MOLECULAR AND CELLULAR BIOLOGY, July 1991, p. 3613–3623
0270-7306/91/073613-11$02.00/0

Vol. 11, No. 7

DNA-Binding Specificity of the fushi tarazu Homeodomain†

BRIAN FLORENCE,[1,2] RICHARD HANDROW,[1] AND ALLEN LAUGHON[1]*

Departments of Genetics and Medical Genetics[1] and Program in Cell and Molecular Biology,[2] University of Wisconsin, Madison, Wisconsin 53706

Received 26 February 1991/Accepted 17 April 1991

The fushi tarazu (*ftz*) gene of *Drosophila melanogaster* encodes a homeodomain-containing transcription factor that functions in the formation of body segments. Here we report an analysis of the DNA-binding properties of the *ftz* homeodomain in vitro. We provide evidence that the homeodomain binds to DNA as a monomer, with an equilibrium dissociation constant of 2.5×10^{-11} M for binding to a consensus binding site. A single *ftz* binding site occupies 10 to 12 bp, as judged by the ability of protein bound at one site to interfere with binding to an adjacent site. These experiments also demonstrated a lack of cooperative binding between *ftz* homeodomains. Analysis of single-nucleotide substitutions over an 11-bp sequence shows that a stretch of 6 bp is critical for binding, with an optimal sequence of 5′$^{C}_{T}$AATTA3′. These data correlate well with recent structural evidence for base-specific contact at these positions. In addition, we found that sequences flanking the region of direct contact have effects on DNA binding that could be of biological significance.

The homeodomain is a conserved DNA-binding domain that occurs in a large number of eukaryotic transcription factors (33, 45, 46). The homeodomain contains a helix-turn-helix structure analogous to that of many prokaryotic transcriptional repressors and activators (28, 39, 42). A number of homeodomain proteins from *Drosophila melanogaster* have been extensively characterized with respect to their DNA-binding properties and ability to activate or repress transcription (15). For a number of *Drosophila* homeodomain proteins, the identities of regulatory target genes have been inferred from genetic data (6, 10, 14, 18, 19, 49). Footprint analysis of sequences near the promoters of putative target genes has led to the identification of consensus binding-site sequences for several homeodomain proteins. In some cases the binding-site sequences that have been identified in vitro have been shown to function in the regulation of transcription in vivo (8, 11, 25, 44, 52).

Consensus binding-site sequences have been identified for the following *Drosophila* homeodomains or homeodomain proteins: even-skipped (*eve*) and engrailed (*en*) (TCAATTA AAT [7, 16]), Antennapedia (*Antp*) (ANNNNCATTA [1, 34]), Ultrabithorax (*Ubx*) (TAA$_n$ [3]), bicoid (*bcd*) (GGGAT TAGA [8]), and fushi tarazu (*ftz*) (CAATTA [35] or GCAAT TAA [41]). These consensus sequences share a 5′ATTA3′ "core" that appears to be critical for homeodomain binding (36). The importance of the ATTA core is evident in the recently determined three-dimensional structures of the *Antp* and *en* homeodomain-DNA complexes (24, 38). Bases within the first half of the core are contacted in the major groove by helix 3, while the second half is contacted in the minor groove by the N-terminal arm. The amino acids that contact the core sequence (Arg-3, Arg-5, Ile-47, and Asn-51) are completely conserved among the homeodomains that have been shown to recognize ATTA (15, 45). In addition to the ATTA core, 2 bp to the 5′ side of the ATTA affect binding and are positioned near the Gln-50 sidechain of helix three (24, 38). The sequences CAATTA and CCATTA are efficiently recognized by homeodomains with glutamine at position 50, whereas the *bcd* homeodomain, with a lysine at position 50, has a preference for the sequence GGATTA (12, 13, 40, 50). Therefore, the recognition of these two bases can be attributed to contact with residue 50.

To learn more about how DNA sequence determines the efficiency of homeodomain binding, we have measured the affinity of the *ftz* homeodomain for a series of binding-site variants. The results serve to pinpoint base pairs that should be critical for discrimination of *cis*-regulatory sequences by the *ftz* homeodomain and its relatives. Our findings fit well with the methylation interference and mutant binding-site data reported for the *ftz* homeodomain by Percival-Smith et al. (40) and with the three-dimensional structures that have been determined for the *Antp* and *en* homeodomain-DNA complexes by Otting et al. (38) and Kissinger et al. (24). In addition, we present a characterization of other aspects of DNA recognition by the *ftz* homeodomain. These include stoichiometry of binding, kinetic and equilibrium analyses, and protein-protein interactions of bound homeodomains. From these data, we found that the *ftz* homeodomain binds to DNA with high affinity as a monomer and that adjacent homeodomains do not bind DNA cooperatively.

MATERIALS AND METHODS

Expression and purification of the *ftz* homeodomain. Two versions of the *ftz* homeodomain were expressed in *Escherichia coli* by using the T7 expression plasmid pET3a (48). All cloning steps were performed according to Sambrook et al. (43). The first version, HD1, was created by insertion of a 267-bp *Nae*I fragment from the *ftz* cDNA clone G20 (28) into the *Bam*HI site of pET3a with 12-bp *Bam*HI linkers (New England BioLabs). The resulting protein has a predicted molecular mass of 12,737 Da, with its homeodomain flanked by 18 N-terminal and 11 C-terminal *ftz* residues that are preceded and followed by 15 and 6 phage or linker-derived amino acids, respectively. The second construct, HD2, encodes the homeodomain preceded by the sequence Met-Val-Ser. Polymerase chain reaction (PCR) amplification was used to position an *Nhe*I site and TAG stop codon at the N- and C-terminal ends of the homeodomain, respectively, before cloning into pUC18. The subsequent *Nhe*I-*Bam*HI fragment was cloned into pET3a.

* Corresponding author.

† University of Wisconsin, Madison Department of Genetics publication number 3198.

One liter of BL21(DE3)lysE cells containing plasmids encoding either HD1 or HD2 was grown at 37°C to an A_{600} of 0.6 before being induced to express T7 RNA polymerase with 0.2 mM IPTG (isopropylthiogalactopyranoside), followed by an additional 2.5 h of incubation. The cultures were then centrifuged, and the cell pellet was suspended in 20 ml of TE (10 mM Tris [pH 7.5], 1 mM EDTA [pH 8.0]). The cells were frozen at −80°C and then lysed by rapid thawing while being swirled in a 37°C bath. Then 2 ml of 5 M NaCl was added to the lysate, and the insoluble material was collected after 10 min of centrifugation at 8,000 rpm in a Sorvall SA600 rotor at 4°C. The pellet, containing mostly HD1 or HD2 protein in the form of inclusion bodies (31), was washed and disrupted in a Dounce homogenizer twice in TE–500 mM NaCl and then three times in TE. The insoluble pellet was denatured in 8 ml of 4 M guanidine-HCl (GuHCl)–5 × TE. At this stage, HD1 preparations were fractionated by gel filtration, while HD2 preparations were treated as follows. Isopropanol (8 ml) was added dropwise, and the resulting nucleic acid precipitant was removed by centrifugation. Four volumes of methanol were added to the supernatant, and the precipitant was isolated by centrifugation. The resulting pellet was washed with methanol and then dissolved in 4 M GuHCl–50 mM Tris (pH 7.5). For both HD1 and HD2, insoluble material was cleared by centrifugation before 4 ml was applied to a column (50 by 2.5 cm) of Sephacryl S200 (Pharmacia) equilibrated with 4 M GuHCl–50 mM Tris (pH 7.5). The column was eluted with the same buffer, and fractions containing the homeodomain peak were pooled and diluted with 1 volume of TE. The protein was then dialyzed against 200 ml of 2 M GuHCl–20 mM Tris (pH 7.5)–1 mM EDTA–1 mM dithiothreitol (DTT)–20% glycerol for 1 h. The GuHCl concentration was reduced slowly by dropwise addition to the dialysis of 1.8 liters of 20 mM Tris (pH 7.5)–100 mM NaCl–1 mM EDTA–1 mM DTT–20% glycerol (buffer R) over 8 h. Samples were then dialyzed against 500 ml of buffer R overnight at 4°C. The resulting preparations were judged to be highly pure, as no other proteins were visible on sodium dodecyl sulfate (SDS)-polyacrylamide gels after staining with Coomassie brilliant blue and no protein was evident in the equivalent fractions from control lysates. Protein concentrations were calculated by using coefficients of 9.3×10^{-5} M/A_{280} for HD1 and 6.5×10^{-5} M/A_{280} for HD2. Bio-Rad protein assays yielded similar results. The protein preparations were stored at −20°C in buffer R that had been adjusted to 50% glycerol.

Construction of synthetic binding sites. The head-to-head (HH), tail-to-tail (TT), and head-to-tail (HT) binding sequences shown in Fig. 2C were synthesized as complementary oligomers, annealed, and cloned into the *Sma*I site of pEMBL19$^+$. T0 was constructed by removing sequences between a *Bcl*I site in the middle of the cloned TT2 site and a flanking *Bam*HI site. H0 was constructed similarly by using a *Bgl*II site within HH6. All clones were sequenced to confirm the insertion of single oligomers.

A pool of 32-bp DNA fragments, degenerate at six positions adjacent to ATTA, were generated by extension of the oligomer primer A (5′TACCCGGGGATCCTTAAT3′) annealed to the oligomer ATTA-32 (5′ACTCTAGANNNNNN ATTAAGGATCCCCGGGTA3′). An electrophoretic mobility shift assay (EMSA; see below) with HD1 was used to purify fragments containing high-affinity binding sites from the degenerate-sequence pool. Purified fragments were cleaved with *Bam*HI and *Xba*I and cloned between these two sites in the plasmid pUC18. The consensus binding-site sequence was generated by extension of primer-A annealed to the oligomer CON (5′ACTCTAGAAAAGCAATTAAGG ATCCCCGGG3′), followed by cleavage and cloning between the *Bam*HI and *Xba*I sites of pUC18. Single-base-pair substitutions within the sequence AAGCA of the consensus sequence were generated by extension of primer-A that had been annealed to five different oligomers identical to CON except that each was triply degenerate at one of the desired positions. Substitutions of the consensus within the sequence ATTAAG were created similarly, by extension of primer-P (5′GGGTCTAGAAAAGCA3′) that had been annealed to six different oligomers, each of which was triply degenerate for one of the underlined nucleotides (5′GGGG ATCCTTAATTGCTTTTCTAGACCC3′).

EMSA. The assay conditions for EMSA experiments evolved during the course of our investigation for several reasons. Concentrations of bovine serum albumin (BSA) and glycerol were adjusted to improve binding affinity. Although it is not clear why higher BSA concentrations raise binding affinity, the increase in affinity caused by higher glycerol concentrations could be due to solvent exclusion (34a). Competing plasmid vector DNA that was present in the initial binding assay mixes (see Table 2) was removed in later experiments so that absolute values for binding constants could be obtained. The temperature was changed from 4°C in early experiments (see Table 2 and Fig. 2) to 20°C in accordance with the normal environment of the protein.

Synthetic binding-site fragments for use in EMSAs were generated either by *Eco*RI and *Hin*dIII cleavage of cloned DNAs (see Fig. 3 and 4 and Tables 1 and 2) or by PCR amplification (20 cycles) of cloned sites with primers B and C (5′GCTGCCTGCAGGTCGAC3′ and 5′CGAGCTCGGTAC CCGGGG3′, respectively; data in Table 3). The experiments summarized in Table 2 were performed on binding-site fragments excised from the parent plasmid with *Eco*RI and *Hin*dIII. The fragment was not separated from the vector, and thus the binding reactions were done in the presence of competitor plasmid DNA. In all other experiments, binding-site fragments were fractionated away from plasmid vector DNA by electrophoresis in 2% agarose gels and recovered from gel slices by electroelution. The isolated fragments were precipitated and quantitated by UV absorption. DNA was labeled by using Klenow fragment and [α-^{32}P]dATP to fill in the sticky ends or by using T4 polynucleotide kinase and [γ-^{32}P]ATP to phosphorylate dephosphorylated 5′ ends. The reaction mixes were extracted once with phenol and then passed over G50-80 spin columns equilibrated in TE. For the experiments described below, unless otherwise specified, 20 μl of protein was added to 1 μl of labeled DNA, so that the concentration of protein was at least 10 times greater than that of DNA.

Measurements of binding affinities for the selected clones in Table 2 were obtained as follows. HD1 was diluted to 4×10^{-8} M in 20 mM Tris (pH 7.5)–100 mM NaCl–5 mM $MgCl_2$–1 mM EDTA–1 mM DTT–10% glycerol and incubated on ice for 1 h. Protein was added to the labeled DNA, and the mixture was incubated on ice for 1 h. Samples were then electrophoresed in 10% polyacrylamide gels (0.15 by 16 by 18 cm) containing 0.5× TBE (43) for 3 h at 190 V and 5°C. Gels were fixed, dried, and used to expose X-ray film. DNA bands were cut out of the dried gels, and radioactivity was quantitated with a scintillation counter.

Measurements of binding affinities for tandemly arranged binding sites (see Fig. 3 and 4 and Table 1) were obtained as follows. HD2 protein was diluted into cold 20 mM *N*-tris[hydroxymethyl]methyl-2-aminoethanesulfonic acid (TES, pH 7.5)–100 mM KCl–1 mM EDTA–1 mg of BSA (dialyzed

Pentax V) per ml–5% glycerol and kept on ice for 1 h. Protein was brought to 20°C before the addition of DNA. Samples were incubated for 2.5 h at 20°C and electrophoresed at room temperature. Gels were fixed and dried, and the DNA in each band was quantitated with a Betascope. The data in Table 3 were obtained in the same way as those in Table 1 except that HD2 was diluted to 4×10^{-10} M in cold 20 mM TES (pH 7.5)–75 mM KCl–50 μg of BSA (dialyzed Pentax V) per ml–10% glycerol and incubated for 4 h at 20°C prior to electrophoresis.

The data used to calculate association constants (see Fig. 5) were obtained as follows. HD2 protein was diluted to 10^{-9} M under the conditions specified above. Protein (140 μl) was added to 7 μl of labeled CON or HT6. At 30-s intervals, for a total of 3 min, 20-μl aliquots were removed and 1 μl of buffer containing a 200-fold excess of unlabeled CON or HT6 was added. DNA in bound and unbound samples was quantitated as described above. The data used to calculate dissociation constants (see Fig. 6) were obtained as follows. HD2 was diluted as above and brought to equilibrium with either labeled CON or HT6. Every 30 min for 2 h, a 2-μl aliquot was diluted 10-fold into buffer with a 100-fold excess of unlabeled CON or HT6. DNA in the samples was quantitated as above.

Calculation of binding constants. All equilibrium and kinetic constants were calculated from three replications unless otherwise noted. Equilibrium dissociation constants were defined by the formula K_d = [DNA][HD]/[HD-DNA], where [DNA] is the concentration of free DNA, [HD] is the concentration of free protein, and [HD-DNA] is the concentration of the protein-DNA complex. Protein was kept at least 10-fold in excess over DNA, allowing the approximation [HD] ≈ [HD_{total}] to be used in calculating equilibrium binding constants. The value of [DNA]/[HD-DNA] was determined by taking the ratio of free counts per minute (cpm) to bound cpm. For the binding to DNA with two binding sites, K_1 describes the K_d for the singly-bound form and K_2 ([HD-DNA][HD]/2[HD_2-DNA]) describes the K_d for the doubly-bound form, where HD_2-DNA is the complex containing two bound homeodomains. The factor of 2 is present to account for the ability of either of the two homeodomains to dissociate (5). The mean equilibrium binding constants shown in Table 3 were compared by analysis of variance with the least significant difference to determine which values differed from that of the consensus sequence at the $\alpha = 0.05$ significance level (47).

Association kinetics for monomer species were determined by the approximation $-k_a[HD_T] = \ln([DNA_t]/[DNA_0])$, where $[DNA_t]/[DNA_0]$ is the ratio of free DNA at time t to free DNA at time 0. Dissociation kinetics were determined by $-k_d t = \ln([\text{HD-DNA}_t]/[\text{HD-DNA}_0])$.

RESULTS

The *ftz* homeodomain binds DNA with high affinity as a monomer. Two versions of the *ftz* homeodomain were used in experiments to investigate homeodomain-DNA interactions in vitro. Both were synthesized in *E. coli* with the T7 RNA polymerase system of Studier et al. (48). The first protein, HD1, contains the homeodomain flanked by 18 N-terminal and 11 C-terminal *ftz* residues and additional vector-derived amino acids at the N and C termini. HD2 contains the homeodomain plus three vector-derived N-terminal amino acids. Both proteins were purified from insoluble fractions of *E. coli* lysates. The predicted molecular masses of HD1 and HD2 are 12,737 and 8,138 Da, respectively. Gel filtration chromatography was performed to determine whether HD1 exists as a monomer or a dimer in solution. Figure 1 shows that under nonreducing conditions, HD1 eluted with an apparent molecular mass of 10,000 Da, indicating that the *ftz* homeodomain exists as a monomer in solution.

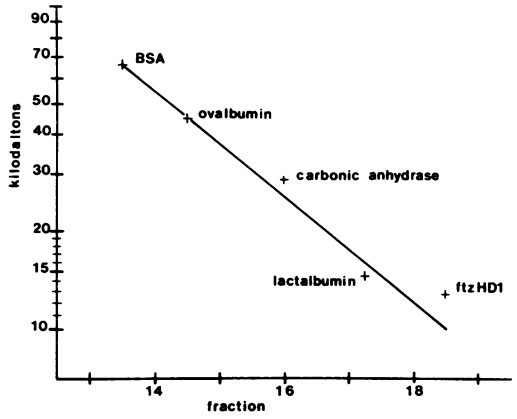

FIG. 1. Estimation of *ftz* homeodomain size in solution by gel filtration. The centers of the elution peaks for each protein are plotted against their predicted molecular masses. HD1 and the indicated proteins (Sigma) (20 μg of each) were applied to a Sephacryl S200 column (1 by 20 cm) in 100 μl containing 20 mM Tris (pH 7.5), 0.5 M NaCl, 5 mM $MgCl_2$, 1 mM DTT, and 5% glycerol. Fractions were collected, and 5-μl samples from each were fractionated on a 17.5% polyacrylamide–SDS gel. Protein bands were visible after staining with Coomassie brilliant blue.

The stoichiometry of *ftz* homeodomain-DNA complexes was determined in two ways. First, a limiting amount of HD1 protein was incubated with an excess of a synthetic DNA fragment containing a single natural *ftz* binding site (from near the *Antp* P2 promoter [35]), and the quantities of DNA in the resulting HD1-DNA complexes were measured by an EMSA. At saturating concentrations of DNA fragment, the molar ratio of HD1 to DNA in the complex was very close to 1 (Fig. 2), suggesting that the *ftz* homeodomain bound to DNA as a monomer. This experiment also showed that virtually all of the HD1 protein was able to bind to DNA.

Second, because of differences in molecular mass, HD1-DNA complexes migrate more slowly than HD2-DNA complexes in an EMSA. As an independent means of determining whether the *ftz* homeodomain binds to DNA as a monomer, complexes formed in the presence of both HD1 and HD2 were assayed by EMSA to determine whether a heterodimer with intermediate mobility was formed (17). Heterodimers were not observed (Fig. 3), again supporting the conclusion that the *ftz* homeodomain bound to DNA as a monomer.

The *ftz* homeodomain does not bind DNA cooperatively. Naturally occurring homeodomain-binding sites have often been found in clusters (16, 35), and the binding-site consensus for *en* and *eve*, TCAATTAAAT, exhibits partial symmetry (7). Both of these findings suggested the possibility that homeodomain proteins interact cooperatively in binding to DNA sequences. To determine whether the *ftz* homeodomain bound to DNA cooperatively, we measured the affinity of HD2 for two adjacent sites aligned either in the same direction or in the two possible opposing directions. In each orientation, sites were positioned at several distances. Thus, it was possible to detect any increase in affinity for a second

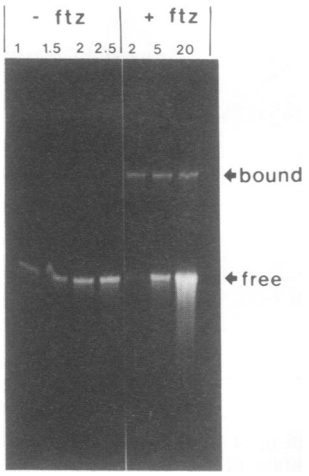

FIG. 2. Monomeric binding of the *ftz* homeodomain to DNA. HD1 was titrated with increasing amounts of a 34-bp DNA fragment containing a single high-affinity binding site. Either 1 μg of HD1 (+*ftz*) or no protein (−*ftz*) was mixed with the amount (in micrograms) of the 34-bp fragment indicated above each lane and electrophoresed in a 15% polyacrylamide gel containing 1× TBE at 4°C. After being stained with ethidium bromide, DNA was visualized on a UV transilluminator. Bound and free DNA are indicated. When DNA was in excess, approximately 2 μg of the fragment was bound by the homeodomain at a molar ratio of 1:1. The sequence of the 34-bp fragment was 5′CGTATATAATATATAAGCAATTAAGG TAAACAGT3′ (binding site underlined).

site due to cooperative interactions between adjacent bound homeodomains, as well as to determine how closely binding sites could be spaced before steric hindrance would prevent the binding of a second protein. A series of oligonucleotides containing two binding sites in three orientations were used: HH, TT, and HT. Head and tail refer, respectively, to the 3′ and 5′ ends of the previously reported *ftz* consensus sequence, 5′CAATTA3′. Spacing of the HT sites was varied in 1-bp increments, and that of HH and TT sites was in 2-bp increments. EMSAs were used to determine the equilibrium dissociation constants for each sequence at four protein dilutions over a 10-fold range of HD2 concentrations.

A single copy of the sequence used in the HH series (H0)

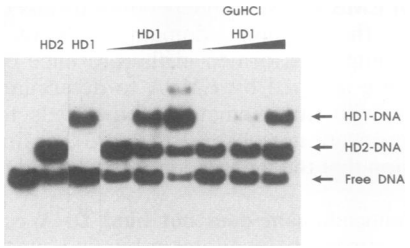

FIG. 3. HD1 and HD2 do not form heterodimers that bind to DNA. Shown is an autoradiograph of an EMSA in which HD1, HD2, or both proteins were allowed to bind to end-labeled TT2 DNA (sequence in Fig. 4). Lanes 1 to 3, free DNA, DNA plus HD2, and DNA plus HD1, respectively. Lanes 4 to 6 are successive threefold increases of HD1 concentration in mixtures containing fixed amounts of HD2 and DNA. Lanes 7 to 9 are the same as lanes 4 to 6 except that the proteins were denatured together in GuHCl and allowed to renature before DNA was added. Some binding of two homeodomains is visible as an additional band at the highest concentration of HD1.

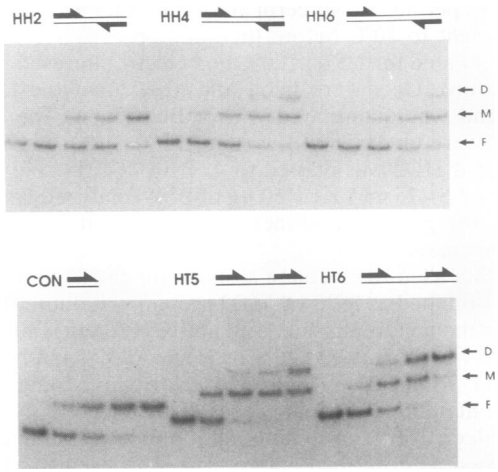

FIG. 4. Distance of closest approach of two *ftz* homeodomains on single DNA fragments. Shown are autoradiographs of EMSAs in which 0, 50, 100, 200, and 400 pM HD2 (five lanes, respectively, from left to right for each DNA) was allowed to bind to DNA fragments containing two binding sites. The top and bottom panels show results for HH and HT DNAs, respectively, with arrows above the lanes indicating the orientations of the binding sites and the arrows at the sides indicating the mobilities of DNAs that are free (F) or bound by one (M) or two (D) homeodomains.

had a K_d of 2.2×10^{-10} M. All of the HH constructs had an affinity similar to that of H0 for the binding of one HD2 monomer. This result would be expected if the rate of association per DNA fragment was not increased by the additional binding site. This interpretation is supported by the association kinetics described below. HH constructs that contained CAATTAs spaced less that 4 bp apart were unable to bind two homeodomains simultaneously (HH-2, HH0 [data not shown], and HH2, Fig. 4 and Table 1). Both HH4 and HH6 were able to bind a second homeodomain with the same efficiency. The increase in spacing from 4 to 6 bp did not decrease inhibition, i.e., a second homeodomain bound with 60% of the affinity of the first ($\omega \approx 0.6$; also true for HT10 and a double insert of TT4 [data not shown]). For this reason, $\omega \approx 0.6$ appears to correspond to a lack of inhibition. The center-to-center distance between sites in HH4 is 10 bp, or about one turn of the DNA helix.

The single-site sequence used in the TT series (T0) had a K_d of 1.6×10^{-10} M (Table 1). Fragments with two CAAT TA sequences spaced less than 2 bp apart were unable to bind two homeodomains simultaneously (TT-2 [data not shown] and TT0). Sites spaced 2 bp apart could be occupied simultaneously with significant inhibition, while 4-bp spacing (TT4) allowed binding to both sites without interference. Again, the center-to-center distance necessary for lack of inhibition was about one helical turn (10 bp).

In contrast to the 10-bp center-to-center distance of the two opposing orientations, DNA fragments with two tandem sites spaced 10 bp apart, center-to-center (HT4), could not accommodate two homeodomains at the same time (data not shown). Binding of a second homeodomain was still inhibited for HT5 (Fig. 4 and Table 1). Only when sites were spaced 12 bp apart, center-to-center (HT6), could both sites be bound simultaneously. Thus, tandem sites occupied 12 bp, while symmetric sites occupied about 10 bp. The single-site sequence (CON) used for the HT constructs was bound with significantly higher affinity (6×10^{-11} M) than that used

TABLE 1. Effect of orientation and spacing on the ability of two homeodomains to simultaneously bind a single sequence

Name	Sequence	Binding constants[a] (10^{-10} M)		ω^b
		K_1	K_2	
T0	CCTTAATTGATCC	1.6 ± 0.2		
TT0	CCTTAATTGCAATTAAGG	1.6 ± 0.1		
TT2	CCTTAATTGATCAATTAAGG	1.9 ± 0.1	6.6 ± 0.7	0.28 ± 0.002
TT4	CCTTAATTGATATCAATTAAGG	1.5 ± 0.1	2.1 ± 0.3	0.66 ± 0.06
H0	CCGCAATTAAGATCC	2.2 ± 0.2		
HH2	CCGCAATTAATTAATTGCGG	2.4 ± 0.2		
HH4	CCGCAATTAACGTTAATTGCGG	2.1 ± 0.2	3.1 ± 0.3	0.60 ± 0.06
HH6	CCGCAATTAAGATCTTAATTGCGG	2.7 ± 0.2	4.3 ± 0.5	0.56 ± 0.06
CON	AAGCAATTAAGG	0.63 ± 0.06		
HT5	AAGCAATTAAGAAGCAATTAAGG	0.53 ± 0.05	2.0 ± 0.1	0.21 ± 0.03
HT6	AAGCAATTAAGGAAGCAATTAAGG	0.58 ± 0.05	0.78 ± 0.06	0.67 ± 0.03

[a] K_1, mean equilibrium dissociation constant ± standard error of the mean for the binding of the first homeodomain; K_2, mean equilibrium dissociation constant ± standard error for the mean for the binding of a second homeodomain.
[b] Mean interference factor ($\omega = K_1/K_2$ [32]) ± standard error of the mean. A ω of <1 for binding to one site inhibits binding to a second site, and a ω of >1 for binding to one site enhances binding to a second site. The mean ω was determined from the ω of each experiment.

for both TT and HH. The decrease in affinity for T0 is mostly attributable to the thymine 5′ to CAATTA, while the decrease in affinity for H0 is due to the CC dinucleotide 5′ to GCAATTA (see below and Tables 1 and 2).

Regardless of orientation or spacing, binding of a second site was always inhibited by binding of the first site ($\omega < 1$). Thus, close contact between any of the surfaces on adjacently bound *ftz* homeodomains did not result in cooperative binding.

Association and dissociation kinetics. The association and dissociation rates of HD2 with DNA were measured for both the CON and HT6 binding sites. Since the pattern of *ftz* protein expression evolves rapidly during embryogenesis, binding kinetics should be important for its function. These measurements were also important in establishing conditions for measuring equilibrium binding constants. The rate of association was measured by incubating HD2 with labeled DNA and removing aliquots over a short time course. Association in each sample was stopped by addition of a 200-fold excess of unlabeled DNA. The amounts of DNA bound were determined by EMSA. Dissociation of homeodomain-DNA complexes was negligible during the time course of these experiments. The kinetic association constants (k_a) of HD2 with CON and HT6 were nearly identical at 1.36 (±0.06) × 10^6 and 1.35 (±0.01) × 10^6 M^{-1} s^{-1} (Fig. 5). The fact that the association rates for fragments with one or two sites were identical explains why the equilibrium constants reported above were also the same for fragments with one or two sites. To measure dissociation rates, an excess of HD2 was brought to equilibrium with labeled DNA. Aliquots were taken at various times, and association was inhibited by 10-fold dilution into buffer containing a 100-fold excess of unlabeled DNA. The results are shown in Fig. 6, with kinetic dissociation constants for CON and HT6 of 1.45 (±0.06) × 10^{-4} and 2.24 (±0.04) × 10^{-4} s^{-1}, respectively. The predicted equilibrium constants (k_d/k_a) of 1.1 × 10^{-10} M (CON) and 1.7 × 10^{-10} M (HT6) were close to the observed value of 6 × 10^{-11} M when CON and HT6 were tested under these conditions.

TABLE 2. Average equilibrium dissociation constants[a]

Clone	Sequence[b] (positions 0–11)	K_d (10^{-8} M) ± SEM	N^c
28	CTAGTAattaag	0.55 ± 0.07	2
51	AAAGCTattaag	0.84 ± 0.10	11
17	AATCTAattaag	0.85 ± 0.07	3
47	AAAGCGattaag	0.89 ± 0.11	3
30	GCTGCAattaag	0.96 ± 0.16	3
19	GAGGTAattaag	1.1 ± 0.4	3
18	AGACCAattaag	1.4 ± 0.7	3
54	AGTACAattaag	1.4 ± 0.2	3
52	CCGACAattaag	1.7 ± 0.2	3
20	TAGATAattaag	3.1 ± 1.3	3
53	AAACCTattaag	3.6 ± 0.5	3
57	AGAGCAattaag	3.6 ± 0.2	3
46	TTTATAattaag	3.6 ± 0.3	3
38	GAGACAattaag	4.2 ± 1.5	3
42	TAAACGattaag	5.2 ± 1.4	3
39	GCGCCGattaag	6.5 ± 0.4	3
41	GGAGCGattaag	6.7 ± 1.8	3
61	AATTGGattaag	8.8 ± 2.3	3
37	GAGACGattaag	9.1 ± 1.6	3
55	CGGACGattaag	10.1 ± 0.1	2
CON	AAAGCAattaag	0.5	1

[a] Dissociation constants for HD1 binding to the highest-affinity clones obtained from a pool of fragments with six degenerate base pairs 5′ of the sequence ATTA were determined. Binding reactions were done in the presence of competitor DNA (see Materials and Methods).
[b] Degenerate positions are shown in capital letters.
[c] N, number of repetitions used to determine equilibrium dissociation constant.

Substitutions within a 6-bp core and flanking sequences affect binding. Previously, it was found that all sequences protected by the *ftz* homeodomain in footprinting experiments contained the sequence ATTA (35), a sequence common to many homeodomain consensus sequences (15). The *ftz* footprints also suggested a weaker sequence preference at flanking positions, with CAATTA emerging as a tentative consensus sequence. To determine the sequence specificity of the *ftz* homeodomain more precisely, we carried out a selection for high-affinity binding sites from a pool of synthetic DNAs that were random for 6 bp 5′ of ATTA (37). A 400-fold excess of synthetic DNA fragments was mixed with HD1, and bound DNA was separated from unbound DNA by EMSA. Bound fragments were cloned and subsequently characterized individually. Table 2 lists the affinities and

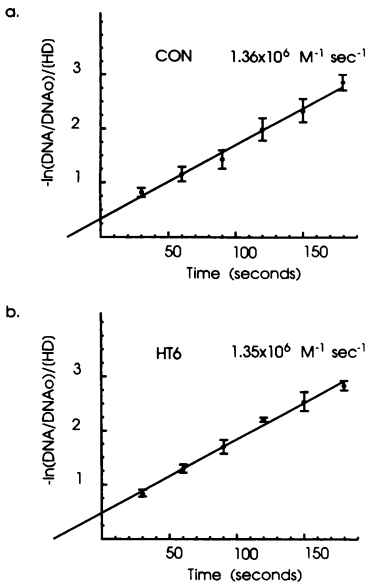

FIG. 5. Association kinetics of the *ftz* homeodomain. Data are for association of (a) HD2 and CON and (b) HD2 and HT6. Association kinetics were determined by plotting the natural logarithm of the ratio of free DNA, to free DNA_0 as a function of time (0 to 3 min), with k_a equaling the slope of the line. The protein concentration was 10^{-9} M, and the DNA concentration was 5×10^{-11} M. Error bars represent the standard error of the mean.

sequences of HD1 for the 20 highest-affinity clones recovered out of a total of 61. Binding affinities ranged 20-fold within the set. Nine of the clones were bound with dissociation constants of less than 2×10^{-8} M. The consensus sequence, AAAGCAATTA, was derived from the frequency of nucleotides at the six random positions in these nine clones and the ATTA core of the previous screen (35). This consensus was synthesized and found to bind with a dissociation constant equivalent to that of the highest-affinity sequences identified in the screen. Because the screen was not saturating (the consensus sequence was not recovered),

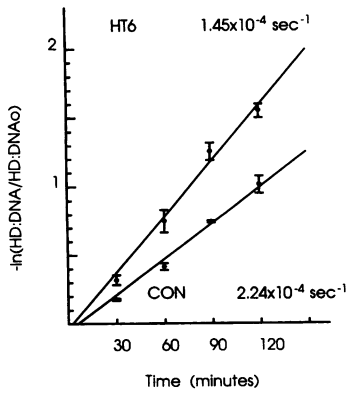

FIG. 6. Dissociation kinetics of the *ftz* homeodomain. Data are for dissociation of HD2-CON (lower line) and HD2-HT6 (upper line) complexes. Dissociation kinetics were determined by plotting the natural logarithm of the ratio of bound DNA, to bound DNA_0 as a function of time (0 to 3 h), with k_d equaling the slope of the line. Error bars represent the standard error of the mean. Data are averages from two trials.

it is a formal possibility that a higher-affinity site(s) exists. However, an analysis of single-nucleotide substitutions performed on this consensus (see below) failed to uncover higher-affinity sites, suggesting that this is the optimal site.

Within the larger set of 20 clones, all positions except position 4 (numbering is from 0 to 5 for AAAGCA for CON) tended to be a purine. Position 4 showed a very strong bias for cytosine or thymine, without any occurrence of adenine. Although 6 of the 10 highest-affinity sites differed from the previous consensus sequence at either position 4 or 5, only clone 61 differed at both positions, and the fairly high binding affinity for this sequence may have been due to the creation of a second site in the opposite orientation. The apparent redundancy of specificity between positions 4 and 5 suggests that the *ftz* homeodomain has some flexibility in its ability to utilize either of these two positions in base-pair recognition.

To determine the contributions of individual base pairs to *ftz* homeodomain recognition, the effects of mutations from a reference binding-site sequence were measured. The sequence AAGCAATTAAG was chosen as the reference sequence (the 2 bp adjacent to the ATTA on the 3′ side show a weak bias towards adenine and cytosine or guanine, respectively [35]). Synthetic binding sites degenerate for individual positions were cloned and screened for all possible mutations for each of the 11 bp. All but 2 of the 33 possible mutations were obtained and used to measure HD2 affinity. The results for repeated EMSA measurements of binding affinities for the consensus and mutant binding sites are listed in Table 3. The positions are numbered 1 through 11, with the ATTA corresponding to positions 6 through 9. The consensus was bound with a equilibrium dissociation constant of 2.5×10^{-11} M. The binding constant variation for the consensus binding site in Tables 1, 2, and 3 is due to the changes in assay conditions, as explained in Materials and Methods.

All but 2 of the 31 substitutions tested over the 11 bp reduced binding affinity significantly ($P < 0.05$). Neither of the two remaining substitutions (thymine 4 and cytosine 11) increased binding, indicating that this consensus is the sequence for which the *ftz* homeodomain has the highest affinity. The effects of point mutations fell roughly into two categories. Substantial reductions in binding affinity of between 6- and 90-fold at positions 4 and 6 through 9 served to identify the sequence CNATTA (or TAATTA, as indicated by data in Table 1) as critical for binding-site recognition. Mutations at flanking base pairs caused weak but statistically significant reductions in binding of less than fourfold. With the exception of position 5, the farther a mutation was from the center of the binding site, the less it affected binding.

The specificity requirements were more stringent at some positions within the binding site than at others. Any substitution for thymine at position 7 (ATTA) resulted in at least a 25-fold reduction in binding affinity, identifying this base as the single most important component of the binding site. At the other extreme, either cytosine or thymine was adequate at position 4, with guanine and adenine reducing binding 3.1- and 6.1-fold, respectively. This was consistent with the absence of both adenine and guanine at position 4 in 19 of the 20 selected synthetic binding sites. At position 3, substitution of thymine for guanine reduced binding affinity 3.6-fold relative to the consensus sequence, while substitution with cytosine and adenine reduced affinity only 1.4- and 2.1-fold, respectively. Again, this is consistent with the absence of thymine at position 3 in 19 of the selected synthetic binding sites (it is worth noting that the exceptions in both instances

TABLE 3. Average equilibrium dissociation constants for HD2 binding to single-base-pair variants from the *ftz* consensus binding site sequence

Position changed	New base	Sequence	K_d (10^{-11} M)[a] ± SEM	K_d/K_d for CON[b]
None (CON)		AAGCAATTAAG	2.5 ± 0.5	1
1	C	CAGCAATTAAG	4.5 ± 0.3	1.8
	G	GAGCAATTAAG	3.9 ± 0.3	1.5
	T	TAGCAATTAAG	3.3 ± 0.6	1.3
2	C	ACGCAATTAAG	7.2 ± 1.2	2.8
	G	AGGCAATTAAG	4.0 ± 0.3	1.6
	T	ATGCAATTAAG	6.5 ± 0.3	2.6
3	A	AAACAATTAAG	3.56 ± 0.01	1.4
	C	AACCAATTAAG	5.2 ± 0.6	2.1
	T	AATCAATTAAG	9.2 ± 0.6	3.6
4	A	AAGAAATTAAG	15.4 ± 0.9	6.1
	G	AAGGAATTAAG	7.8 ± 0.2	3.1
	T	AAGTAATTAAG	3.0 ± 0.3	1.2
5	C	AAGCCATTAAG	4.1 ± 0.3	1.6
	G	AAGCGATTAAG	3.8 ± 0.1	1.5
	T	AAGCTATTAAG	6.4 ± 0.4	2.5
6	C	AAGCACTTAAG	10.5 ± 0.7	4.2
	G	AAGCAGTTAAG	13.3 ± 0.2	5.3
	T	AAGCATTTAAG	35.3 ± 3.6	14
7	A	AAGCAAATAAG	162 ± 25	64
	C	AAGCAACTAAG	232 ± 31	92
	G	AAGCAAGTAAG	63 ± 6	25
8	A	AAGCAATAAAG	8.5 ± 0.2	3.4
	C	AAGCAATCAAG	50 ± 3	20
	G	AAGCAATGAAG	48 ± 9	19
9	C	AAGCAATTCAG	23 ± 2	8.9
	T	AAGCAATTTAG	19 ± 2	7.6
10	C	AAGCAATTACG	3.1 ± 0.2	1.2
	G	AAGCAATTAGG	9.6 ± 0.6	3.8
	T	AAGCAATTATG	7.9 ± 0.5	3.1
11	C	AAGCAATTAAC	2.9 ± 0.3	1.2
	T	AAGCAATTAAT	4.08 ± 0.03	1.6
Multiple cloning site			347 ± 37	140

[a] Equilibrium dissociation constants ± standard error of the mean. Substitution of cytosine for guanine at position 11 or a thymine for cytosine at position 4 had no significant effect on binding affinity ($P > 0.05$). All other mutations significantly decreased binding affinity ($P < 0.05$).

[b] Ratio of the equilibrium dissociation constant for substituted sequences to the equilibrium dissociation constant for CON.

are derived from clone 61, which appears to have a relatively high affinity site in the opposite orientation).

The multiple-cloning-site fragment of pUC18 served as the control for HD2 affinity for nonspecific DNA sequences (all of the synthetic binding sites were cloned into the multiple cloning site of pUC18 or pEMBL18). HD2 bound to the multiple-cloning-site fragment with a dissociation constant of 3.5×10^{-9} M, or 138-fold weaker than for the consensus sequence under the same conditions. This value reflected the homeodomain's binding affinity for the best site or sites within the 48-bp multiple-cloning-site fragment (a 3-bp match

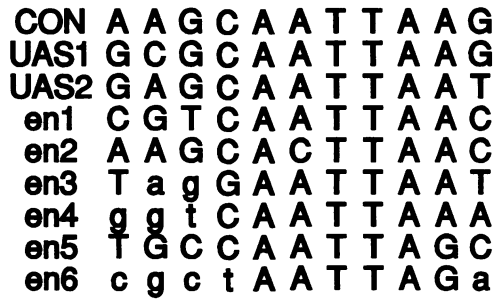

FIG. 7. Conserved matches to the *ftz* homeodomain binding-site consensus sequence within the *ftz*-activated *ftz* UAS and *en* intron enhancers. Sequences are from references 22, 23, and 41. Sites en4 through en6 correspond to the engrailed protein footprints e4 through e6 of Kassis et al. (22). Capitalized bases are conserved between *D. melanogaster* and *D. virilis*.

to the 6-bp consensus core exists within the pUC18 polylinker), and therefore the average affinity per individual site may be lower.

DISCUSSION

Under conditions in which the *ftz* homeodomain is fully active, it binds to a 59-bp fragment containing the consensus sequence AAGCAATTAAG with a K_d of 2.5×10^{-11} M. Percival-Smith et al. (40) obtained a K_d of 8×10^{-10} M for binding of the *ftz* homeodomain to an 18-bp fragment containing the sequence AAGCCATTAGA, a 32-fold difference in affinity from that found in the present study. We have shown that the differences between the two sequences at positions 5 and 10 individually cause reductions in affinity of 1.6- and 3.8-fold, respectively. Thus, these reductions account for a sixfold difference in binding affinity. The affect of adenine at position 11 and the different sizes of the DNAs assayed (59 versus 18 bp) could also have contributed to the remaining fivefold discrepancy in binding affinities.

The high affinity that *ftz* homeodomain monomers have for DNA suggests that the *ftz* protein does not need to dimerize in order to bind effectively to specific DNA sequences in vivo. Perhaps not coincidentally, two *ftz*-regulated enhancers, the *ftz* upstream activation sequence (UAS) and the *en* intron enhancer (21, 22, 41), contain very good matches to our *ftz* monomer-binding site consensus sequence (Fig. 7). These binding-site sequences are conserved in distantly related fly species (23, 29), an indication that they may be functional. In neither case are there symmetrically oriented sites that could be recognized by an *ftz* homodimer. Together, these data favor the possibility that the *ftz* protein functions as a monomer in vivo.

The consensus sequences for *Ubx*, *ftz*, *eve*, and *en* were deduced, in part, from large DNase I footprints that contained multiple homeodomain-binding sites (3, 7, 16, 35, 41). The synergistic effects of multiple homeodomain-binding sites on activation or repression of transcription (20, 51) suggest that the organization of naturally occurring binding-site clusters may be important. Our data suggest that the interactions of *ftz* homeodomains on such sequences are noncooperative. The structures of the *Antp* and *en* homeodomain-DNA complexes suggest physical reasons for this observation. The approach of two homeodomains on a tail-to-tail sequence is completely obstructed when the sites are 6 bases or less apart from center to center, which is likely

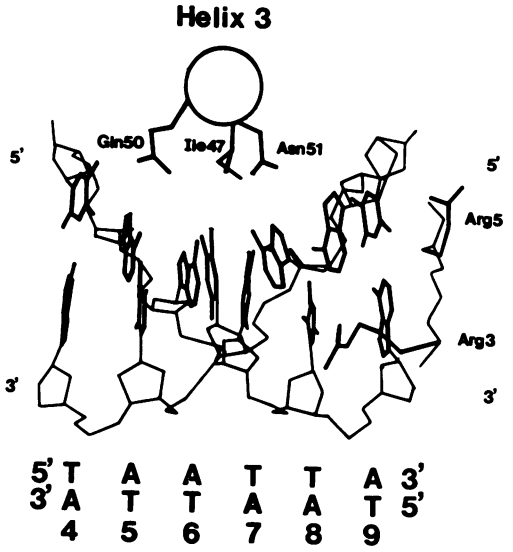

FIG. 8. Base-specific contacts of the engrailed homeodomain showing helix 3 and the N-terminal arm bound to the sequence TAATTA. The numbering corresponds to that in Tables 1 and 2. The amino acid residues that make base-specific contacts are Arg-3 and Arg-5 of the N-terminal arm and Ile-47, Gln-50, and Asn-51 of helix 3. Coordinates, kindly provided by C. Kissinger and C. Pabo, were from reference 24.

to be the result of steric interference between the N terminus of helix 3 and helix 2 of the neighboring homeodomain. Sites that are further separated (8 or 10 bp) show increasingly less inhibition. The approach of two homeodomains on head-to-head sequences is limited to four bases apart (10 bp center to center). The inhibition observed when sites are placed closer together may be the result of steric hindrances between the N termini of the homeodomains. The inhibition observed for simultaneous binding to HT4 may be due an electrostatic repulsion between the positively charged side chains of Arg-28, Arg-29, and Arg-30 of one homeodomain and the side chain of Arg-10 of the neighboring homeodomain. This interaction would have to be strong enough to inhibit approach on sites less than 12 bases apart center to center (HT6). Alteration of these side chains might allow a closer approach, perhaps even cooperative binding. Full-length homeodomain proteins might have distances of closest approach different from that of the homeodomain alone. Such differences between proteins could allow a cluster of binding sites to be recognized more efficiently by one protein than by another, even though each might bind to individual sites with the same affinity. Outside the distance of closest approach, the homeodomain has shown no inherent ability to cooperatively bind another homeodomain. However, within the context of the complete protein, cooperative interactions may take place.

The data on binding-site substitutions presented in this article, together with ethylation and methylation interference experiments (40), show that the *ftz* homeodomain monomer recognizes a span of 6 bp (CAATTA in our consensus sequence). The same base pairs make close contact with specific homeodomain side chains in the structures derived for the *Antp* and *en* homeodomain-DNA complexes (24, 38) (Fig. 8). The bases CAAT (positions 4 to 7) are contacted in the major groove by helix 3, and the bases TA (positions 8 and 9) are contacted in the minor groove by an N-terminal arm. In what follows, the effects of individual base-pair substitutions at these positions will be discussed in terms of specific base pair-amino acid contacts.

Bp 3. A change from guanine to thymine at bp 3 reduces binding affinity almost fourfold, while changes to adenine or cytosine have only a twofold effect. In addition, methylation of N-7 at guanine 3 dramatically interferes with binding of the *ftz* homeodomain (40). However, from the structures of the *Antp* and *en* homeodomain-DNA complexes, it appears that bp 3 is too far away to be contacted by helix 3, suggesting that the effects of sequence variation are indirect, perhaps through alterations in DNA structure that affect the conformations of bp 4 or 5. In contrast to our result for the *ftz* homeodomain, the binding-site consensus derived for both the *en* and *eve* proteins has a thymine at position 3 (7, 16), suggesting that this position may be recognized differently by homeodomain proteins with similar overall specificity.

Bp 4 and 5. In the context of our consensus, either cytosine or thymine is allowed at position 4, while adenine or guanine reduces binding 6.1- and 3.1-fold, respectively. Changes at position 5 have little effect on binding in the context of our consensus sequence. However, others have shown that in certain contexts, substitutions at position 5 can have major effects (13, 40). In both the *Antp* and *en* homeodomain-DNA complexes, the γCH_2 of Gln-50 makes hydrophobic contact with position 4 (C-5 of cytosine 5 in *Antp* and C-7 of thymine 4 in *en*). This contact may serve to tether the amide group of Gln-50, allowing it to pivot near the plane of bp 5 with enough freedom to allow hydrogen bonding with any nucleotide at position 5 (Fig. 9). The resulting binding affinities for substitutions at position 5 would be expected to vary somewhat because the positions of hydrogen bond donors and acceptors vary for each base relative to cytosine 4.

A thymine at position 4 would change the position of the amide and γCH_2 moieties of Gln-50, since the C-7 methyl group of thymine protrudes farther into the major groove than the C-5 of cytosine. The data in Table 2 suggest that such repositioning constrains position 5 to adenine. In the presence of thymine at position 4, either guanine or thymine at position 5 results in greatly reduced affinity (27a). Both guanine and thymine provide only hydrogen bond acceptors in the approximate position of adenine's N-6, suggesting that, in the context of thymine at position 4, Gln-50 is constrained to recognition of a proton donor at position 5. Similar hydrophobic and hydrogen bond contacts involving glutamine have been reported for the phage 434 repressor (2). This model would predict that the untested sequence TC would also be acceptable at positions 4 and 5.

Alternatively, Gln-50 might only interact directly with the adenine at bp 5, as proposed by Hanes and Brent (13). However, this model fails to account for high-affinity binding to sites lacking adenine at position 5 and the effects of substitutions at position 4 in the context of adenine at position 5.

Bp 6. In the *en* complex, Ile-47 makes hydrophobic contact with the methyl group of the thymine opposite adenine at position 6. No other base provides a methyl group for hydrophobic contact, and constructs with substitutions for thymine at this position are bound 4- to 14-fold more weakly.

Bp 7. The *en* structure shows Asn-51 in position to form hydrogen bonds with N-7 and N-6 of adenine on the strand opposite from thymine 7. A change from adenine to cytosine at this position reduces affinity 25-fold. This change might

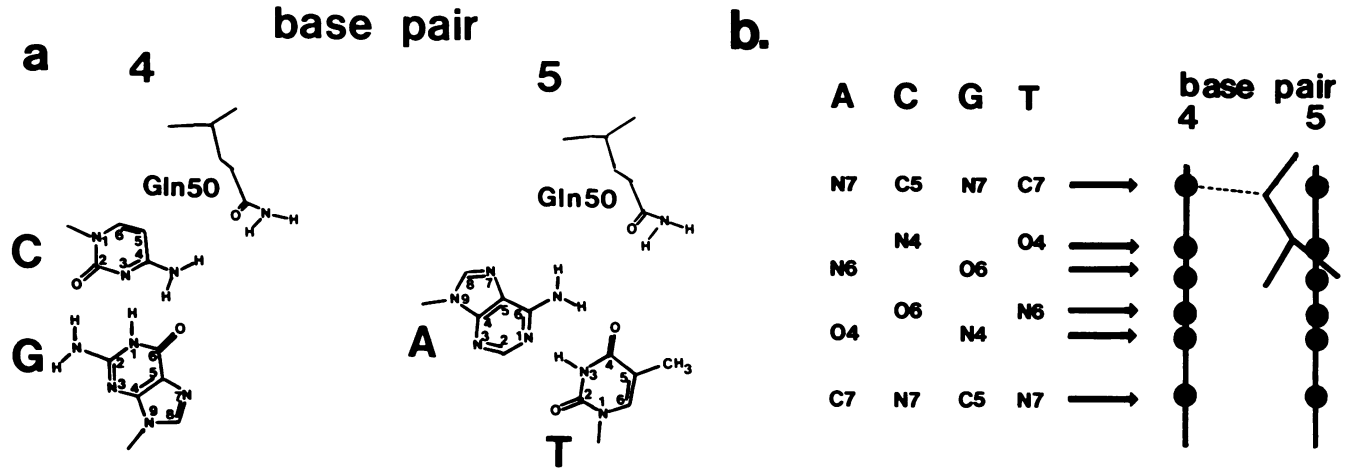

FIG. 9. (a) Position of Gln-50 with respect to possible contacts at bp 4 and 5. The Gln-50 amide is centered about one-third of the distance from bp 4 to bp 5 (also see Fig. 8). The positions of base pairs were adapted from the coordinates for T4 and A5 of Kissinger et al. (24). (b) Schematic showing the functional groups of all base pairs at positions 4 and 5 with respect to Gln-50. The dashed line represents the proposed hydrophobic contact between Gln-50 and cytosine or thymine 4.

still allow one hydrogen bond to form with N-4 of cytosine. The O-4 of thymine and N-7 of guanine could be potential hydrogen bond acceptors; however, the methyl group of thymine and O-6 of guanine would then be in position to interfere with the carbonyl of Asn-51. Affinities are reduced 73- and 100-fold when adenine is changed to thymine or guanine, respectively. The effects of mutations at this position are more severe than at any other position, consistent with the highly specific contact evident in the *en*-DNA complex and with sequence conservation of Asn-51.

Bp 8 and 9. The structure of the *en* complex shows that the thymines of bp 8 and 9 are contacted in the minor groove by Arg-3 and Arg-5, respectively. Both contacts appear to be hydrogen bonds to O-2 of thymine. A change from thymine to adenine at position 8 reduced binding only 3.4-fold, suggesting that Arg-3 is flexible enough to contact O-2 on the opposite strand. Even though cytosine also provides O-2, cytosine or guanine at this position reduced affinity 20-fold. This result could be explained if Arg-3 makes an additional minor-groove contact that is not provided by a CG base pair. Indeed, the *en* crystal structure shows that the η-amine of Arg-3 is only slightly farther from the N-3 of adenine than from the O-2 of thymine at position 8. Approach to both the O-2 of cytosine and the N-3 of guanine in a CG base pair might be inhibited by the N-2 of guanine. Arg-5 appears to be both less flexible and less critical for binding, since either cytosine or thymine at position 9 reduced affinity about eightfold. Recognition at both positions 8 and 9 may be influenced by the residues that flank the homeodomain in normal proteins. As noted by Kissinger et al. (24), the homeodomain is the only transcription factor known to contact DNA in the minor groove, yet our data show that such contacts can contribute substantially to DNA-binding specificity.

Most changes from the consensus sequence at bp 1, 2, 3, 5, 10, and 11 result in about a twofold reduction in binding affinity. Changes in binding affinity caused by substitutions at these positions are consistent with the ethylation and methylation interference data from Percival-Smith et al. (40). With the exception of bp 5, as discussed above, it seems unlikely that the *ftz* homeodomain makes direct contact with these bases in either the major or minor groove. These effects may be due to changes in the structure of DNA transmitted over one to several base pairs. Although the mechanisms responsible for such "context effects" are not well understood, the data in Tables 2 and 3 show that, for homeodomains, sequences flanking regions of base-specific contact have substantial effects on binding affinity. Specifically, binding-site clone 47 is bound with a 6- to 11-fold-higher affinity than clones 37, 39, 41, 42, and 55, even though each differs only outside the region of base-specific contact. Since regulation by some homeodomain proteins (i.e., *Ubx*, *Scr*, ftz^{Ual}) is dose dependent, it is conceivable that relatively minor variations in sequence affinity could have significant effects on regulation of transcription. A better understanding of how sequences flanking the core binding site influence the binding of homeodomain-containing transcription factors in vivo may help in identifying bona fide targets of regulation.

We have shown that, in isolation, an *ftz* homeodomain monomer recognizes primarily 6 bp which, from structural studies, appear to be sites of base-specific homeodomain contact. The amino acids involved directly in base-specific contact (N terminus and helix 3) are conserved among most of the characterized homeodomain proteins, including the products of the Antennapedia and Bithorax complex homeotic genes. This fact could be taken to suggest that these proteins should recognize very similar, if not identical, DNA sequences. This view is somewhat in conflict with recent experiments, which show that the different regulatory specificities of several homeotic proteins lie mainly within or near their homeodomains (9, 27, 30). Perhaps differences exist in the intrinsic DNA-binding specificities of these closely related homeodomains, possibly the result of conformational differences that alter the positions of base-contacting amino acid side chains.

A second possibility is that homeodomains also interact specifically with cofactor proteins, perhaps through residues that are less conserved than those of helix 3. Indeed, this appears to be the case for interaction between a mammalian homeodomain protein, Oct-1, and α-TIF (26). It is also known that the helix-turn-helix lambda repressor protein activates transcription through helix 2 (4). A requirement for cofactors is indicated by experiments which show that *en*

and *ftz* binding sites alone are insufficient to express a basal promoter in embryos under the direct control of *en*, *ftz*, or other closely related homeodomain proteins (35a, 51). The DNA-binding specificities of homeodomain proteins may also be determined in part by sequences flanking the homeodomains. This has been shown to be the function of the POU-specific domain of Oct-1 (26).

ACKNOWLEDGMENTS

We thank Stephen Lo for sequencing of clones; Sean Carroll, F. Michael Hoffmann, and members of our lab for comments on the manuscript; Claude Desplan for suggestions on protein purification; and both Charles Kissinger and Carl Pabo for providing engrailed structural coordinates. We also thank Fred Blattner and Shuang-En Chuang for generously sharing their Betascope.

This work was supported by Public Health Service grant HD23820 and by Basil O'Conner Junior Investigator and Shaw Scholar awards to A.L.

REFERENCES

1. **Affolter, M., A. Percival-Smith, M. Muller, W. Leupin, and W. J. Gehring.** 1990. DNA binding properties of the purified *Antennapedia* homeodomain. Proc. Natl. Acad. Sci. USA **87:** 4093–4097.
2. **Anderson, J. E., M. Ptashne, and S. C. Harrison.** 1987. Structure of the repressor-operator complex of bacteriophage 434. Nature (London) **326:**846–852.
3. **Beachy, P. A., M. A. Krasnow, E. R. Gavis, and D. S. Hogness.** 1988. An *Ultrabithorax* protein binds sequences near its own and the *Antennapedia* P1 promoters. Cell **55:**1069–1081.
4. **Bushman, F. D., C. Sheng, and M. Ptashne.** 1989. A single glutamic acid residue plays a key role in the transcriptional activation function of lambda repressor. Cell **58:**1163–1171.
5. **Cantor, C. R., and P. Schimmel.** 1980. Biophysical chemistry. W.H. Freeman & Co., New York.
6. **Carroll, S. B., and M. P. Scott.** 1986. Zygotically active genes that affect the spatial expression of the *fushi tarazu* segmentation gene during early Drosophila development. Cell **45:**113–126.
7. **Desplan, C., J. Theis, and P. H. O'Farrell.** 1988. The sequence specificity of homeodomain-DNA interaction. Cell **54:**1081–1090.
8. **Driever, W., and C. Nusslein-Volhard.** 1989. The *bicoid* protein is a positive regulator of *hunchback* transcription in the early Drosophila embryo. Nature (London) **337:**138–143.
9. **Gibson, G., A. Schier, P. K. LeMotte, and W. J. Gehring.** 1990. The specificities of *Sex combs reduced* and *Antennapedia* are defined by a distinct portion of each protein that includes the homeodomain. Cell **62:**1087–1103.
10. **Hafen, E., M. Levine, and W. J. Gehring.** 1984. Regulation of *Antennapedia* transcript distribution by the *Bithorax* complex in *Drosophila*. Nature (London) **307:**287–289.
11. **Han, K., M. S. Levine, and J. L. Manley.** 1989. Synergistic activation and repression of transcription by Drosophila homeobox proteins. Cell **56:**573–583.
12. **Hanes, S. D., and R. Brent.** 1989. DNA specificity of the *bicoid* activator protein is determined by homeodomain recognition helix residue 9. Cell **57:**1275–1283.
13. **Hanes, S. D., and R. Brent.** 1991. A genetic model for interaction of the homeodomain recognition helix with DNA. Science **251:**426–430.
14. **Harding, K., C. Wedeen, W. McGinnis, and M. Levine.** 1985. Spatially regulated expression of homeotic genes in *Drosophila*. Science **229:**1236–1242.
15. **Hayashi, S., and M. P. Scott.** 1990. What determines the specificity of action of Drosophila homeodomain proteins? Cell **63:**883–894.
16. **Hoey, T., and M. Levine.** 1988. Divergent homeo box proteins recognize similar DNA sequences in *Drosophila*. Nature (London) **332:**858–861.
17. **Hope, I. A., and K. Struhl.** 1987. GCN4, a eukaryotic transcriptional activator protein, binds as a dimer to target DNA. EMBO J. **6:**2781–2784.
18. **Howard, K. R., and P. W. Ingham.** 1986. Regulatory interactions between the segmentation genes *fushi tarazu*, *hairy*, and *engrailed* in the Drosophila blastoderm. Cell **44:**949–957.
19. **Ingham, P. W., and A. Martinez-Arias.** 1986. The correct activation of Antennapedia and bithorax complex genes requires the *fushi tarazu* gene. Nature (London) **324:**592–597.
20. **Jaynes, J. B., and P. H. O'Farrell.** 1988. Activation and repression of transcription by homeodomain-containing proteins that bind a common site. Nature (London) **336:**744–749.
21. **Kassis, J. A.** 1990. Spatial and temporal control elements of the Drosophila *engrailed* gene. Genes Dev. **4:**433–443.
22. **Kassis, J. A., C. Desplan, D. K. Wright, and P. H. O'Farrell.** 1989. Evolutionary conservation of homeodomain-binding sites and other sequences upstream and within the major transcription unit of the *Drosophila* segmentation gene engrailed. Mol. Cell. Biol. **9:**4304–4311.
23. **Kassis, J. A., S. J. Poole, D. K. Wright, and P. H. O'Farrell.** 1986. Sequence conservation in the protein coding and intron regions of the *engrailed* transcription unit. EMBO J. **5:**3583–3589.
24. **Kissinger, C. R., B. Liu, E. Martin-Blanco, T. B. Kornberg, and C. O. Pabo.** 1990. Crystal structure of an *engrailed* homeodomain-DNA complex at 2.8Å resolution: a framework for understanding homeodomain-DNA interactions. Cell **63:**579–590.
25. **Krasnow, M. A., E. E. Saffman, K. Kornfeld, and D. S. Hogness.** 1989. Transcriptional activation and repression by *Ultrabithorax* proteins in cultured Drosophila cells. Cell **57:**1031–1043.
26. **Kristie, T. M., and P. A. Sharp.** 1990. Interactions of the Oct-1 POU subdomains with specific DNA sequences and with the HSV A-trans-activator protein. Genes Dev. **4:**2383–2396.
27. **Kuziora, M. A., and W. McGinnis.** 1989. A homeodomain substitution changes the regulatory specificity of the *Deformed* protein in Drosophila embryos. Cell **59:**563–571.

27a. **Laughon, A.** Unpublished data.

28. **Laughon, A., and M. P. Scott.** 1984. Sequence of a *Drosophila* segmentation gene: protein structure homology with DNA-binding proteins. Nature (London) **310:**25–31.
29. **Maier, D., A. Preiss, and J. R. Powell.** 1990. Regulation of the segmentation gene *fushi tarazu* has been functionally conserved in Drosophila. EMBO J. **9:**3957–3966.
30. **Mann, R. S., and D. S. Hogness.** 1990. Functional dissection of *Ultrabithorax* proteins in D. melanogaster. Cell **60:**597–610.
31. **Marston, F. A. O.** 1987. The purification of eukaryotic polypeptides expressed in *Escherichia coli*, p. 59–88. *In* D. M. Glover (ed.), DNA cloning, vol. III. IRL Press, Oxford.
32. **McGhee, J. D., and P. H. von Hippel.** 1974. Theoretical aspects of DNA-protein interactions: co-operative and non-co-operative binding of large ligands to a one-dimensional homogeneous lattice. J. Mol. Biol. **86:**469–489.
33. **McGinnis, W., R. L. Garber, J. Wirz, A. Kuroiwa, and W. J. Gehring.** 1984. A homologous protein-coding sequence in Drosophila homeotic genes and its conservation in other metazoans. Cell **37:**403–408.
34. **Muller, M., M. Affolter, W. Leupin, G. Otting, K. Wuthrich, and W. J. Gehring.** 1988. Isolation and sequence-specific DNA binding of the *Antennapedia* homeodomain. EMBO J. **7:**4299–4304.

34a. **Na, G. C., and S. N. Timasheff.** 1981. Interaction of calf brain tubulin with glycerol. J. Mol. Biol. **151:**165–178.

35. **Nelson, H. B., and A. Laughon.** 1990. The DNA binding specificity of the Drosophila *fushi tarazu* protein: a possible role for DNA bending in homeodomain recognition. New Biol. **2:**171–178.

35a. **Nelson, H. B., and A. Laughon.** Unpublished data.

36. **Odenwald, W. F., J. Garbern, H. Arnheiter, E. Tournier-Lasserver, and R. A. Lazzarini.** 1989. The Hox-1.3 homeo box protein is a sequence-specific DNA-binding phosphoprotein. Genes Dev. **3:**158–172.
37. **Oliphant, A. R., and K. Struhl.** 1987. The use of random-sequence oligonucleotides for determining consensus se-

quences. Methods Enzymol. **155**:568–582.
38. **Otting, G., Y. Q. Qian, M. Billeter, M. Muller, M. Affolter, W. J. Gehring, and K. Wuthrich.** 1990. Protein-DNA contacts in the structure of a homeodomain-DNA complex determined by nuclear magnetic resonance spectroscopy in solution. EMBO J. **9**:3085–3092.
39. **Pabo, C. O., and R. T. Sauer.** 1984. Protein-DNA recognition. Annu. Rev. Biochem. **53**:293–321.
40. **Percival-Smith, A., M. Muller, M. Affolter, and W. J. Gehring.** 1990. The interaction with DNA of wild-type and mutant *fushi tarazu* homeodomains. EMBO J. **9**:3967–3974.
41. **Pick, L., A. Schier, M. Affolter, T. Schmidt-Glenewinkel, and W. J. Gehring.** 1990. Analysis of the ftz upstream element: germ layer-specific enhancers are independently autoregulated. Genes Dev. **4**:1224–1239.
42. **Qian, Y. Q., M. Billeter, G. Otting, M. Muller, W. J. Gehring, and K. Wuthrich.** 1989. The structure of the *Antennapedia* homeodomain determined by NMR spectroscopy in solution: comparison with prokaryotic repressors. Cell **59**:573–580.
43. **Sambrook, J., E. F. Fritsch, and T. Maniatis.** 1989. Molecular cloning: a laboratory manual, 2nd ed. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
44. **Samson, M.-L., L. Jackson-Grusby, and R. Brent.** 1989. Gene activation and DNA binding by Drosophila *Ubx* and *abd-A* proteins. Cell **57**:1045–1052.
45. **Scott, M. P., J. W. Tamkun, and G. W. Hartzell III.** 1989. The structure and function of the homeodomain. Biochim. Biophys. Acta **989**:25–48.
46. **Scott, M. P., and A. J. Weiner.** 1984. Structural relationships among genes that control development: sequence homology between the *Antennapedia*, *Ultrabithorax*, and *fushi tarazu* loci of *Drosophila*. Proc. Natl. Acad. Sci. USA **81**:4115–4119.
47. **Snedecor, G. W., and W. G. Cochran.** 1967. Statistical methods. The Iowa State University Press, Ames, Iowa.
48. **Studier, F. W., A. H. Rosenberg, J. J. Dunn, and J. W. Dubendorf.** 1990. Use of T7 RNA polymerase to direct the expression of cloned genes. Methods Enzymol. **185**:60–88.
49. **Tautz, D.** 1988. Regulation of the *Drosophila* segmentation gene *hunchback* by two maternal morphogenetic centers. Nature (London) **332**:281–284.
50. **Treisman, J., P. Conczy, M. Vashishtha, E. Harris, and C. Desplan.** 1989. A single amino acid can determine the DNA binding specificity of homeodomain proteins. Cell **59**:553–562.
51. **Vincent, J.-P., J. A. Kassis, and P. H. O'Farrell.** 1990. A synthetic homeodomain binding site acts as a cell type specific, promoter specific enhancer in Drosophila embryos. EMBO J. **9**:2573–8.
52. **Winslow, G. M., S. Hayashi, M. A. Krasnow, D. S. Hogness, and M. P. Scott.** 1989. Transcriptional activation by the *Antennapedia* and *fushi tarazu* proteins in cultured Drosophila cells. Cell **57**:1017–1030.