 flanking amino acids at the amino and carboxyl ends, respectively (Fig. 1). The ability of FHB4 to bind DNA was tested by incubating the extracts containing FHB4 with ^{32}P-labelled restriction fragments, followed by precipitation of the fusion protein with anti-β-galactosidase antibodies coupled to sepharose beads, essentially as described by Desplan *et al.* (1985). After washing, bound DNA was released with SDS and phenol extraction, followed by precipitation with ethanol. Bound fragments were analysed by electrophoresis on a 5 % nondenaturing polyacrylamide gel. In the first set of experiments, we used *Hae*III fragments from pEMBL18 and from a plasmid (ftzKS7) containing 7 kb of *ftz* sequence inserted into pUC18 (Fig. 3). In a low ionic strength buffer containing 25 mM-NaCl, FHB4 binds to most of the restriction fragments in the mixture. The concentration of fusion protein is limiting in these experiments and addition of more FHB4 extract increases both the amount and number of fragments bound, indicating that *ftz* has general affinity for DNA but that it binds preferentially to particular sequences. Increasing the NaCl concentration to 200 mM reduces the binding of fragments differentially such that a small number bind with a much higher affinity than the remainder in the mixture (Fig. 3). Between 250 and 300 mM-NaCl, DNA-binding is almost completely abolished. Using this assay, unfused β-galactosidase from a strain containing the plasmid pUR290 shows no detectable DNA-binding activity in either high or low salt buffer (data not shown).

The full-length FTZ3 fusion protein was also tested for specific and nonspecific DNA-binding activity and found to be indistinguishable from FHB4 (Fig. 3). A tentative conclusion from this experiment would be that all of the DNA-binding activity of *ftz* protein resides within the homeodomain. It is clear from Fig. 2 that the FTZ3 protein in our extracts is substantially degraded, possibly lacking amino acids that affect the DNA-binding activity of the homeodomain or that could confer additional DNA-binding activities. However, proteins cleaved on the amino terminal side of the homeodomain would not be immunoprecipitated, only leaving in doubt the importance of the 100 amino acids on the carboxy terminal side of the *ftz* homeodomain. The concentration of FHB4 is also higher than FTZ3 in extracts made in parallel (Fig. 2). Equal extract volumes were used in the binding experiments and the increased amount of DNA bound using FHB4 compared to
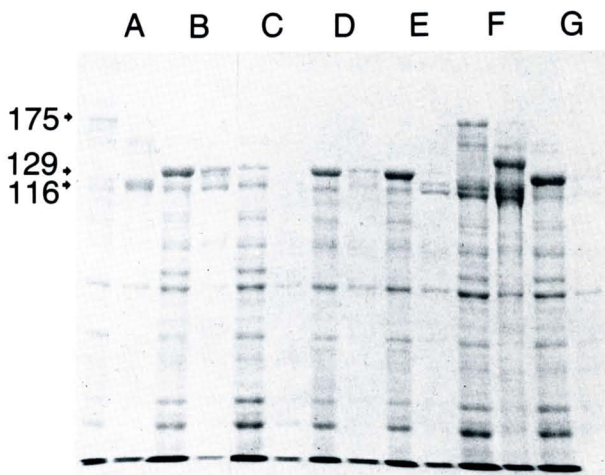
FTZ3 in Fig. 3 is due to the difference in fusion protein concentration and not to a difference in relative binding affinity.

Two deletion derivatives of FHB4 were constructed in an effort to localize further the DNA-binding activity within the *ftz* homeodomain. One, FHB100, lacks 16 codons at the 5′ end of the homeobox and the other, FHB200, is missing 12 codons from the 3′ end of the homeobox. Specific or nonspecific affinity for DNA is undetectable for either protein, within the sensitivity of our binding assay (Fig. 3). FHB100 deletes a portion of the homeodomain outside of the helix-turn-helix DNA-binding fold postulated to exist on the basis of similarity to bacterial regulatory proteins (Laughon & Scott, 1984), while FHB200 removes almost half of

what should be the α helix that is predicted to make sequence-specific contacts within the major groove of DNA. The failure of either protein to bind to DNA, even nonspecifically, suggests that the homeodomain does in fact constitute a domain structure that must be relatively intact if the DNA-binding site is to be functional.

A ftz mutation affects DNA binding

The relevance of *in vitro* DNA-binding experiments to the *in vivo* function of *ftz* is supported by exper-

Fig. 3. Binding of DNA fragments by *ftz* and *Dfd* fusion proteins. A 24 h autoradiographic exposure of ^{32}P-labelled restriction fragments fractionated by electrophoresis in a nondenaturing 5 % acrylamide gel is shown. Bound DNA fragments were recovered after immunoprecipitation with anti-β-galactosidase antibody coupled to sepharose. Except for the pairs of lanes marked T, E and K, the left lane in each pair contains *Hae*III restriction fragments from a 7 kb *Kpn*I to *Sal*I *ftz* subclone in pUC18 (*Kpn*I is 6 kb upstream and *Sal*I is within the *ftz* coding region) that were bound by fusion protein in a buffer containing 25 mM-NaCl (low-salt buffer). The right lane in each pair contains fragments that were bound by fusion protein in a buffer containing 200 mM-NaCl (high-salt buffer). ftz 3, 2 μl of FTZ3 extract; ftz 4, 5 μl of FHB4 extract from cells grown at 18°C; 18, 5 μl of SHB1 extract from cells grown at 18°C; 37, 5 μl of SHB4 extract from cells grown at 37°C; 100, 2 μl of FHB100 extract; 200, 2 μl of FHB200 extract; Dfd 1, 2 μl of DFD1 extract; Dfd 13, 2 μl of DFD13 extract. T, total labelled *Hae*III digest of pEMBL18 (left) or ftzKS7 (right) added to each binding reaction; E, pEMBL18 *Hae*III fragments bound by FHB4 (1 μl of extract) in low salt (left) and high salt (right); K, ftzKS7 *Hae*III fragments bound by FHB4 (1 μl of extract) in low salt (left), and an empty lane (right). Arrows mark the positions of 400 and 1000 bp fragments bound in high-salt buffer. Preparation of fusion protein extracts was according to the protocol described in Carroll & Laughon (1987). Labelling of DNA fragments with T$_4$ DNA polymerase, and immunoprecipitation and electrophoresis of DNA fragments bound to fusion proteins were performed according to Desplan *et al.* (1985) except that 20 μl of sepharose coupled to 20 μg of anti-β-galactosidase was added per 20 μl binding reaction instead of using soluble antibody plus *Staphylococcus aureus* cells. DNA-binding reactions contained 20 mM-Tris pH 7·5, 25 or 200 mM-NaCl, 5 mM-MgCl$_2$, 100 μg ml^{-1} BSA, 10 ng of ^{32}P-labelled restriction fragments and 1–2 μl of fusion protein extract or diluted extract. Washes were done using the same buffer lacking DNA and protein extract. All binding experiments were done at 4°C, unless otherwise noted in the text.
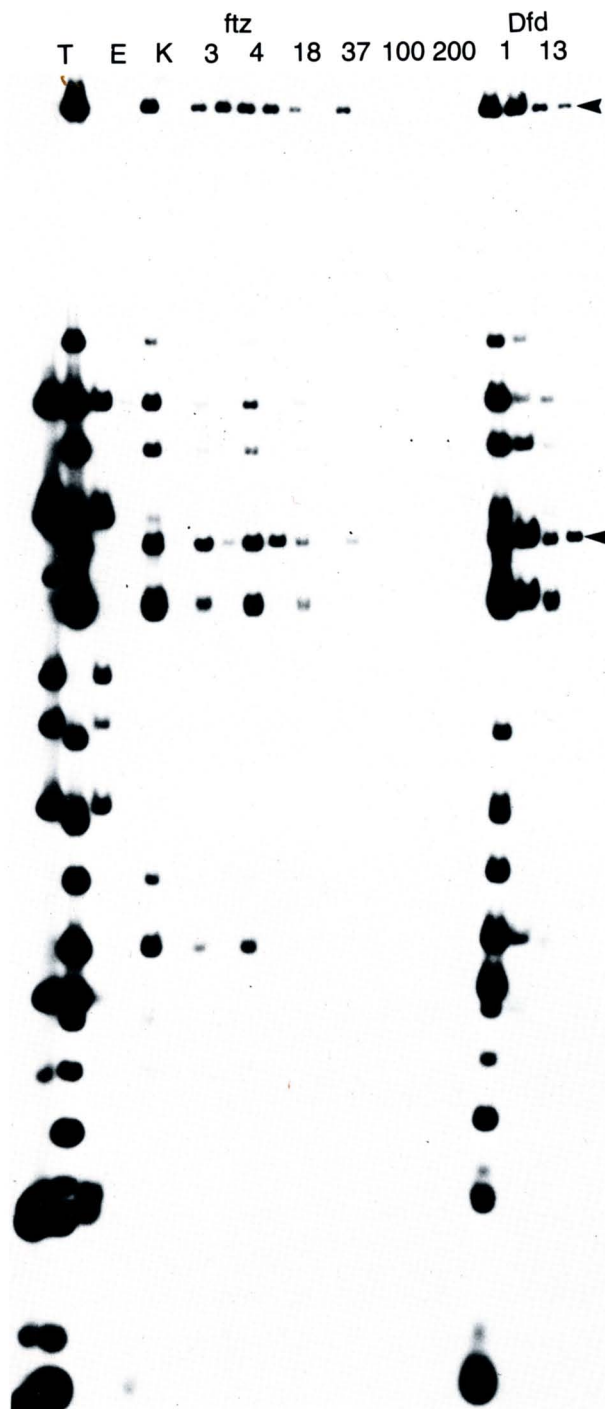
iments which test the effect of a mutant *ftz* allele on DNA binding. Two *ftz* mutations have been characterized at the molecular level (Laughon & Scott, 1984). One, ftz^{f47ts}, is a temperature-sensitive allele that results from the conversion of a conserved alanine to a valine in the first helix of the putative DNA-binding motif. Alanine is conserved at the corresponding position in helix-turn-helix bacterial DNA-binding proteins where it forms part of a hydrophobic pocket thought to be important for the alignment of the two helices (Pabo & Sauer, 1984). SHB1, a derivative of FHB4 containing the alanine-to-valine change, was constructed to test the effect of the temperature-sensitive mutation on DNA binding. The permissive and restrictive temperatures for the mutation in flies are 18°C and 29°C, respectively. Therefore, binding experiments were performed using extracts from *E. coli* cultures grown at 18°C. Binding assays were performed at 4°C, 18°C and 29°C (Fig. 4). The nonspecific binding activity of both FHB4 and SHB1 in low-salt buffer was essentially identical at the three temperatures, with only a slight drop in activity at 29°C for both proteins. This result is in contrast to the deleted forms of the homeodomain described above, which do not have detectable nonspecific DNA-binding activity. In high-salt buffer, the specific binding activity of FHB4 was reduced twofold going from 4 to 18°C and by another sevenfold going from 18 to 29°C, while SHB1 completely lacked detectable specific binding activity at all three temperatures. While these experiments have not detected a temperature-sensitive effect of the mutation on DNA binding, it is clear that, compared to the wild-type homeodomain, the mutant protein has a reduced sequence-specific affinity for DNA, correlating DNA binding with the *in vivo* function of the gene.

Deformed *encodes a protein with DNA-binding specificity closely related to that of* ftz

Given the high degree of sequence homology among ANT-C and BX-C homeoboxes, it is important to find out whether other homeodomains within this group have DNA-binding specificities related to that of the *ftz* homeodomain. To test this possibility, we investigated the DNA-binding properties of the *Dfd* homeodomain, a member of the ANT-C homeodomain group that is 73 % identical in sequence to the *ftz* homeodomain (Laughon *et al.* 1985; Regulski *et al.* 1987). pDFD1 is a pUR *lacZ* fusion starting two codons 5′ of the *Dfd* homeobox and ending 21 codons past its 3′ end (Fig. 1). A larger, but not full-length, fusion, pDFD13, begins 202 codons upstream of the homeobox and ends past the 3′ end of the coding region, encoding 427 amino acids of *Dfd* protein (Laughon *et al.* 1985). The full-length *Dfd* product
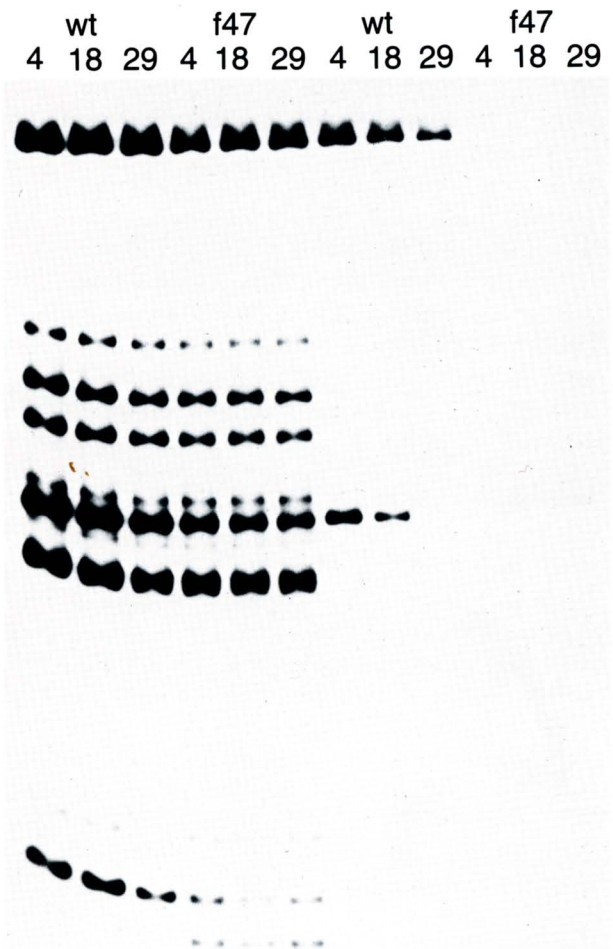

Fig. 4. Affect of the ftz^{f47ts} mutation on DNA binding. Shown is a 24 h autoradiograph of ftzKS7 *Hae*III fragments that were bound by FHB4 (wt) or SHB4 (f47) and fractionated on a 5 % native polyacrylamide gel. Extracts were prepared from cultures grown at 18°C and binding experiments were done at 4, 18 or 29°C in low-salt buffer (left half of figure) or high salt buffer (right half of figure).

contains an additional 160 amino acids at its amino-terminal end (Regulski *et al.* 1987). The β-galactosidase–*Dfd* fusion proteins were expressed at levels comparable to those of the *ftz* fusions (Fig. 2).

The results of *Dfd* protein-binding experiments done in parallel with those described for *ftz* fusion proteins are shown in Fig. 3. In low-salt buffer, DFD13 binds DNA with the same apparent limited specificity as FHB4 and FTZ3. In contrast, DFD1 efficiently bound to all of the ftzKS7 fragments in low salt buffer. However, the high concentration of DFD1 in extracts relative to DFD13, FHB4 and FTZ3, (about fivefold higher) might account for the difference since the amount of extract is limiting in

these experiments (data not shown). In the high-salt buffer, DFD13 binds specifically to the same two *ftz* DNA fragments as do the *ftz* fusion proteins. However, DFD1 binds strongly to these fragments and more weakly to five additional fragments, the same set of fragments that are preferentially bound by DFD13, FHB4 and FTZ3 in low-salt buffer. The appearance of fragments bound by DFD1 but not by the *ftz* fusions in these experiments could be due to a difference in DNA-binding specificity or to insufficient washing of the sepharose beads prior to elution of the bound DNA. The second explanation is likely given the failure of DFD13 to bind to these additional fragments in high-salt buffer.
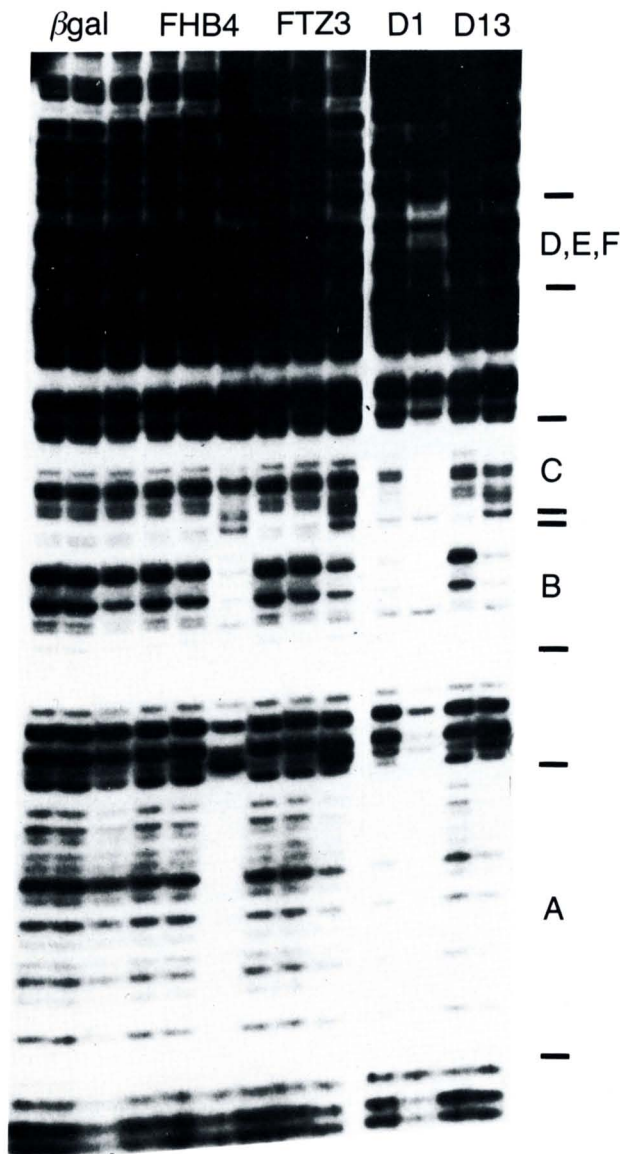
Footprint analysis of ftz- and Dfd-binding sites

The comparison of *ftz* and *Dfd* DNA-binding specificity was extended with the use of DNase I footprinting (Fig. 5). For this analysis, we chose a 495 bp *Hinf*I fragment located 2135 bp upstream of the *Antp* P1 promoter, a fragment that is bound extremely well in precipitation binding assays (see below). The FTZ3, FHB4, DFD13 and DFD1 proteins each generate an almost identical set of six footprints, three of which are well resolved in Fig. 5 and are lettered A, B and C. These are 18, 17 and 17 bp in length, separated by 8 and 1 bp. FTZ3 and FHB4 also cause enhanced DNaseI cleavage at a thymidines at one end of site A, and between B and C and of an A–T pair at the end of C. DFD1 completely protects sites A, B and C at a concentration of $200\,\mu g\,ml^{-1}$ and at $20\,\mu g\,ml^{-1}$

Fig. 5. Identification of homeodomain binding sites by DNaseI footprinting. Shown is a 72 h autoradiograph of an end-labelled *Antp* DNA fragment, 577R2.2-*Hinf*I, fractionated on a sequencing gel after partial digestion with DNaseI. Prior to the addition of DNaseI, the fragment was incubated with the following volumes (in μl) of protein extracts (left to right): β-gal, 0·01, 0·1, 1; FHB4, 0·01, 0·1, 1; FTZ3, 0·01, 0·1, 1; D1 (DFD1), 0·1, 1; D13 (DFD13), 0·01, 0·1. A–F indicate the positions of sequences protected from DNaseI cleavage to a varying degree by the *ftz* and *Dfd* fusion proteins. DNaseI footprinting was performed as follows: 5 ng of DNA fragment, ^{32}P-labelled at a single 5' end, was incubated on ice with varying amounts of fusion protein extract in $20\,\mu$l containing 50 mM-NaCl, 20 mM-Tris pH 7·5, 5 mM-MgCl$_2$, $100\,\mu g\,ml^{-1}$ BSA and $2\,\mu g\,ml^{-1}$ dIdC (Sigma). After 30 min, 0·5 units of DNaseI (Worthington) was added and incubation at 4°C was continued for 5 min, at which time the reaction was stopped by addition of $25\,\mu$l of 250 mM-NaCl, 20 mM-Tris pH 7·5, 10 mM-EDTA, 1 % SDS. Following extraction with phenol, DNA was recovered by precipitation with ethanol, dissolved in deionized formamide plus 0·08 % xylene cyanol and electrophoresed on a 10 % denaturing acrylamide gel along with sequencing ladders of the same labelled fragment generated by chemical cleavage.

$(3\times10^{-8}\,\text{M tetramers})$ sites A and C are protected in about half of the fragments in the binding reaction $(10^{-9}\,\text{M}$ total labelled DNA fragments). Assuming that the fusion protein binds as a β-galactosidase tetramer and that nonspecific affinity for DNA is low enough for the vast majority of the protein to be free in solution, we can derive an apparent equilibrium binding constant of $3\times10^{-8}\,\text{M}$. The actual binding affinity would be higher if a portion of the DFD1 protein detected on SDS gels is inactive with respect to DNA binding.

Comparison of the three footprints indicates that *ftz* and *Dfd* show a preference for AA/TT dinucleotides but does not provide enough information to derive a consensus sequence for the preferred homeodomain binding site (H. Nelson, M.P.S. and A.L., unpublished data). This also suggests that none of these sequences constitutes an optimal consensus binding site for the *ftz* or *Dfd* homeodomains and it

follows that these homeodomains should be capable of binding affinities in excess of the $\approx 10^{-8}$ M value we have estimated from these imperfect binding sites.

Identification of ftz *homeodomain binding sites in genes regulated by* ftz

To be able to investigate further how homeodomain proteins regulate gene expression, it is essential to identify the *cis*-acting sequences that they normally act upon. As a first approach to this problem, we have searched for *ftz* homeodomain binding sites at the *Antp* locus. *Antp* P2 expression is regulated by *ftz* at blastoderm (Ingham & Martinez-Arias, 1986). Two genes of the BX-C, *Ultrabithorax* (*Ubx*) and *abdominal A* (*abdA*), repress expression of *Antp* in parasegments 5 to 13 (Hafen *et al.* 1984b; Carroll *et al.* 1986). These genes encode homeodomains that are as closely related to *ftz* as is *Dfd*. Therefore, based on the related DNA-binding specificity of *ftz* and *Dfd*, it is possible that binding sites for the *Ubx* and *abdA* proteins (including those that control expression of *Antp*) will also be recognized by the *ftz* protein.

Immunoprecipitation of restriction fragments bound to FHB4 protein in 200 mM-NaCl buffer was used to identify binding sites within 90 kb of *Antp* DNA that extends from 16 kb 5′ of the P1 transcriptional start to 7 kb 3′ of the P2 start site (Laughon *et al.* 1986; Schneuwly *et al.* 1986; Stroeher *et al.* 1986). Subclones of *Eco*RI or *Xba*I fragments spanning the region were screened twice, once cleaved into fragments with *Hin*fI and once with *Dde*I. Data for sequences in the immediate vicinity of the P1 and P2 transcriptional start sites are shown in Fig. 6 and the results from all of the experiments are summarized in Fig. 7. A total of eight *Eco*RI or *Xba*I subclones contained one or more *Hin*fI or *Dde*I fragments that were bound by FHB4 in 200 mM-NaCl buffer. For fragments within the sequenced regions surrounding the P1 and P2 starts sites, the exact positions of the bound fragments were deduced by restriction mapping. Six fragments containing *ftz* homeodomain binding sites are clustered over a 7 kb region surrounding the P1 start site whereas for P2 only one 200 bp region was found 550 bp 5′ of the start of transcription. Three fragments with binding sites were found in the 70 kb interval between the P1 cluster and the single P2 site. Three additional fragments were found between 11 and 17 kb 5′ of the P1 start for a total of 13 fragments containing binding sites in 90 kb of DNA.

R4 R2.2 R4.6

Fig. 6. Identification of *ftz* homeodomain binding sites near the *Antp* P1 and P2 promoters. Immunoprecipitation DNA binding assays using FHB4 extract were performed on two *Antp* subclones, 577R4 (R4) and 577R2.2 (R2.2), that contain DNA extending from 4 kb 5′ to 2 kb 3′ of the P1 start site, and one clone, 448R4.6 (R4.6) containing sequences from 3·3 kb 5′ to 1·3 kb 3′ of the P2 start site. Fragments bound in high-salt buffer appear in the 1st, 3rd and 5th lanes (left to right) and total labelled DNA appears in the 2nd, 4th and 6th lanes.

Discussion

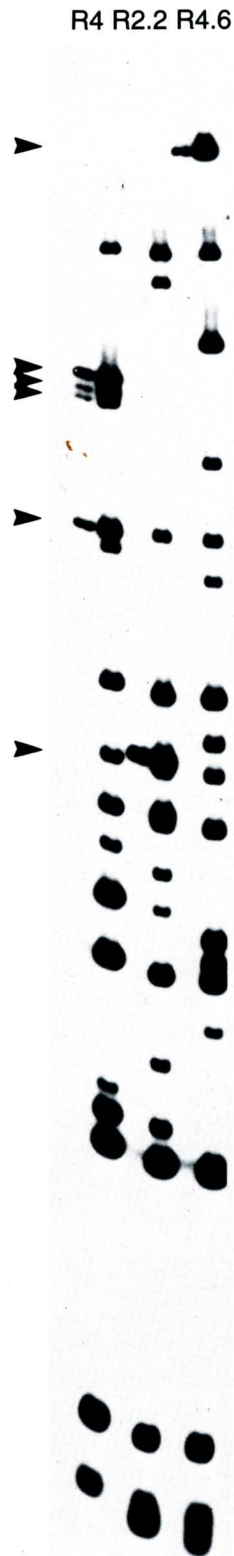The results of our experiments with the *ftz* and *Dfd* homeodomains are consistent with the DNA-binding properties described for the *en* homeodomain (Desplan *et al.* 1985) and generally support the notion that the conserved function of the homeodomain is to bind to specific DNA sequences. In both cases the homeodomains alone are sufficient for the sequence-specific binding activity. Our experiments indicate that none
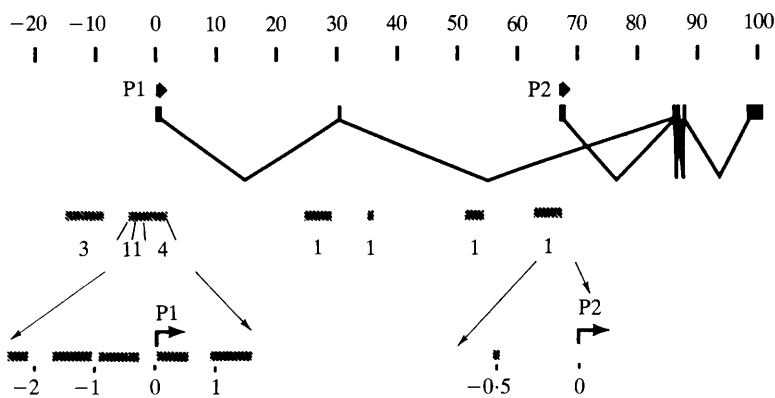
Fig. 7. Mapping of *ftz* homeodomain binding sites within *Antp*. Presented is a summary of immunoprecipitation experiments performed on DNA fragments extending from −15 to +75 kb, measured from the P1 transcriptional start. The positions of *Eco*RI subclones containing either *Hin*fI or *Dde*I fragments with high-affinity sites for FHB4 are shown as hatched bars with the number of high-affinity fragments just below. At the bottom, the regions around the P1 and P2 starts are shown expanded. *Antp* exons are drawn are solid bars and distances are in kilobases.

of the sequences on the amino terminal side of the *ftz* homeodomain contribute to the specificity of binding *in vitro*. However, binding specificity could be influenced by other proteins in a way that depends on sequences either within or outside of the homeodomain and we take our results to mean only that the amino terminal half of the *ftz* protein, (not all of the amino terminus of the *Dfd* protein was present in our longer fusion) outside of the homeodomain, is not involved in directly contacting DNA. It is also possible that the attachment of *ftz* or *Dfd* to *β*-galactosidase affected the ability of the *ftz* or *Dfd* nonhomeodomain sequences to function.

The high degree of sequence homology among ANT-C and BX-C homeodomains suggests that they could have closely related DNA-binding specificity (Laughon & Scott, 1984). The region that appears to correspond to the recognition helix in bacterial DNA-binding proteins is invariant among ANT-C and BX-C homeodomains. The closely related DNA-binding specificity of the *ftz* and *Dfd* proteins supports the model based on bacterial protein homologies and predicts that all of the seven highly conserved ANT-C and BX-C homeodomains (*Antp*, *ftz*, *Scr*, *Dfd*, *Ubx*, *abdA*, *AbdB*) will bind to DNA with similar specificity *in vitro*. Again, this does not mean that these proteins will bind to the same sequences *in vivo*, only that in the absence of other proteins the intrinsic specificity of their homeodomains should be very closely related. Of course these experiments only address the question of DNA-binding specificity and provide no information on other features that may be important for the function of these proteins in activating or repressing transcription.

The observed distribution of restriction fragments containing *ftz* homeodomain binding sites across 90 kb of sequence suggests that sites may be clustered around the P1 promoter but that the total number of high-affinity binding sites in the genome is probably very large. Assuming that sites are clustered near the *Antp* promoters, we can still use the three fragments

containing sites within the 70 kb between P1 and P2 to extrapolate to 6500 fragments per haploid gene. The number of individual binding sites could be even greater since we have found multiple homeodomain binding sites on a single fragment. Is it possible that the presumably limited number of genes that are regulated by *ftz* actually compete with such a large number of binding sites? Perhaps enough *ftz* protein is available for thousands of sites to be occupied. The other explanation is that functional binding sites are rare but have substantially higher binding affinities for *ftz* than the sites we have identified in the *Antp* gene. This issue will be investigated by testing the ability of the *ftz* homeodomain binding sites identified *in vitro* to confer transcriptional regulation by *ftz* on an exogenous gene.

References

AKAM, M. (1985). Segments, lineage boundaries and the domains of expression of homeotic genes. *Phil. trans. Roy. Soc. Lond.* B **312**, 179–187.

AKAM, M. (1987). The molecular basis for metameric pattern in the Drosophila embryo. *Development* **101**, 1–22.

CARROLL, S. B. & LAUGHON, A. (1987). Production and purification of polyclonal antibodies to the foreign segment of *β*-galactosidase fusion proteins. In *DNA Cloning III. A Practical Approach* (ed. D. M. Glover). Oxford: IRL Press.

CARROLL, S. B., LAYMON, R. A., MCCUTCHEON, M. A., RILEY, P. D. & SCOTT, M. P. (1986). Localization and regulation of *Antennapedia* protein expression in Drosophila embryos. *Cell* **47**, 113–122.

CARROLL, S. B. & SCOTT, M. P. (1985). Localization of the *fushi tarazu* protein during Drosophila embryogenesis. *Cell* **43**, 47–57.

DESPLAN, C., THEIS, J. & O'FARRELL, P. H. (1985). The *Drosophila* developmental gene, engrailed, encodes a sequence-specific DNA binding Activity. *Nature, Lond.* **318**, 630–635.

DINARDO, S. & O'FARRELL, P. H. (1987). Establishment and refinement of segmental pattern in the Drosophila embryo: spatial control of engrailed expression by pair-rule genes. *Genes Dev.* **1**, 1212–1225.

DOYLE, H. J., HARDING, K., HOEY, T. & LEVINE, M. (1986). Transcripts encoded by a homeo box gene are restricted to dorsal tissues of *Drosophila* embryos. *Nature, Lond.* **323**, 76–79.

DUNCAN, I. M. (1986). Control of bithorax complex functions by the segmentation gene *fushi tarazu* of Drosophila melanogaster. *Cell* **47**, 297–309.

FRIGERIO, G., BURRI, M., BOPP, D., BAUMGARTNER, S. & NOLL, M. (1986). Structure of the segmentation gene paired and the Drosophila PRD gene set as part of a gene network. *Cell* **47**, 735–746.

FROHNHÖFER, H. G. & NÜSSLEIN-VOLHARD, C. (1986). The organization of anterior pattern in the *Drosophila* embryo by the maternal gene bicoid. *Nature, Lond.* **324**, 120–125.

HAFEN, E., KUROIWA, A. & GEHRING, W. J. (1984a). Spatial distribution of transcripts from the segmentation gene *fushi tarazu* of Drosophila. *Cell* **37**, 825–831.

HAFEN, E., LEVINE, M. & GEHRING, W. J. (1984b). Regulation of Antennapedia transcript distribution by the *bithorax* complex in Drosophila. *Nature, Lond.* **307**, 287–289.

HIROMI, Y. & GEHRING, W. J. (1987). Regulation and function of the Drosophila segmentation gene fushi tarazu. *Cell* **50**, 963–974.

HOWARD, K. & INGHAM, P. (1986). Regulatory interactions between the segmentation genes *fushi tarazu, hairy*, and *engrailed* in the Drosophila blastoderm. *Cell* **44**, 949–957.

INGHAM, P. W. & MARTINEZ-ARIAS, A. (1986). The correct activation of Antennapedia and bithorax complex genes requires the *fushi tarazu* gene. *Nature, Lond.* **324**, 592–597.

KAUFMAN, T. C. (1983). The genetic regulation of development in Drosophila. In *Time, Space and Pattern in Embryonic Development*, pp. 365–383. New York: Alan R. Liss.

LAUGHON, A., BOULET, A. M., BERMINGHAM, J. R., LAYMON, R. A. & SCOTT, M. P. (1986). The structure of transcripts from the homeotic *Antennapedia* gene of Drosophila: Two promoters control the major protein-coding region. *Molec. cell. Biol.* **6**, 4676–4689.

LAUGHON, A., CARROLL, S. B., STORFER, F. A., RILEY, P. D. & SCOTT, M. P. (1985). Common properties of proteins encoded by the Antennapedia complex genes

of *Drosophila melanogaster. Cold Spring Harbor Symp. quant. Biol.* **50**, 253–262.

LAUGHON, A. & SCOTT, M. P. (1984). Sequence of a *Drosophila* segmentation gene: protein structure homology with DNA-binding proteins. *Nature, Lond.* **310**, 25–31.

LEWIS, E. B. (1978). A gene complex controlling segmentation in *Drosophila. Nature, Lond.* **276**, 565–570.

MARSTON, F. A. O. (1987). The purification of eukaryotic polypetides expressed in Escherichia coli. In *DNA Cloning III. A Practical Approach* (ed. D. M. Glover). Oxford: IRL Press.

MCGINNIS, W., LEVINE, M., HAFEN, E., KUROIWA, A. & GEHRING, W. J. (1984). A conserved DNA sequence in homeotic genes of the *Drosophila* Antennapedia and bithorax complexes. *Nature, Lond.* **308**, 428–433.

NÜSSLEIN-VOLHARD, C. & WIESCHAUS, E. (1980). Mutations affecting segment number and polarity in Drosophila. *Nature, Lond.* **287**, 795–801.

PABO, C. O. & SAUER, R. T. (1984). Protein–DNA recognition. *A. Rev. Biochem.* **53**, 293–321.

REGULSKI, M., MCGINNIS, N., CHADWICK, R. & MCGINNIS, W. (1987). Developmental and molecular analysis of Deformed; a homeotic gene controlling Drosophila head development. *EMBO J.* **6**, 767–777.

RÜTHER, V. & MÜLLER-HILL, B. (1983). Easy identification of cDNA clones. *EMBO J.* **2**, 1791–1794.

SCHNEUWLY, S., KUROIWA, A., BAUMGARTNER, P. & GEHRING, W. J. (1986). Structural organization and sequence of the homeotic gene *Antennapedia of Drosophila melanogaster. EMBO J.* **5**, 733–739.

SCOTT, M. P. & CARROLL, S. B. (1987). The segmentation and homeotic gene network in early Drosophila development. *Cell* **51**, 689–698.

SCOTT, M. P. & WEINER, A. J. (1984). Structural relationships among genes that control development: Sequence homology between the *Antennapedia, Ultrabithorax* and *fushi tarazu* loci of Drosophila. *Proc. natn. Acad. Sci. U.S.A.* **81**, 4115–4119.

SHEPHERD, J. C. W., MCGINNIS, W., CARRASCO, A. E., DEROBERTIS, E. D. & GEHRING, W. J. (1984). Fly and frog homoeo domains show homologies to yeast mating type regulatory proteins. *Nature, Lond.* **310**, 70–71.

STROEHER, V. L., JORGENSEN, E. M. & GARBER, R. L. (1986). Multiple transcripts from the *Antennapedia* gene of *Drosophila. Molec. cell. Biol.* **6**, 4667–4675.

WAKIMOTO, B. T. & KAUFMAN, T. C. (1981). Analysis of larval segmentation in lethal genotypes associated with the Antennapedia gene complex in Drosophila melanogaster. *Devl Biol.* **81**, 51–64.

WAKIMOTO, B. T., TURNER, F. R. & KAUFMAN, T. C. (1984). Defects in embryogenesis in mutants associated with the Antennapedia gene complex of *Drosophila melanogaster. Devl Biol.* **102**, 147–172.