The EMBO Journal vol.9 no.12 pp.3967-3974, 1990

The interaction with DNA of wild-type and mutant *fushi tarazu* homeodomains

Anthony Percival-Smith[1], Martin Müller, Markus Affolter and Walter J.Gehring

Biozentrum, University of Basel, Department of Cell Biology, Klingelbergstr. 70, CH-4056 Basel, Switzerland

[1]Present address: Banting and Best Institute for Medical Research, University of Toronto, Toronto, Ontario M5G 1L6, Canada

Communicated by W.J.Gehring

The *in vitro* DNA binding properties of wild-type and mutant *fushi tarazu* homeodomains (*ftz* HD) have been analysed. The DNA binding properties of the *ftz* HD are very similar to those of the *Antp* HD. In interference experiments with mutant *ftz* HDs, close approaches between specific portions of the *ftz* HD peptide and specific regions of the binding site DNA were mapped. A methylation interference, G7 on the β strand of BS2, is absent from the interference pattern with a mutant *ftz* HD [*ftz* (R43A) HD] in which the Arg43 at the second position of helix III (the recognition helix) is replaced by an Ala. This indicated that Arg43 of the *ftz* HD is in close proximity to the N7 of G7 of the β strand of BS2 in the major groove. The methylation and ethylation interference patterns with the *ftz* (NTD) HD, in which the first six amino acids of the homeodomain were deleted, were extensively altered relative to the *ftz* HD patterns. Methylation of A11 and G12 of the α strand and ethylation of the phosphate of nucleotide A12 of the α strand no longer interfere with binding. This indicated that the first six amino acids of the homeodomain of *ftz* interact with A11 of the α strand in the minor groove, the phosphate of the nucleotide A13 on the α strand and G12 of the α strand in the adjacent major groove of BS2. In a binding study using a change of specificity mutation [*ftz* (Q50K) HD], in which the Gln50 at the ninth position of the third helix is exchanged for a Lys (as in the *bicoid* HD), and variant binding sites, we concluded that position 50 of the *ftz* HD and the *ftz* (Q50K) HD peptides interacts with base pairs at positions 6 and 7 of BS2. These three points of contact allowed us to propose a crude orientation of the *ftz* HD within the protein-DNA complex. We find that the *ftz* HD and the *Antp* HD peptides contact DNA in a similar way.

Key words: DNA binding proteins/*fushi tarazu* homeodomains/*Drosophila*/protein-DNA interaction

Introduction

The homeobox is a 180 bp DNA sequence that encodes a 60 amino acid protein domain, the homeodomain. The homeobox was first identified as a sequence present in a number of *Drosophila* developmental regulatory genes (McGinnis *et al.*, 1984a,b; Scott and Weiner, 1984; Gehring, 1987; Scott *et al.*, 1989). Subsequently, the homeobox was identified in vertebrate genomes, which indicates extensive evolutionary conservation, and possible functional conservation (McGinnis *et al.*, 1984b). Recent screening of the *Caenorhabditis elegans* genome for homeobox-containing genes has indicated that ~60 genes, ~1% of genes of *C.elegans*, contain homeoboxes (Bürglin *et al.*, 1989). This makes the homeodomain a very common protein domain. The homeobox has also been found in genes that encode transcription factors (for reviews, see Levine and Hoey, 1988; Scott *et al.*, 1989; Affolter *et al.*, 1990b). *In vitro* DNA binding assays have shown that the homeodomain is a DNA binding domain of homeodomain-containing proteins (Desplan *et al.*, 1988; Hoey and Levine, 1988; Müller *et al.*, 1988).

Amino acid sequence comparison of the homeodomain with prokaryotic DNA binding proteins indicated that the homeodomain might contain a helix-turn-helix motif (Laughon and Scott, 1984; Shephard *et al.*, 1984). The determination of the three-dimensional structure of the *Antp* homeodomain by nuclear magnetic resonance (NMR) spectroscopy demonstrated the existence of the postulated helix-turn-helix motif (Otting *et al.*, 1988; Qian *et al.*, 1989; Billeter *et al.*, 1990; Güntert *et al.*, 1990; for a review see Gehring *et al.*, 1990). Structure determination of several prokaryotic DNA binding protein-DNA complexes has shown that the N-terminal residues in the second helix of the helix-turn-helix motif make a number of specific amino acid-base pair interactions (Pabo *et al.*, 1990). However, recent genetic results imply a direct involvement in DNA binding for different amino acid residues of the helix-turn-helix motif of the homeodomain than might have been anticipated relative to prokaryotic DNA binding proteins. In particular, a single amino acid substitution at the ninth residue of helix III switched the DNA binding specificities of different homeodomains (Hanes and Brent, 1989; Treisman *et al.*, 1989). Also in studies of the DNA binding properties of *Antp* HD, it was shown that *Antp* HD bound with a high affinity to a site of 10-12 bp as a monomer (Affolter *et al.*, 1990a). These data suggested that the *Antp* HD and prokaryotic repressor DNA binding domains interact with DNA in a substantially different manner.

In this paper, we present data on the interaction of *ftz* HD peptides with DNA. We used this data to orient the homeodomain in the complex relative to DNA and also compared it with the structure of the *Antp* HD-DNA complex determined by NMR spectroscopy (Otting *et al.*, 1990).

Results

The ftz HD binds to DNA in a similar manner to the Antp HD

The homeodomain (HD) used in this study was derived from the *Drosophila fushi tarazu* (*ftz*) gene (see Materials and methods) and expressed in a T7 expression vector (Studier

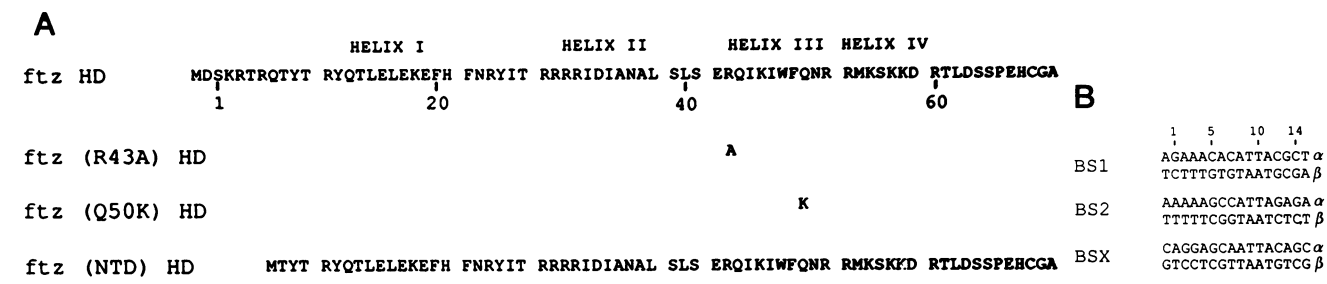

Fig. 1. (**A**) Amino acid sequences of the wild-type *ftz* HD and the three mutant peptides used in this study are shown. Helices I–IV are indicated and the peptide sequence is numbered according to Qian *et al.* (1989). *ftz* (R43A) and *ftz* (Q50K) are single amino acid exchanges in the third helix of the homeodomain. In *ftz* (R43A) Arg43 is replaced by an Ala and in *ftz* (Q50K) Gln50 is replaced by a Lys. *ftz* (NTD) is a deletion mutant which misses the first seven amino acids present in the *ftz* HD peptide. (**B**) The sequences of the three *in vitro* defined binding sites BS1, BS2 and BSX used for the ethylation and methylation interference experiments shown in Figure 2 are indicated. The individual bases are labelled according to the nomenclature of Otting *et al.* (1990). The upper strand of each binding site is the α strand (5′–3′), the lower is the β strand (3′–5′). The bases of the α and the β strand are numbered from left to right from 1 to 14.

and Moffat, 1986). The transcript encodes a 73 amino acid peptide that consists of, from the N- to the C-termini, a starting methionine derived from the expression plasmid, an Asp residue derived from the *ftz* gene, the 60 amino acid long *ftz* homeodomain, and an 11 amino acid C-terminal tail derived from the *ftz* gene (Figure 1A). This peptide is referred to as the *ftz* HD throughout the text. The *ftz* HD was overexpressed and purified in a similar way to the *Antp* HD containing peptide (Müller *et al.*, 1988). Since the results of this study will be discussed relative to a structure of an *Antp* HD DNA complex determined by NMR (Otting *et al.*, 1990), it was first necessary to show that the *ftz* HD and the *Antp* HD interact with the binding sites in a similar manner.

We have described previously the DNA binding properties of an *Antp* HD (Affolter *et al.*, 1990a) used for the NMR structure determination of an *Antp* HD–DNA complex (Otting *et al.*, 1990). The first property of the *ftz* HD and the *Antp* HD we compared were their equilibrium dissociation constants (K_D). These were determined in saturation experiments using the oligodeoxy nucleotide BS2-18 (see Materials and methods) as the ligand. The K_D of the *ftz* HD–DNA interaction was determined in two independent experiments as $(6.0 \pm 1.0) \times 10^{-10}$ M and $(8.0 \pm 3.0) \times 10^{-10}$ M at 20°C. The K_D for the *ftz* HD–DNA interaction is only 2-fold lower than the *Antp* HD K_D of 1.6×10^{-9} M (Affolter *et al.*, 1990a).

To determine whether the *ftz* HD was interacting with the target DNA in a different way relative to the *Antp* HD, ethylation and methylation interference experiments were performed using the *ftz* HD, and the results compared with those of a previous study using the *Antp* HD (Affolter *et al.*, 1990a). In an interference experiment a 3′ end labelled DNA fragment containing a binding site was modified either with ethyl nitrosourea (ENU) for an ethylation interference experiment, or with dimethyl sulphate (DMS) for a methylation interference experiment. The modified DNA was then incubated with the *ftz* HD in a binding reaction and the reaction electrophoresed through a native polyacrylamide gel to separate the *ftz* HD–DNA complex from the free DNA. The bound DNA and the unbound DNA were isolated, chemically cleaved at the modifications and run on a denaturing polyacrylamide gel. Modifications that inhibit *ftz* HD interactions with DNA were visualized on the autoradiogram of the sequencing gel as a reduction in intensity or a loss of bands in the 'bound' lane relative to the 'unbound' lane.

The *ftz* HD ethylation interference experiment was performed with DNA fragments containing the BS2 binding site, as had been done for the *Antp* HD (Affolter *et al.*, 1990a). Phosphate triester formation with ethyl nitrosourea adds an ethyl group to the sugar phosphate backbone and removes a negative charge. Hence, an inhibition of *ftz* HD binding can be due to either steric interference, or loss of an electrostatic interaction between the *ftz* HD and a charged phosphate. On the α strand of BS2, ethylation of the phosphates of the nucleotides A4, G5, C6, C7, A8 and A13 interfered with *ftz* HD binding; on the β strand, ethylation of the phosphates of the nucleotides G7, T8, A9, A10, T11 and C12 interfered with *ftz* HD binding (Figure 2A). The positions that interfere with *Antp* HD binding also interfered with *ftz* HD binding (Affolter *et al.*, 1990a). However, there were two minor differences: ethylation of the phosphate of the nucleotide C12 on the β strand interfered weakly with *ftz* HD binding, but not *Antp* HD binding; and ethylation of the phosphate of the nucleotide A13 on the α strand interfered strongly with *ftz* HD binding, but only weakly with *Antp* HD binding.

We performed methylation interference experiments with the *ftz* HD on three binding sites: BS1 and BS2 from the *engrailed* (*en*) promoter (Müller *et al.*, 1988), and BSX from the *ftz* enhancer (Pick *et al.*, 1990). As described previously (Affolter *et al.*, 1990a), three binding sites were used to increase the surface area of the DNA assayed for methylation interferences. On BS2, methylations at A3, A4, G5, A11, G12 on the α strand and G7, A9, A10 on the β strand interfered with *ftz* HD binding (Figure 2B), as was found with the *Antp* HD. On BS1, methylations at A3, A4, A11 on the α strand, and G5, G7, A9, A10, G12 on the β strand interfered with *ftz* HD binding (Figure 2C). Methylation of A3 on the α strand and G12 on the β strand did not interfere with *Antp* HD binding, but did interfere with *ftz* HD binding. Methylation at G13 on the α strand of BS1 did interfere with *Antp* HD binding, but not with *ftz* HD binding. On BSX, methylations at G5 and A11 on the α strand, and G6, A9, A10, G12 on the β strand interfered with *ftz* HD binding (Figure 2D). Methylation of G12 on the β strand did not interfere with *Antp* HD binding, but methylation of G13 on the α strand did interfere with *Antp* HD binding. All these changes in the methylation interference patterns of *Antp* HD

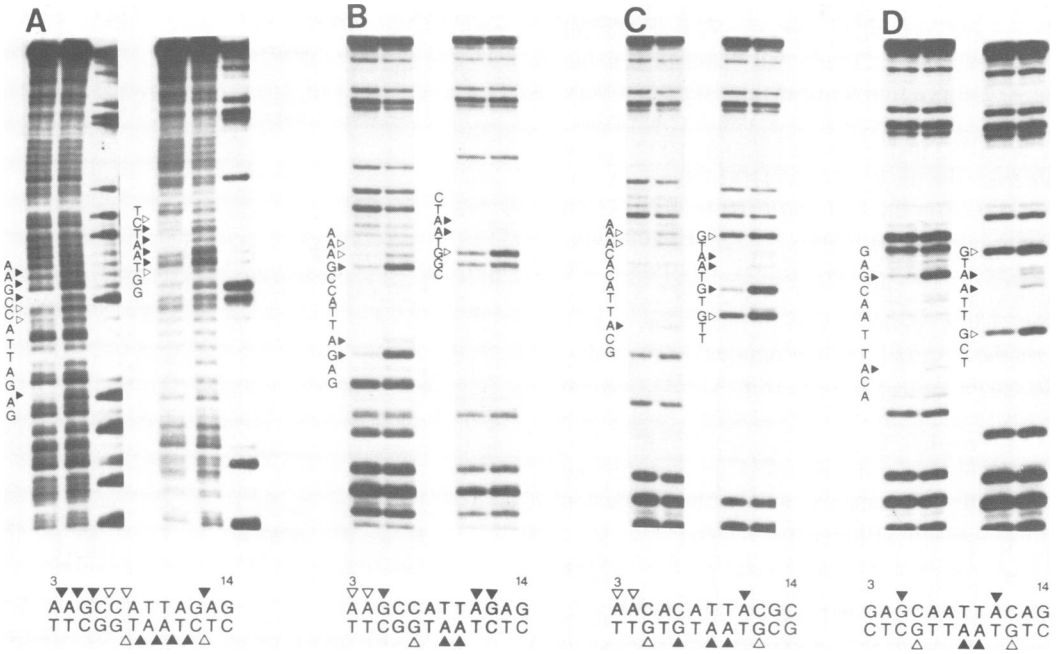

Fig. 2. Ethylation and methylation interference experiments with the *ftz* HD. From left to right in the autoradiograms of the eight sequencing gels shown, *Hin*dIII−*Xba*I DNA fragments (3′ end-labelled at the *Hin*dIII site) were isolated from pBS2-26a, pBS2-26b, pBS2-26a, pBS2-26b, pBS1-26a, pBS1-26b, pBSX-18a and pBSX-18b. (**A**) DNA modified with ethyl nitrosourea. (**B**−**D**) DNA modified with dimethyl sulphate. The left-hand gel in each panel is an interference experiment on the α strand of DNA, and the right-hand gel in each panel is an interference experiment on the β strand of DNA. The first lane of each gel is the DNA isolated from the bound complex, and the second lane is the unbound DNA. The third lane in the gels shown in A is the marker lane of dimethyl sulphate treated DNA. The sequences of regions where interferences occur are shown on the left of each gel. The modified phosphates or bases that interfere strongly are indicated by closed arrowheads and the ones that interfere weakly are indicated by open arrowheads. The summary of the interference pattern is shown below the autoradiograms.

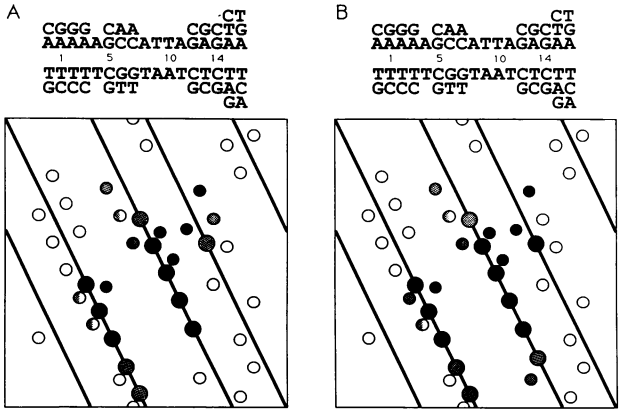

Fig. 3. Projection diagram summaries of the ethylation and methylation interference data for both *Antp* HD (**A**) and *ftz* HD (**B**). The diagrams represent the projection of the chemical groups onto a cylindrical surface, which is then cut along its *Z* axis and folded out. The sequences at the top are that of BS2 with the different bases present in BS1 and BSX above and below it. The shared 5′-ATTA-3′ motif was used to align the three sites relative to one another. The ethylation interferences on BS2 are indicated as the larger shaded circles along the sugar phosphate backbone (the diagonal lines). The positions of the N-3 of the adenines and the N-7 of the guanines that were assayed for methylation interference are indicated in the minor and major grooves, respectively, by the smaller circles. Positions that did not interfere in the methylation studies are not filled in. Phosphates that did not interfere when ethylated are not shown. Weak interferences and strong interferences are indicated by stippled circles and solid circles, respectively. Interferences that behave differently, depending on the binding site the group is in, are indicated as half-filled circles.

and *ftz* HD are weak interferences. Examination of the projection diagrams of the ethylation and methylation interferences of *Antp* HD and *ftz* HD shows that both proteins interact with DNA in a similar manner (Figure 3). Any differences that occur are at the periphery of the interfering region.

Interference patterns of mutant ftz HD proteins

A way of mapping close approaches between a DNA binding protein and DNA is to examine the interference pattern of mutant derivatives that still retain some DNA binding activity. The idea behind this approach is that a mutant protein may be able to accommodate a modification into the complex that was not accommodated with the wild-type protein may be able to accommodate a modification into the complex that was not accommodated with the wild-type protein, or a mutant protein may no longer contact the base or sugar phosphate backbone that when modified interfered with binding of the wild-type protein. We examined the DNA binding properties of a number of *ftz* HD mutations and found two that gave altered interference patterns.

Purified *ftz* (R43A) HD gave an altered methylation interference pattern on BS2. The amino acid sequence of the *ftz* (R43A) HD is identical to that of the *ftz* HD except for the exchange of an Arg43 for an Ala at the second position of helix three of the homeodomain (Figure 1). This exchange increases the K_D for BS2 relative to the *ftz* HD ~2-fold to $(1.4 \pm 0.2) \times 10^{-9}$ M. This increase in the K_D does not seem to be deleterious *in vivo*, because a *ftz*

gene harbouring the same mutation, when reintroduced into the *Drosophila* genome, rescued in a *ftz* rescue assay (Hiromi *et al.*, 1985; Müller, 1990). These data indicate that the presence of an Arg at position 43 of the *ftz* HD is not very important for *ftz* activity. However, in an interference assay, the methylations of BS2 that interfered with *ftz* HD binding (Figure 4A) interfered with *ftz* (R43A) HD binding, except for methylation of G7 on the β strand (Figure 4B). G7 on the β strand of BS2 interfered weakly with *ftz* HD binding, but not with *ftz* (R43A) HD binding. We interpreted the result in the following way. The *ftz* HD−BS2 DNA complex cannot accommodate a methyl group in the major groove on N7 of G7 on the β strand, but the removal of the bulky Arg and replacement with Ala in *ftz* (R43A) HD resulted in the formation of a protein−DNA complex that could accommodate this methyl group. This would mean that Arg43 and the N7 of G7 on the β strand are close to one another in the *ftz* HD−BS2 complex, and that this interaction would occur in the major groove of B-DNA.

The *ftz* (NTD) HD gave an extensively altered interference pattern on BS2. The amino acid sequence of the *ftz* (NTD) HD is identical to that of the *ftz* HD except that amino acids −1 to 6 were deleted fusing the starting methionine to position 7 of the *ftz* HD (Figure 1) (NTD, N-terminal deletion). This removes a portion of the homeodomain that was found with the *Antp* HD to be unstructured in solution (Qian *et al.*, 1989). The methylation interference pattern on BS2 of the *ftz* (NTD) HD did not include interferences at A11 and G12 on the α strand (Figure 5B), as found with the *ftz* HD (Figure 5A). The ethylation interference pattern on BS2 of the *ftz* (NTD) HD differed from that of the *ftz* HD by the loss of the interference with binding on ethylation of the phosphate of A13 on the α strand (Figure 6). Methylation interference at A11 on the α strand of BS1 was also not observed with the *ftz* (NTD) HD (Percival-Smith and Gehring, unpublished observations). As expected if these interferences represented protein−DNA close contacts, the K_D of the *ftz* (NTD) HD for BS2 increased 130-fold, relative to the K_D of the *ftz* HD for BS2, to (1.14 ± 0.49) × 10^{-7} M. These data indicate that the N-terminus of the homeodomain may become structured on complex formation making interactions on the α strand of BS2 with A11 in the minor groove, the sugar phopshate backbone around G12 and A13, and G12 in the adjacent major groove.

Binding studies with the ftz HD relative to the ftz (Q50K) HD

It has been demonstrated previously that position 50 of the homeodomain, commonly referred to as position 9 of the third helix, is an important determinant of DNA binding specificity. Changing the lysine at position 9 in the *bicoid* homeodomain to glutamine, as found in the *Antp* homeodomain at position 9, changes the specificity of *bicoid* dependent transcription activation, in yeast, from reporter gene fusions containing *bicoid* binding sites to a reporter gene fusion containing *Antp* binding sites (Hanes and Brent, 1989). In addition, *in vitro* DNA binding experiments with *paired* protein showed that when the serine at position 9 of the *paired* homeodomain was changed to a glutamine, as found in the *ftz* homeodomain, or to a lysine, as found in the *bicoid* homeodomain, the modified *paired* proteins specifically recognized *ftz* binding sites, or *bicoid* binding sites, respectively (Treisman *et al.*, 1989). Change of specificity muta-

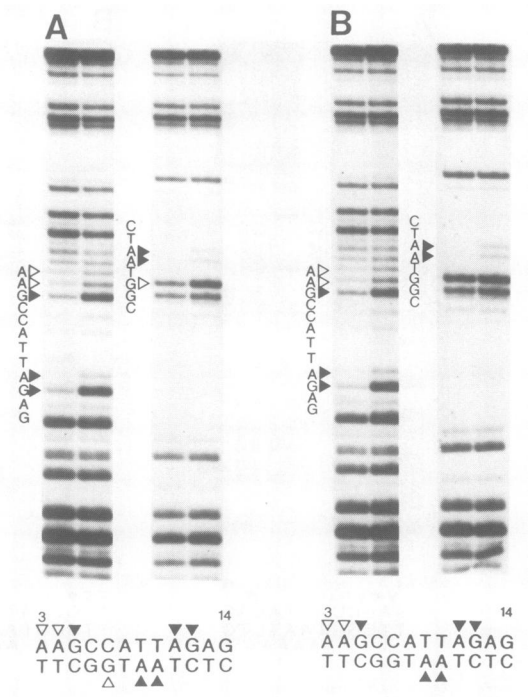

Fig. 4. Methylation interference experiments with the *ftz* HD (**A**) and the *ftz* (R43A) HD (**B**) on BS2. The left-hand gel in each panel is an interference experiment on the α strand of BS2 and the right-hand gel on the β strand of BS2. The first lane of each gel is the DNA isolated from the bound complex; the second lane is the unbound DNA. The summary of the interference patterns for the *ftz* HD and the *ftz* (R43A) HD are shown below the autoradiograms. The bases that interfere strongly are indicated by closed arrowheads, and weakly by open arrowheads.

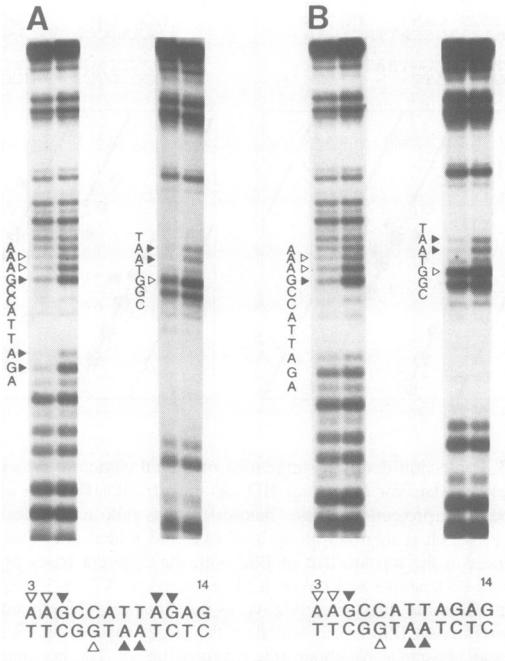

Fig. 5. Methylation interference experiments with the *ftz* HD (**A**) and the *ftz* (NTD) HD (**B**) on BS2. The left-hand gel in each panel is an interference experiment on the α strand of BS2, and the right-hand gel of the β strand of BS2. The first lane of each gel is the DNA isolated from the bound complex; the second lane is the unbound DNA. The summary of the interference patterns for the *ftz* HD and the *ftz* (NTD) HD are shown below the autoradiograms. The bases that interfere strongly are indicated by closed arrowheads and weakly by open arrowheads.

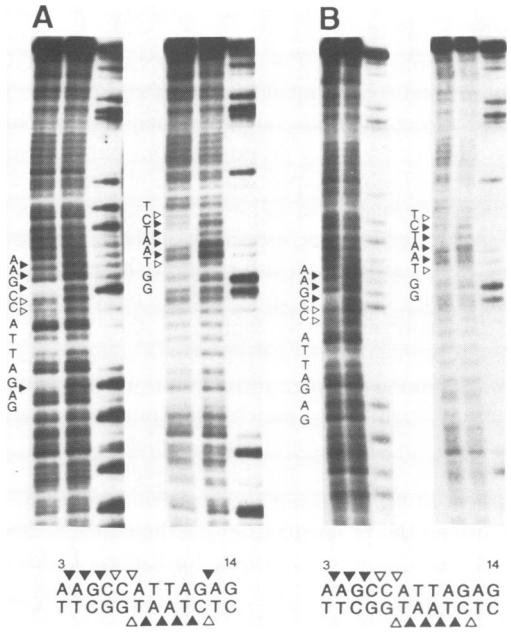

Fig. 6. Ethylation interference experiments with *ftz* HD (**A**) and the *ftz* (NTD) HD (**B**) on BS2. The left-hand gel in each panel is an interference experiment on the α strand of BS2, and the right-hand gel on the β strand of BS2. The first lane of each gel is the DNA isolated from the bound complex; the second lane is the unbound DNA. The third lane is the marker lane of dimethyl sulphate treated DNA. The summary of the interference patterns for the *ftz* HD and the *ftz* (NTD) HD are shown below the autoradiograms. The bases that interfere strongly are indicated by closed arrowheads and weakly by open arrowheads.

tions of prokaryotic DNA binding proteins have been used to define specific amino acid–base pair interactions (Wharton and Ptashne, 1987; Bass *et al.*, 1988; Lehming *et al.*, 1990).

In this study, we partially purified a homeodomain, the *ftz* (Q50K) HD, which has an identical amino acid sequence to the *ftz* HD, except that the glutamine at position 50 (position 9 of helix III) of the homeodomain has been changed to a lysine, as found in the *bicoid* homeodomain at position 50 (Berleth *et al.*, 1988). We compared the affinities of the *ftz* HD and the *ftz* (Q50K) HD for a variety of binding sites to determine the important base pairs required for differential recognition by *ftz* HD and *ftz* (Q50K) HD. The number of binding sites that needed to be analysed was simplified by the observation that the sequence of the *bicoid* consensus binding site in GGGATTAGA (Driever and Nüsslein-Volhard, 1989) and the sequence of the BS2 binding site from position 5 to 13 is GCCATTAGA differing only at positions 6 and 7 (GG versus CC) (Müller *et al.*, 1988). Therefore, we synthesized variant BS2 binding sites with all different combinations of G-C base pairs at position 6 and 7 (see Table I, and Materials and methods). Saturation experiments were used to determine the K_D for the *ftz* HD and the *ftz* (Q50K) HD using the four double-stranded deoxyoligonucleotides as ligands (Table I). The K_Ds of the *ftz* HD and *ftz* (Q50K) HD for the four double-stranded deoxynucleotides can be arranged in an almost reciprocal manner: the *ftz* HD K_Ds are BS2 < BS2v1 ≤ BS2v2 << BS2v3; whereas *ftz* (Q50K) HD K_Ds are BS2v3 = BS2v2 < BS2v1 << BS2 (Table I). However, the only drastic

Table I. K_Ds of the interaction of *ftz* HD and *ftz* (Q50K) HD with a number of binding sites

		ftz HD	*ftz* (Q50K) HD
BS2	GCCATTA[a]	0.84 ± 0.28 nM 1[b]	41.0 ± 11.0 nM 65
BS2v1	GCGATTA	1.6 ± 0.3 nM 2	3.7 ± 0.8 nM 6
BS2v2	GGCATTA	3.1 ± 0.8 nM 4	0.44 ± 0.12 nM 0.7
BS2v3	GGGATTA	131.0 ± 29.0 nM 160	0.63 ± 0.15 1[c]

[a]Positions 5–11 of the α strand of the binding sites used are indicated.
[b]The relative K_Ds for the *ftz* HD, indicated below the K_D, are relative to the K_D of the *ftz* HD–BS2 complex.
[c]The relative K_Ds for the *ftz* (Q50K) HD are relative to the K_D of the *ftz* (Q50K) HD–BS2v3 complex.

differences in K_D of ~100-fold are observed when both base pairs are changed, i.e. CC <-> GG, at position 6 and 7. A surprising result was that the increase in K_Ds of the binding sites with single base pair changes (BS2v1 and BS2v2) do not add up to the increase seen between the binding sites with the double base pair changes (BS2 and BSv3) with *ftz* HD or *ftz* (Q50K) HD. This indicates that two base pairs are responsible for the low affinities: C6, C7 for the *ftz* (Q50K) HD and G6, G7 for the *ftz* HD. Hence, the amino acid at position 9 in *ftz* (Q50K) HD and *ftz* HD interacts with the base pairs at position 6 and 7.

Discussion

The *Antp* HD used to determine the structure of a homeodomain–DNA complex is a mutant peptide (Güntert *et al.*, 1990; Otting *et al.*, 1990); relative to normal *Antp*, it contains a substitution of a serine for a cysteine at position 39 of the homeodomain to prevent oxidative dimerization of the protein (Müller *et al.*, 1988; Affolter *et al.*, 1990a). This mutant protein we had called the *Antp* HD (Affolter *et al.*, 1990a). The *ftz* HD and the *Antp* HD differ in amino acid sequence at 10 positions within the homeodomain. Although the two homeodomain peptides differ somewhat in amino acid sequence, they both interact with the same binding sites, apart from a few minor differences, in a similar manner (Figure 3). We take this to indicate that the free protein and protein–DNA complexes of *ftz* HD and *Antp* HD are very similar in structure.

In Figure 7, we have summarized the amino acid–base pair close approaches that can be proposed from our data. We have assumed that the structure of the *ftz* homeodomain does not drastically alter in structure on binding B-DNA (see also Otting *et al.*, 1990). Arg43 of the *ftz* HD was placed in the vicinity of N7 of guanosine 7 on the β strand in the major groove. From the analysis of the change of specificity mutation, position 50 was placed as interacting with base pairs 6 and 7. In addition, the first six amino acids of the homeodomain were placed on BS2 in the vicinity of A11 of the α strand in the minor groove, the sugar phosphate backbone and G12 of the α strand in the major groove. These three points of contact allowed us to orient the homeodomain relative to the DNA in the complex. Starting with Arg43 in the vicinity of the N7 of G7 on the β strand puts the N-terminal portion of the third helix in the major groove. Since

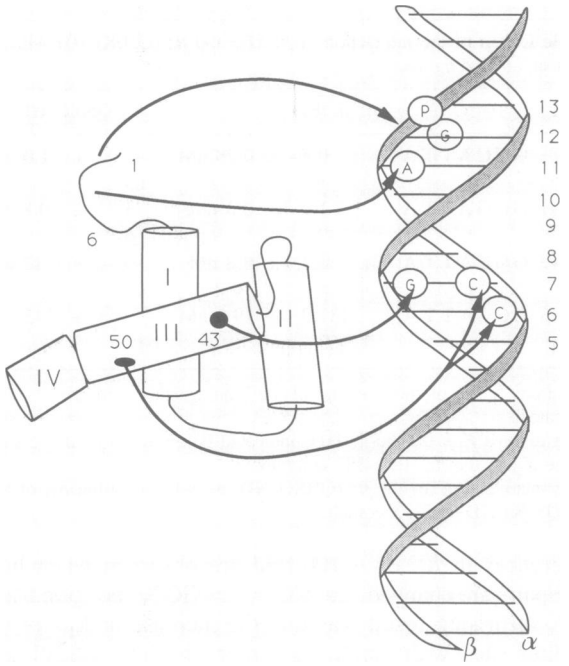

Fig. 7. Summary of the three points of contact between the *ftz* HD and DNA. On the left is a schematic representation of the *ftz* homeodomain and on the right is a schematic representation of B-DNA (BS2). We have assumed that the *Antp* HD (Qian *et al.*, 1989) and the *ftz* HD are similar in structure. The four helixes of the HD are indicated by roman numerals. Positions 1–6 that are deleted in *ftz* HD (NTD) HD, position 43 that is changed in *ftz* (R43A) HD, and position 50 that is changed in *ftz* (Q50K) are indicated on the homeodomain. The numbers alongside the DNA indicate positions of the binding site as described in Otting *et al.* (1990). The positions on DNA that are of interest are indicated by circles. The arrows between the HD and DNA indicate close contacts. It is important to note that for the complex to form, the HD has to be rotated by 180° around a vertical axis in order to dock it against the DNA. This rough model is a good agreement with the structural data obtained by NMR spectroscopy (Otting *et al.*, 1990).

position 50 (position 9 of the third helix) interacts along the base pairs 6 and 7, helix III and IV must be near perpendicular to the DNA axis and oriented from left to right across the major groove of B DNA (see Figure 7). This results in the N-terminus of the homeodomain being in the region of A11 of the α strand in the minor groove. The data are mutually consistent with one another; from any two experimentally determined contact points the third can be derived.

The orientation of the *ftz* homeodomain in the complex proposed from our biochemical and genetic data is in agreement with the orientation determined independently in an NMR investigation of an *Antp* HD–BS2 DNA complex (Otting *et al.*, 1990). In the NMR study, a number of intermolecular contacts between the *Antp* HD and BS2 were identified, which were then used to dock the *Antp* HD against an idealized B DNA model of BS2. In the NMR structure, position 43 is also in the vicinity of G7 of the β strand. A hydrogen of the γCH_2 of Gln50 of the *Antp* HD makes an intermolecular NOE with a ring hydrogen of C7 of the α strand, a region of DNA we had placed position 50 from our data. An intermolecular NOE was identified between Arg5 of the *Antp* HD and the 1′ sugar proton of G12 on the α strand. This indicates that Arg5 reaches in the minor groove. Position 5 of the homeodomain was deleted in the *ftz* (NTD) HD, and as would be expected from the NMR data, interferences around G12 of the α strand were lost with this mutant. It is also interesting to note that all the amino acid residues of *Antp* that have so far been determined to make intermolecular NOEs with DNA are conserved between *ftz* and *Antp* homeodomains (Otting *et al.*, 1990). The close agreement between the orientation of the homeodomain relative to DNA proposed from our data and the NMR data highlights the potential of combined biochemical and genetic approaches in gaining structural information on DNA–protein complexes, especially on those complexes for which direct structural information based on X-ray crystallography or NMR spectroscopy is not available.

The N-terminal amino acids of the homeodomain interacting with DNA is an interesting observation. This suggests that upon complex formation some or all of the six first amino acids of the homeodomain, which are unstructured when uncomplexed (Billeter *et al.*, 1990), become more defined in the complex. This is very reminiscent of the N-terminal arm of the λ repressor which wraps around the DNA making specific contacts down the major groove (Pabo *et al.*, 1982; Jordan and Pabo, 1988). A possible important contact for specificity of binding could be made between an amino acid of the ftz HD N-terminus and the 3′ most base pair of the ATTA motif found in most *ftz* and *Antp* binding sites (Desplan *et al.*, 1988; Müller *et al.*, 1988; Pick *et al.*, 1990). This is proposed because of the loss of the methylation interference at A11 in the minor groove on the α strand of BS2 with the *ftz* (NTD) HD, and the difficulty to imagine from the model of the complex, a side chain from helix III or helix IV reaching the base pair at position 11.

All the mutant *ftz* HDs used in this study retained some DNA binding activity, an absolute necessity for the types of assays we used. But it is interesting to note that a *ftz* HD mutant, in which amino acids C-terminal to position 53 were deleted, did not bind to DNA (Percival-Smith and Gehring, unpublished results). This protein could not be purified by our standard procedure, so we do not know if this peptide is folded properly. However, interestingly Met54 of helix IV makes an intermolecular NOE with C7 of the α strand indicating an important role for helix IV in DNA binding (Otting *et al.*, 1990).

One surprising finding with the change of specificity mutation was that if position 50 was the only determinant of the differential specificities between the *ftz* and *bicoid* proteins (Hanes and Brent, 1989; Treisman *et al.*, 1989), they would have overlapping binding specificites [both the *ftz* HD and the *ftz* (Q50K) HD recognize the BS2v1 and BS2v2 binding sites with similar affinities]. However, it is possible that position 54, which is a Met in *ftz* and an Arg in *bicoid*, may also contribute to differential binding, such that the *bicoid* protein only recognizes with high affinity the BS2v3 binding site. As noted above in the Discussion, Met54 makes an intermolecular NOE with DNA. It is also interesting to note that a mutational analysis of BS2 making single base pair exchanges would not have identified positions 6 and 7 as important for *ftz* HD binding, since it was necessary to change both base pairs to observe a dramatic reduction in affinity.

Materials and methods

Construction of ftz HD expression plasmid Bluefop 1

ftz HD peptides were expressed in a T7 expression vector (Studier and Moffat, 1986) called Blue75+. It combines the advantages of the Bluescript

plasmids (Stratagene) with the T7 expression vector pAR3040 (Rosenberg *et al.*, 1987). Blue 75+ is a Bluescript M13+ derivative, which has a ~80 bp *Xba*I–*Bam*HI fragment of pAR3040 inserted into its polylinker downstream of the T7 promoter. The ~80 bp fragment contains a Shine–Dalgarno sequence (Shine and Dalgarno, 1974) and a start codon necessary for the translation of the mRNA synthesized from the T7 promoter in the *Escherichia coli* strain BL21 (DE3) (Studier and Moffat, 1986). In order to clone the *ftz* homeobox into Blue 75+, a 270 bp *Nae*I fragment containing the *ftz* homeobox was isolated from a *ftz* cDNA clone (Kuroiwa *et al.*, 1984). Linkers providing an in-frame stop codon and a *Bam*HI site were ligated to the *Nae*I fragment. The linkered *Nae*I fragment was digested with *Hin*fI, which cuts immediately 5′ to the homeobox, and the sticky ends were filled-in with Klenow polymerase. The fragment was finally digested with *Bam*HI. The expression vector Blue 75+ was digested with *Nhe*I and the 5′ protruding ends were digested with mung bean nuclease. The linearized vector was cut with *Bam*HI and subsequently ligated with the *ftz* homeobox fragment. The new plasmid was called Bluefop 1 and its sequence was confirmed by sequencing. It encodes a peptide 73 amino acids in length in which only the N-terminal Met is not derived from the *ftz* gene.

Site directed mutagenesis

Site directed mutagenesis was used to introduce specific changes in the *ftz* HD (Amersham oligonucleotide directed *in vitro* mutagenesis system version 2). ssDNA of Bluefop 1 was prepared, and the mutagenesis deoxyoligonucleotides were purified over an Oligonucleotide Purification Cartridge (Applied Biosystems). The sequence of the oligodeoxynucleotides used to mutate Bluefop 1 are:

ftz (R43A) HD	C T T G A T C T G C G C T T C G C T C A G G
ftz (NTD) HD	G A T A T A C A T G/ A C G T A C A C C C
ftz (Q50K) HD	G A T C T G G T T C A A A A C C G A C G

The underlined bases indicate mismatches; the line indicates the fusion junction. The correct sequence of the mutagenized vectors was confirmed by sequencing.

Expression and purification of the peptides

All peptides were expressed in *E.coli* BL21 (DE3) lysogen (Studier and Moffatt, 1986). The *ftz* HD and *ftz* (R43A) HD were purified to near homogeneity as described in Müller *et al.* (1988) for the *Antp* HD. The *ftz* (NTD) HD and the *ftz* (Q50K) HD were partially purified essentially as described in Müller *et al.* (1988) omitting FPLC chromatography.

Oligodeoxynucleotides

16 bases long single-stranded deoxyoligonucleotides were purified over Oligonucleotide Purification Cartridges (Applied Biosystems). The DNA concentration of the single-stranded oligodeoxynucleotides was determined by direct absorption measurements at 260 nm. Equal molar amounts of the appropriate oligodeoxynucleotides were annealed. For the saturation experiments the annealed double-stranded DNA was labelled with [α-^{32}P]dATP using Klenow DNA polymerase in the presence of the dGTP, dCTP and dTTP, and subsequently chased with dATP. The sequence of the α strand of the filled in, double-stranded oligodeoxynucleotides are (5′–3′):

BS2-18	G A G A A A A A G C C A T T A G A G
BS2v1-18	G A G A A A A A G C G A T T A G A G
BS2v2-18	G A G A A A A A G G C A T T A G A G
BS2v3-18	G A G A A A A A G G G A T T A G A G

Mobility shift assays

Binding reactions for the saturation experiments were carried out in 20 μl of 20 mM Tris–HCl pH 7.6/75 mM KCl/50 μg of bovine serum albumin per ml/1 mM dithiothreitol/10% glycerol and incubated at 20 ± 1°C for 2–3 h. The samples were run at room temperature through 15% polyacrylamide gels. Gel slices corresponding to free and bound DNA were cut out and the amount of radioactivity quantified by liquid scintillation spectroscopy.

The binding reaction for the interference studies was the same as above, except it contained 1 ng/μl of Bluescript DNA as competitor and samples were incubated on ice for ~1 h. The samples were run through 15% polyacrylamide gels at 5°C.

Interference studies

For the construction and structure of the six plasmids used see Affolter *et al.* (1990). *Hin*dIII–*Xba*I DNA fragments, 3′ end labelled at the *Hin*dIII site were isolated from pBS1-26a, pBS1-26b, pBS2-26a, pBS2-26b, pBSX-18a and pBSX-18b, treated with ethyl nitrosourea, and subsequently cleaved at the modified phosphates as described previously (Siebenlist and Gilbert, 1980). Dimethyl sulphate treatment and subsequent cleavage with 0.1 M NaOH were done as published (Maxam and Gilbert, 1977).

Acknowledgements

We thank Reinhard Dölz for help with recreational computer modelling. We thank Kurt Wüthrich for his comments on the manuscript. We especially thank Erika Marquardt for careful typing of the manuscript. A. P.-S. was supported by a Medical Research Council (Canada) Fellowship. The work was supported by the Kantons of Basel and the Swiss National Science Foundation for which we are grateful.

References

Affolter,M., Percival-Smith,A., Müller,M., Leupin,W. and Gehring,W.J. (1990a) *Proc. Natl. Acad. Sci. USA*, **87**, 4093–4097.
Affolter,M., Schier,A. and Gehring,W.J. (1990b) *Curr. Opinion Cell Biol.*, **2**, 485–495.
Bass,S., Sorrells,V. and Youderian,P. (1988) *Science* **242**, 240–245.
Berleth,T., Burri,M., Thoma,G., Bopp,D., Richstein,S., Frigerio,G., Noll,M. and Nüsslein-Volhard,C. (1988) *EMBO J.*, **7**, 1749–1756.
Billeter,M., Qian,Y.Q., Otting,G., Müller,M., Gehring,W.J. and Wüthrich,K. (1990) *J. Mol. Biol.*, **214**, 183–197.
Bürglin,T.R., Finney,M., Coulson,A. and Ruvkun,G. (1989) *Nature*, **341**, 239–243.
Desplan,C., Theis,S. and O'Farrell,P.H. (1988) *Cell*, **54**, 1081–1090.
Driever,W. and Nüsslein-Volhard,C. (1989) *Nature*, **337**, 138–143.
Gehring,W.J. (1987) *Science*, **236**, 1245–1252.
Gehring,W.J., Müller,M., Affolter,M., Percival-Smith,A., Billeter,M., Qian,Y.Q., Otting,G. and Wüthrich,K. (1990) *Trends Genet.*, **6**, 323–329.
Güntert,P., Qian,Y.Q., Otting,G., Müller,M., Gehring,W.J. and Wüthrich,K. (1990) *J. Mol. Biol.*, in press.
Hanes,S.D. and Brent,R. (1989) *Cell*, **57**, 1275–1283.
Hiromi,Y., Kuroiwa,A. and Gehring,W.J. (1985) *Cell*, **43**, 603–613.
Hoey,T. and Levine,M. (1988) *Nature*, **332**, 858–861.
Jordan,S.R. and Pabo,C.O. (1988) *Science*, **242**, 893–899.
Kuroiwa,A., Hafen,E. and Gehring,W.J. (1984) *Cell*, **37**, 825–831.
Laughon,A. and Scott,M.P. (1984) *Nature*, **310**, 25–31.
Lehming,M., Sartorius,J., Kisters-Woike,B., von Wilcken-Bergman,B. and Müller-Hill,B. (1990) *EMBO J.*, **9**, 615–621.
Levine,M. and Hoey,T. (1988) *Cell*, **55**, 537–540.
Maxam,A. and Gilbert,W. (1977) *Methods Enzymol.*, **65**, 499–525.
McGinnis,W., Levine,M.S., Hafen,E., Kuroiwa,A. and Gehring,W.J. (1984a) *Nature*, **308**, 428–433.
McGinnis,W., Garber,R.L., Wirz,J., Kuroiwa,A. and Gehring,W.J. (1984b) *Cell*, **38**, 403–409.
Müller,M. (1990) Ph.D. Thesis, University of Basel.
Müller,M., Affolter,M., Leupin,W., Otting,G., Wüthrich,K. and Gehring,W.J. (1988) *EMBO J.*, **7**, 4299–4304.
Otting,G., Qian,Y.Q., Müller,M., Affolter,M., Gehring,W.J. and Wüthrich,K. (1988) *EMBO J.*, **7**, 4305–4309.
Otting,G., Qian,Y.Q., Billeter,M., Müller,M., Affolter,M., Gehring,W.J. and Wüthrich,K. (1990) *EMBO J.*, **9**, 3085–3092.
Pabo,C.O., Krovatin,W., Jeffrey,A. and Sauer,R.T. (1982) *Nature*, **298**, 441–443.
Pabo,C.O., Aggarwal,A.K., Jordan,S.R., Beamer,L.J., Obeysekare,U.R. and Harrison,S.C. (1990) *Science*, **247**, 1210–1213.
Pick,L., Schier,A., Affolter,M., Schmidt-Glenewinki,T. and Gehring,W.J. (1990) *Genes Dev.*, **4**, 1224–1239.
Qian,Y.Q., Billeter,M., Otting,G., Müller,M., Gehring,W.J. and Wüthrich,K. (1989) *Cell*, **59**, 573–580.
Rosenberg,A.H., Lade,B.N., Chui,D.S., Lin,S.W., Dunn,J.J. and Studier,F.W. (1987) *Gene*, **56**, 125–135.
Scott,M.P. and Weiner,A.J. (1984) *Proc. Natl. Acad. Sci. USA*, **81**, 4115–4119.
Scott,M.P., Tamkun,S.W. and Hartzell,G.W.,III (1989) *Biochim. Biophys. Acta Rev. Cancer*, **989**, 25–49.
Shepherd,J.C.W., McGinnis,W., Carrasco,A.E., De Robertis,E.M. and Gehring,W.J. (1984) *Nature*, **310**, 70–71.
Shine,J. and Dalgarno,L. (1974) *Proc. Natl. Acad. Sci. USA*, **71**, 1342–1346.
Siebenlist,U. and Gilbert,W. (1980) *Proc. Natl. Acad. Sci. USA*, **77**, 122–126.

Studier,F.W. and Moffat,B.A. (1986) *J. Mol. Biol.*, **189,** 113–130.
Treisman,J., Gönczy,P., Vashishtha,M., Harris,E. and Desplan, C. (1989) *Cell,* **59,** 553–562.
Wharton,R.P. and Ptashne,M. (1987) *Nature,* **326,** 888–891.

Received on July 23, 1990; revised on August 27, 1990