MOLECULAR AND CELLULAR BIOLOGY, Oct. 2000, p. 7673–7684 Vol. 20, No. 20
0270-7306/00/$04.00+0

Reprogrammable Recognition Codes in Bicoid Homeodomain-DNA Interaction

VRUSHANK DAVE,[1] CHEN ZHAO,[1] FAN YANG,[1] CHANG-SHUNG TUNG,[2] AND JUN MA[1]*

Division of Developmental Biology, Children's Hospital Research Foundation, Cincinnati, Ohio 45229,[1] *and Theoretical Biology and Biophysics (T-10), Theoretical Division, Los Alamos National Laboratory, Los Alamos, New Mexico 87545*[2]

Received 17 March 2000/Returned for modification 1 May 2000/Accepted 18 July 2000

We describe experiments to determine how the homeodomain of the *Drosophila* morphogenetic protein Bicoid recognizes different types of DNA sequences found in natural enhancers. Our chemical footprint analyses reveal that the Bicoid homeodomain makes both shared and distinct contacts with a consensus site A1 (TAATCC) and a nonconsensus site X1 (TAAGCT). In particular, the guanine of X1 at position 4 (TAAGCT) is protected by Bicoid homeodomain. We provide further evidence suggesting that the unique arginine at position 54 (Arg 54) of the Bicoid homeodomain enables the protein to recognize X1 by specifically interacting with this position 4 guanine. We also describe experiments to analyze the contribution of artificially introduced Arg 54 to DNA recognition by other Bicoid-related homeodomains, including that from the human disease protein Pitx2. Our experiments demonstrate that the role of Arg 54 varies depending on the exact homeodomain framework and DNA sequences. Together, our results suggest that Bicoid and its related homeodomains utilize distinct recognition codes to interact with different DNA sequences, underscoring the need to study DNA recognition by Bicoid-class homeodomains in an individualized manner.

A homeodomain is an evolutionarily conserved domain found in many DNA-binding transcription factors that control such biological processes as cell type specification and embryonic pattern formation (26). The homeodomain is responsible for recognizing specific DNA sequences to bring the transcription factors to proper target genes. This 60-amino-acid domain is composed of three helices and a flexible amino-terminal arm (27, 73). The DNA-binding specificity of a homeodomain is determined primarily by its third helix, called the recognition helix, which inserts itself into the major groove of the recognition site. In addition, the flexible amino-terminal arm wraps around DNA and makes specific contacts in the minor groove. The second helix of a homeodomain also makes DNA backbone contacts, further contributing to specific homeodomain-DNA interactions. The recognition sites for most homeodomains have a common "TAAT" core, which is followed by two residues that confer differential binding specificity (63). It has been proposed that the 9th position of the recognition helix (the 50th position of the homeodomain) plays a critical role in differential DNA recognition (34, 62). In particular, homeodomains containing a glutamine residue at the 50th position (referred to as the Q50 class) prefer a TAATGG sequence. In contrast, homeodomains containing a lysine residue at this position (referred to as the K50 class) recognize TAATCC.

Bicoid (Bcd), a *Drosophila* homeodomain-containing protein, is required for establishing the polarity along the anterior-posterior axis of the early embryo (10). The protein is encoded by the maternal gene *bicoid* (4) and is distributed along an anterior-to-posterior gradient in the embryo (14). The Bcd gradient instructs the formation of the anterior structures, including the head and thorax, by activating zygotic genes in a concentration-dependent manner (11–13, 15, 53, 59). The homeodomain of Bcd, which is of the K50 class, recognizes DNA sequences found in enhancer elements of Bcd-responsive genes such as *hunchback* (*hb*), *knirps* (*kni*), *buttonhead* (*btd*), *runt* (*run*), *hairy* (*h*), *orthodenticle* (*otd*) and *even-skipped* (*eve*) (6, 13, 25, 34, 37, 58, 65, 70, 71, 77). A comparison of the natural Bcd sites (13) and in vitro site selection experiments (70, 76) have revealed a consensus site, TAATCC. However, the Bcd homeodomain can also recognize sequences that deviate from this consensus, including those that do not even have a TAAT core. At least three types of nonconsensus sites can be classified according to their core sequences: TAAGCC, TGATCC, and AAATCC (13, 53, 76).

Previous studies have demonstrated that nonconsensus DNA sites play an important role in mediating Bcd function. For example, a Bcd-responsive enhancer element of the *kni* gene, which is activated by Bcd, does not have any perfect TAATCC sequence (53). In addition, multimerized nonconsensus sites taken from an enhancer element of the Bcd-responsive gene *hb* can respond to the Bcd gradient in *Drosophila* embryos (15). Our site-directed mutagenesis analysis has revealed a particularly important role of nonconsensus sites in supporting transcriptional activation by Bcd in the context of the natural *hb* enhancer element (42, 76). More recently, we have shown that a homeodomain protein derivative, Ftz (Q50K), which fails to recognize nonconsensus sites, also fails to activate transcription from natural enhancer elements (78). Together, these findings suggest that the ability of Bcd to recognize nonconsensus sites is an essential function in executing its biological activity.

Despite their biological importance in mediating Bcd function, nothing is known about how nonconsensus sites are recognized by the Bcd homeodomain. Recent structural studies of other types of DNA-binding proteins have revealed two strikingly different strategies for a given protein to recognize different types of sites (16, 17, 49, 56). In some cases (e.g., estrogen receptor and Zif268) the protein uses different recognition codes to interact with different DNA sequences, whereas in other cases (e.g., TATA box-binding protein) the molecular interactions with different sequences remain virtually identical.

* Corresponding author. Mailing address: Division of Developmental Biology, Children's Hospital Research Foundation, 3333 Burnet Ave., Cincinnati, OH 45229. Phone: (513) 636-7977. Fax: (513) 636-4317. E-mail: jun.ma@chmcc.org.

Thus, two different models can be proposed to explain how the Bcd homeodomain recognizes different sequences. A "rigid" model proposes that the Bcd homeodomain employs an identical (or similar) recognition code for both consensus and nonconsensus sites. Consequently, some specific interactions are either completely lost or severely compromised at the deviating nucleobases of a nonconsensus site. Alternatively, an "adaptive" model proposes that the Bcd homeodomain can adjust itself structurally to establish a different recognition code for a nonconsensus site. According to this model, the deviating nucleobases in a nonconsensus site represent novel opportunities for the Bcd homeodomain to make specific new contacts.

In this report, we describe experiments to probe the interactions between the Bcd homeodomain and two different types of sites, A1 and X1 from the *hb* enhancer element. While A1 has a consensus Bcd-binding site of TAATCC, X1 exemplifies a nonconsensus site (TAAGCT), with a TAAG core followed by only one C. Our chemical footprint assays reveal both shared and distinct contacts with A1 and X1, suggesting that the Bcd homeodomain docks on these sites with a similar overall structure but different sets of interactions. We provide further evidence suggesting that arginine at position 54 of the Bcd homeodomain (Arg 54) enables the protein to recognize X1 by specifically interacting with the guanine at position 4 (TAAGCT). These results support the adaptive model and suggest that the Bcd homeodomain uses reprogrammable recognition codes for different DNA sites. We also demonstrate that the role of Arg 54 in DNA recognition varies depending on the homeodomain framework and DNA sites, and we suggest that different homeodomains use distinct recognition codes to interact with a given DNA sequence.

MATERIALS AND METHODS

Plasmid construction and mutagenesis. Mutations at positions 50 and 54 in all the homeodomains were generated by a PCR-mediated method. BcdTN3 was used as the template (11) for Bcd homeodomain mutation. pCZ2046, pVD47, and pVD48 have R54A, R54K, and K50R mutations to the Bcd homeodomain, respectively. The Otd, Boz, Pitx2, and Ftz homeodomains (pVD50, pVD51, pVD52, and pCZ58) and their mutants containing Arg 54 (pVD53, pVD54, pVD55, and pVD56) were generated by PCR using their full-length cDNA as templates, respectively (20, 31, 57). Full-length Boz (Dharma) was a gift from Wolfgang Driever. For PCR, C-terminal primers that covered positions 50 and/or 54 were used and the products were cleaved by *Eco*RI and cloned at the *Eco*RI site in pGEX-1λT vector. A modified PCR was used in which the middle primer carried the required site-specific mutation. The PCR reaction and identification of the mutant fragment was carried out according to Ma et al. (42). The DNA products were cut with *Eco*RI and cloned in pGEX-1λT vector and confirmed by DNA sequencing. Plasmids pFY403, pCZ2076, and pCZ2078 were used to express full-length wild-type Bcd, the Bcd(R54A) mutant, and the Bcd(K50R) mutant, respectively. All these plasmids are based on pAc5.1/V5-HisC (Invitrogen) with an in-frame hemagglutinin (HA) tag at the N terminus of Bcd sequences. The construction of pFY403 is described elsewhere (78). pCZ2076 was created in two steps: BcdTN3 was first used in PCR reaction as the template to generate a point mutation at position 54 of the Bcd homeodomain; the resulting PCR product was then inserted into pFY7003 (78) as a *Nde*I-*Sac*II fragment to generate pCZ2067. To make pCZ2076, the Bcd sequence was excised from pCZ2067 as a *Hin*dIII-*Xba*I fragment and subcloned into pFY404. pFY404 is a derivative of pAc5.2/V5-HisC (Invitrogen) made by inserting a *Xba*I stop codon-containing linker into the *Eco*RI site. pCZ2077 was generated similarly to pCZ2076, except that a different mutagenesis primer for PCR was initially employed. Mutations to Bcd in pCZ2076 and pCZ2077 were confirmed by DNA sequencing. The reporter constructs pCZ3005, containing native *hb-CAT*, and pCZ3007 containing *hb(6A)-CAT*, were created from G1E1bCAT (39). pCZ3005 contains a 250-bp *hb* natural enhancer element upstream of the adenovirus E1b TATA sequence (41). The detailed construction of pCZ3005 and pCZ3007 is described elsewhere (78).

Recombinant homeodomains and gel mobility shift assays. All homeodomain proteins (containing 60 amino acids) used in our assay were expressed in bacteria using the pGEX-1-λT expression system (Amersham-Pharmacia Biotech). After the respective glutathione *S*-transferase fusion proteins were purified, the glutathione *S*-transferase tag was removed using thrombin as specified by the manufacturer. The cleaved homeodomains were then dialyzed against our gel shift binding buffer (20 mM Tris [pH 7.5], 50 mM NaCl, 0.5 mM EDTA) and stored at −20°C. The Bcd(R54A) mutant homeodomain was stored at −80°C in thrombin cleavage buffer (60 mM Tris [pH 7.5], 150 mM NaCl, 1.5 mM EDTA) for improved solubility and stability; this protein was diluted threefold immediately prior to each binding assay. Gel mobility shift assays were carried out using Bcd and related homeodomains and various ^{32}P-labeled double-stranded DNA probes. The oligonucleotides used for making the A1, X1, and X3 probes, for both sense and antisense strands (5′ to 3′), were CTAGGACCACCAACGTAATCCCCATAG plus AGCTCTATGGGGATTACGTTGGTC, CTAGCTCGCTGCTAAGCTGGCCAT plus AGCTATGGCCAGCTTAGCAGCGAG, and CTAGATCTGCTCTGATCCAGAATG plus TCGACATTCTGGATCAGAGGCAGAT, respectively. For Hybrid-A1 used in the methylation interference assay, the sense and antisense probes were CTAGCTCGCTGCTAATCCGGCCAT and TCGAATGGCCGGATTAGCAGCGAG, respectively. For Hybrid-A1 used in the methylation protection assay, the sense and antisense probes were CTAGCTCGCTGCTAATCCGGCCATTCGA and TCGAATGGCCGGATTAGCAGCGAGCTAG, respectively. Briefly, 1 μg of annealed double-stranded DNA was 5′-end labeled in a standard T4 polynucleotide kinase reaction mixture containing 10 μl of [γ-^{32}P]ATP and purified over a G-25 Sephadex column to remove free nucleotide and salt (44). The binding reactions were carried out at room temperature for 30 min in 30 μl of Bcd binding buffer containing 1 nM active homeodomain proteins and 1 nM DNA probe unless stated otherwise. The bound protein was separated from the free probe on a 15% polyacrylamide gel (acrylamide/bisacrylamide ratio, 29:1) containing 0.5× TBE by polyacrylamide gel electrophoresis (PAGE) at 10 V/cm, and the images were obtained on a Molecular Dynamics PhosphorImager system. The active homeodomain proteins present in the preparations were estimated by a gel mobility shift assay at a saturating concentration of A1 site probe (5×10^{-6} M).

Methylation interference assay. Dimethyl sulfate (DMS) methylates preferentially the N-7 position of guanine in the major groove and to a much lesser extent the N-3 position of adenine in the DNA minor groove (45). This chemical modification of a particular base, when carried out prior to protein binding, will interfere with DNA-protein complex formation either due to steric hindrance if a particular amino acid is nearby or due to a loss of contact if the base directly participates in a specific interaction. In a methylation interference assay, uniquely 5′-end-labeled (as mentioned above) sense and antisense double-stranded probes were partially methylated in a standard Maxam-Gilbert DMS reaction and used for preparative gel mobility shift assays. Typically, 10 pmol of DNA was dissolved in 200 μl of cacodylate buffer (50 mM sodium cacodylate [pH 8.0], 1 mM disodium EDTA) and placed on ice for 20 min. The methylation reaction was started by addition of 1 μl of DMS solution (Maxam-Gilbert sequencing kit; Sigma Chemical Co.) and continued for 20 min. The reaction was stopped by the addition of 50 μl of stop buffer (1.4 M β-mercaptoethanol, 20 μg of yeast tRNA, 1.5 M sodium acetate [pH 7.1]) followed by the immediate addition of 750 μl of chilled absolute ethanol and incubation on dry ice for 20 min. The precipitated DNA was subsequently dissolved in 250 μl of 0.3 M sodium acetate (pH 5.3), reprecipitated, washed twice with 2 volumes of 100 and 70% ethanol, respectively, and, after being dried, used directly in a preparative gel mobility shift reaction with ~10 nM protein and 1 μg of poly(dI-dC). Using autoradiography, the gel areas containing bound and free DNA probes were identified and cut out, and the DNA was eluted (2). Eluted DNA was further concentrated on a vacuum dryer, ethanol precipitated, washed with 70% ethanol, dried, dissolved in 100 μl 1 M piperidine, and heated at 90°C for 30 min. Piperidine was removed by extensive vacuum drying, and the DNA was dissolved in Sequencing Load Buffer (90% formamide in 10 mM Tris.Cl [pH 8.0], 0.05% [wt/vol] bromophenol blue, and 0.05% xylene cyanol). Approximately 2,000 cpm of DNA per lane was loaded on a prerun 20% polyacrylamide (29:1) sequencing gel. The gels were exposed overnight on a PhosphorImager screen, and the images were obtained on a PhosphorImager.

Methylation protection assay. In the methylation protection assay, the methylation of guanines (and to a lesser extent adenines) is carried out using DMS in a binding-reaction mix in which the DNA is already bound by the protein. The protection assay for the Bcd homeodomain on A1, X1, Hybrid-X1, and X3 was carried out in a similar manner to the DMS interference assay, except that the DNA was subjected to methylation by DMS after protein binding. Briefly, 1 μl of DMS was added to 300 μl of preincubated binding-reaction mixture at room temperature containing 10 nM protein, 1 μg of poly(dI-dC), and 10 pmol of labeled DNA and the reaction was continued for 2 min. The reaction mixture was immediately subjected to fast-running polyacrylamide gel electrophoresis at 30 V/cm. The subsequent steps were the same as in methylation interference assays.

Thymine-specific interference footprinting. $KMnO_4$ releases permanganate ions, a strong oxidant that reacts with DNA bases selectively on a single-stranded DNA (1). In unbuffered solution or water, thymine is by far the most commonly $KMnO_4$-oxidized base in single-stranded DNA, although a faint background of oxidized cytosines and guanosines and a position-dependent variability in the reactivity of individual thymines have been described (54). However, in the presence of 30 mM Tris buffer (pH 8.0), such position effects are no longer observed (64), a condition that we used for our experiments. It is known that $KMnO_4$ attacks the double bond of thymine between C-5 and C-6 in single-stranded DNA and forms a glycol, thereby altering its electronic state (28, 64). After the $KMnO_4$ reaction, the single-stranded DNA is annealed to its antisense strand and the double-stranded DNA obtained is used for protein binding and

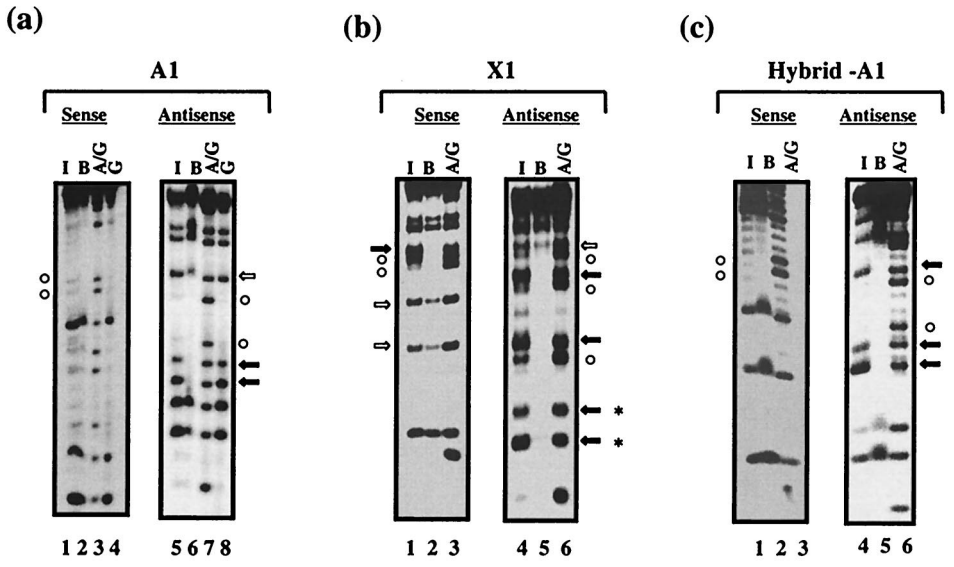


FIG. 1. A1 and X1 exhibit different methylation interference patterns for Bcd homeodomain binding. (a and b) A methylation interference analysis (see Materials and Methods for details) was performed using the Bcd homeodomain on both strands of DNA probes containing A1 (a) or X1 (b). (c) A third probe, Hybrid-A1, contains the recognition site A1 in the flanking sequences of X1; therefore, Hybrid-A1 and X1 probes are identical except for their recognition sequences. I and B represent the methylation profiles of DNA isolated from interfered (unbound) and bound fractions, respectively. In this assay, a band missing in the bound fraction indicates that methylation at this position interferes with (prevents) Bcd homeodomain binding. Strongly interfered guanine positions are marked with solid arrows, while partially interfered guanines are marked with open arrows. The interfered adenines are marked with open circles. The methylation interference data are summarized with the DNA sequences in Fig. 4. Two guanines on the antisense strand of X1 are highlighted with asterisks; these two positions exhibit strong methylation interference on X1 probe (b) but not on the Hybrid-A1 probe (c). A/G and G are Maxam-Gilbert DNA sequencing ladders.

subsequent interference experiments. Briefly, 10 pmol of 5′-end-labeled single-stranded DNA dissolved in 5 μl of 30 mM Tris.Cl (pH 8.0) was treated with 20 μl of 0.25 mM $KMnO_4$ and incubated for 10 min at 20°C. The reaction was stopped by addition of 50 μl of DMS stop buffer (see above) and 175 μl of chilled H_2O. The DNA was immediately precipitated in 2 volumes of chilled ethanol, washed with 70% ethanol, and dried. The treated single-stranded DNA was annealed to its opposite strand in 50 μl of a solution containing 25 mM NaCl and 10 mM $MgCl_2$ and confirmed to be present in double-stranded form by PAGE on a native 20% polyacrylamide gel before the interference experiment was carried out. The Bcd homeodomain binding reaction was carried out as described above, free and bound DNAs obtained from preparative gel mobility shift assays were eluted for piperidine cleavage, and sequencing gels were run to identify the interfered thymines. A control $KMnO_4$ reaction with single-stranded DNA (24-mer oligonucleotide) alone was carried out with subsequent piperidine cleavage to make sure that the DNA had not undergone any intramolecular interaction (base pairing) that would result in a protected thymine pattern complicating our interference experiment with the Bcd homeodomain.

Cell culture and transfection. *Drosophila* Schneider S2 cells were cultured at 25°C in DES expression medium (Invitrogen) supplemented with 10% fetal bovine serum. The cells were seeded in 60-mm-diameter tissue culture plates at roughly 5×10^6 cells/plate. After 24 h, transfection was performed by the calcium phosphate coprecipitation method (Gibco BRL kit). A total of 10 μg of DNA containing 1 μg of reporter, 0.2 μg of effector plasmid, 1 μg of p*Copia-lacZ* plasmid as an internal control, and 7.8 μg of empty pAc5.1 vector (Invitrogen) was used per transfection. The transfected cells were harvested 48 h later, and cell lysates were prepared by the freeze-thaw method as described previously (2). The transfection efficiency was determined by monitoring the β-galactosidase activity, and the amount of lysate used in the chloramphenicol acetyltransferase (CAT) assay and Western blotting was normalized accordingly. Derivatized chloramphenicol was quantitated with a volume integration function on a PhosphorImager. For Western analysis, cell lysates were separated on a sodium dodecyl sulfate–10% polyacrylamide gel and transferred to a cellulose membrane. The appropriate protein bands were visualized with an anti-HA monoclonal antibody (HA.11 [Babco]; 1:600 final dilution) and subjected to enhanced chemiluminescence analysis (Pharmacia Amersham Biotech).

Molecular modeling of the Bcd homeodomain-DNA complex. The structure of the Bcd homeodomain-DNA complex was determined using a homology-modeling approach (67), with the template being the crystal structure of the Engrailed (En) homeodomain-DNA complex (22). In the region we modeled, the two homeodomains (Bcd and En) can be aligned with a 45% sequence identity involving no insertion or deletion. The DNA molecule in the target complex contains the TAATCC, the consensus site A1 (see Fig. 9a), or TAAGCT, the nonconsensus site X1 (see Fig. 9b). The modeled structure of the complex was energy minimized using AMBER (69). In both structures, Arg 54 (colored in magenta) points in the major groove of the DNA (colored in yellow) and is well positioned to form hydrogen bonds with the bases. For A1, the third-base adenine N7 (TA<u>A</u>TCC; the corresponding base pair is colored in cyan) is positioned to form a single hydrogen bond with Arg 54. For X1, the Arg 54 side chain is translated vertically to form bidentate hydrogen bonds, one with N-7 of the third-base adenine and other with O-6 of the fourth-base guanine (TAA<u>G</u>CC) (the corresponding base pairs are colored in cyan).

RESULTS

A1 and X1 exhibit both shared and distinct chemical interference positions. To understand how the Bcd homeodomain recognizes A1 and X1, we carried out a methylation interference analysis. This assay identifies specific guanines (and, to a lesser extent, adenines) that prevent the protein from binding when methylated (see Materials and Methods). The results shown in Fig. 1a and b reveal both similar and different positions in A1 and X1 that inhibit Bcd homeodomain binding when methylated (see Fig. 4 for a summary of all our methylation interference data). First, methylation of guanines at the positions 5 and 6 on the antisense strand of A1 (3′ ATTA<u>GG</u> 5′) interfered strongly with Bcd homeodomain binding (Fig. 1a), consistent with their proposed role in Bcd recognition (32, 33). Similar to A1, methylation of the fifth-position guanine on the antisense strand of X1 (3′ ATTC<u>G</u>A 5′) inhibited Bcd homeodomain binding (Fig. 1b). Second, methylation of several adenines within the recognition sequences of both A1 and X1 interfered with Bcd homeodomain binding similarly. Third, the unique fourth-position guanine on the sense strand of X1 (TAA<u>G</u>CT) interfered with Bcd binding when methylated, suggesting the importance of this position in Bcd homeodomain binding (also see below). Our experiments also show that while methylation interference is restricted largely to positions within the A1 recognition sequence, X1 had an extended interference pattern beyond its recognition sequence. This property is associated with the recognition site of X1, rather than the flanking sequences, because a hybrid probe (Hybrid-A1)

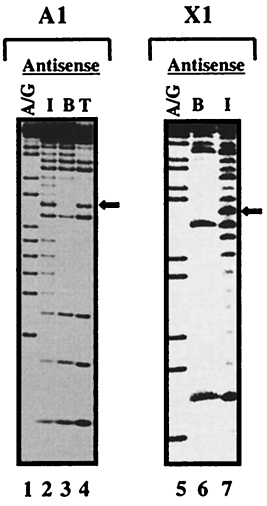


FIG. 2. The Bcd homeodomain makes a conserved thymine contact on A1 and X1. DNA samples with one strand treated with $KMnO_4$ (which modifies thymines) were used in binding assays to separate bound (B) and interfered (I; unbound) fractions. Thymine ladders (T) represent the positions of the modified thymines, used here as a reference. A band missing (marked by solid arrows in the bound fraction) indicates that modification of this thymine interferes with Bcd homeodomain binding. See the legend to Fig. 1 for further details.

containing the A1 recognition site and X1 flanking sequences did not show such an extended interference pattern (Fig. 1c; also see below).

Previous structural studies of other homeodomain proteins show that the amino-terminal arm makes specific contacts with thymines in the minor groove (36, 74). To determine how the Bcd homeodomain makes contacts with thymines, we conducted $KMnO_4$ interference experiments for the antisense strands of both A1 and X1. $KMnO_4$ interference is analogous to methylation interference but, unlike methylation interference, determines specifically how modifications of thymines interfere with protein-DNA interactions (64). The results shown in Fig. 2 demonstrate that modification of the second-position thymine in both A1 (3′ ATTAGG 5′) and X1 (3′ ATTCGA 5′) interfered with Bcd binding similarly. In contrast, modification of the third-position thymine did not inhibit Bcd homeodomain binding for either site. These results suggest that the Bcd homeodomain makes a conserved contact in the minor groove of the core sequences of both A1 and X1.

The Bcd homeodomain protects different guanines in A1 and X1. To further determine how the Bcd homeodomain interacts with A1 and X1, we conducted a methylation protection analysis. In this analysis, guanine residues that are contacted by (or in close proximity to) the Bcd homeodomain are protected specifically. Our results (Fig. 3) demonstrate that two guanines at the fifth and sixth positions on the antisense strand of A1 (3′ ATTAGG 5′) were strongly protected by the Bcd homeodomain (Fig. 3a). No protection was observed on the sense strand. In contrast, X1 showed a different protection pattern, with three guanines protected by the Bcd homeodomain (Fig. 3b). Two of these guanines were strongly protected and are located within the recognition sequence of X1: the fourth-position guanine on the sense strand (TAAGCT) and the fifth-position guanine on the antisense strand (3′ ATT CGA 5′). A third, partially protected guanine is located on the antisense strand immediately upstream (−1 position) of the recognition sequence of X1. When an antisense guanine was artificially placed upstream of A1 in the hybrid probe, it was not protected by the Bcd homeodomain (Fig. 3c). Except for this −1 position guanine in X1, no additional protection was observed outside the recognition sequences of either A1 or X1. This suggests that the methylation interference observed outside the recognition sequence of X1 (Fig. 1b) results from steric interference due to the bulky methyl groups rather than from loss of base-specific contacts, a conclusion further supported by our purine-specific missing-contact analysis (data not shown). A summary of all the methylation interference and protection data is shown in Fig. 4.

Arginine at position 54 of the Bcd homeodomain plays an important role in DNA recognition. Our experiments described thus far suggest that the Bcd homeodomain makes

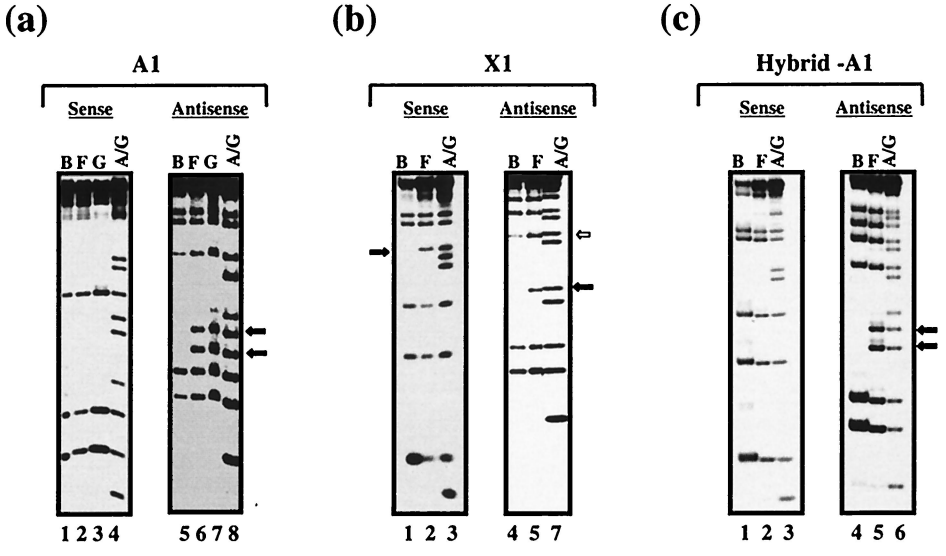


FIG. 3. The Bcd homeodomain makes different guanine contacts on A1 and X1. A methylation protection analysis was performed on both strands of DNA probes containing A1 (a), X1 (b), or Hybrid-A1 (c). F and B represent the methylation profiles of DNA isolated from free (unbound) and bound fractions, respectively. In this assay, a band missing in the bound fraction indicates that the Bcd homeodomain protects this position from being methylated by DMS. See the legend to Fig. 1 for further details.

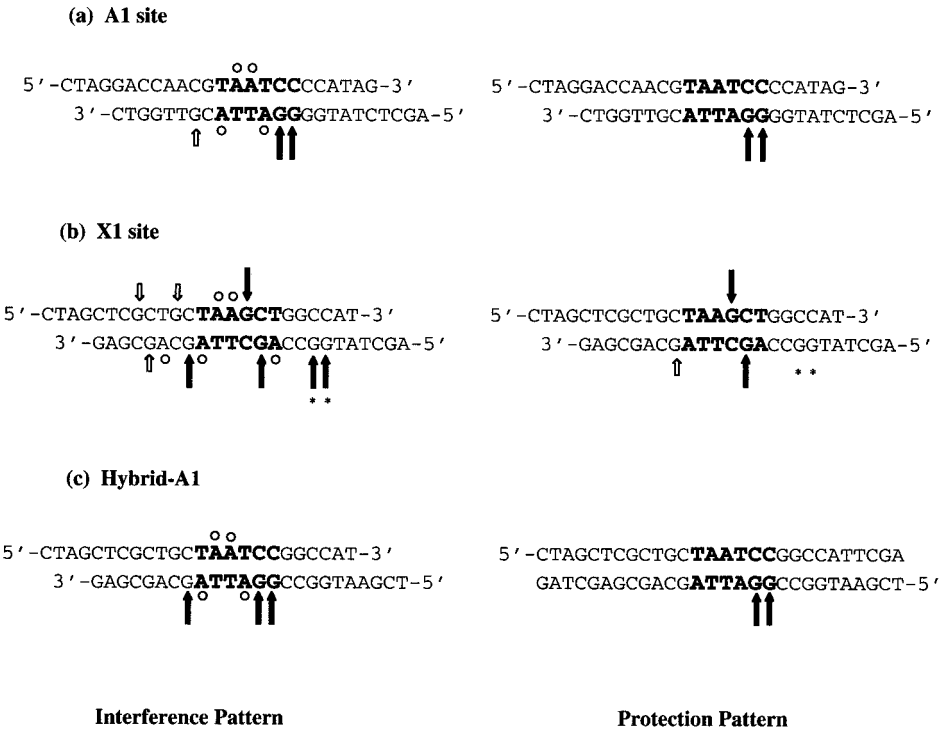


FIG. 4. Summary of interference and protection patterns on A1 and X1. Shown are the DMS interference and protection patterns on A1, X1, and Hybrid-A1. The solid and open arrows show guanines that are strongly and weakly interfered or protected, respectively. Open circles show adenine interference. Asterisks indicate interference at guanines on X1 that is not observed in Hybrid-A1. These two guanines are not protected by the Bcd homeodomain either (see the text for further details). The recognition sequences are in bold.

both shared and distinct contacts with the consensus site A1 and the nonconsensus site X1. In particular, the fourth-position guanine (TAAGCT) unique to X1 is specifically protected by the Bcd homeodomain, suggesting that this residue may be part of the specificity determinant of X1. Since guanines are highly electronegative in the major groove and ideally structured to interact with arginines (43, 46, 60), we focused our attention on the arginine residue at position 54 (Arg 54). Among all the natural K50 homeodomains in a recent database (3), only the Bcd homeodomain contains an arginine at this position (Fig. 5). It has been proposed that residues 50 and 54 of homeodomains may have coevolved to determine the DNA-binding specificity (7, 9, 50). In addition, previous structural studies have shown that residue 54 in other homeodomains can make either base-specific or phosphate contacts with DNA (5, 21, 29, 30, 35, 68, 74). In particular, structural studies of the yeast Matα2 homeodomain show that Arg 54 interacts specifically with a fourth-position guanine on the antisense strand in the major groove (38, 74). These observations, as well as our molecular modeling studies (see Discussion), suggest that Arg 54 of the Bcd homeodomain may participate in recognizing X1 by specifically contacting the fourth-position guanine.

To determine the role of Arg 54 of the Bcd homeodomain in DNA recognition, we changed this position to alanine (R54A) and analyzed its DNA-binding activity in gel mobility shift assays. Our experiments (Fig. 6a) show that the Bcd(R54A) homeodomain exhibits a decreased but detectable binding to A1 (compare lanes 2 and 3). In contrast, this derivative has a much more severe defect in recognizing X1 under the same condition (lane 7). We also tested another nonconsensus site, TGATCC (X3s), also from the *hb* enhancer element (76). Our experiments show that, like A1 but unlike X1, X3s is recognized by the Bcd(R54A) homeodomain with a modestly decreased efficiency (compare lanes 10 and 11). Our measurements of dissociation constants (Fig. 6b) further confirmed that the R54A mutation of the Bcd homeodomain preferentially affects its binding to X1.

Our transient-transfection experiments with *Drosophila* Schneider S2 cells demonstrate that the Bcd(R54A) mutant full-length protein has a reduced ability to activate transcription from the natural *hb* enhancer compared with the wild-type protein (Fig. 7). In addition, when the three nonconsensus sites in the *hb* enhancer element, two of which have a TAAG core, were converted to consensus sites (Fig. 7a), the activity of Bcd(R54A) was restored (Fig. 7b). In our experiments, the Bcd proteins accumulated to similar levels in *Drosophila* cells, as determined by Western blot analysis (Fig. 7c). Together, these experiments further demonstrate the importance of Arg 54 in Bcd function, particularly in recognizing nonconsensus sites in the natural *hb* enhancer element.

Arg 54 of the Bcd homeodomain makes base-specific contacts. To determine whether Arg 54 of the Bcd homeodomain makes base-specific or phosphate contacts, we generated an arginine-to-lysine mutation (R54K). We reasoned that since both lysine and arginine are positively charged, this mutation might not disrupt DNA binding if the only role of Arg 54 is to make phosphate contacts. However, our experiments show that Bcd(R54K) has a greatly reduced ability to recognize all three DNA sites tested (Fig. 6a, lanes 4, 8, and 12). As a control, we analyzed the result of a lysine-to-arginine mutation at the position 50 (K50R), a position that confers Bcd specificity presumably by making base-specific contacts (32, 34). As expected, the Bcd(K50R) homeodomain failed to bind DNA (data not shown). Together, these results are consistent with the idea, but do not demonstrate directly, that Arg 54 of the

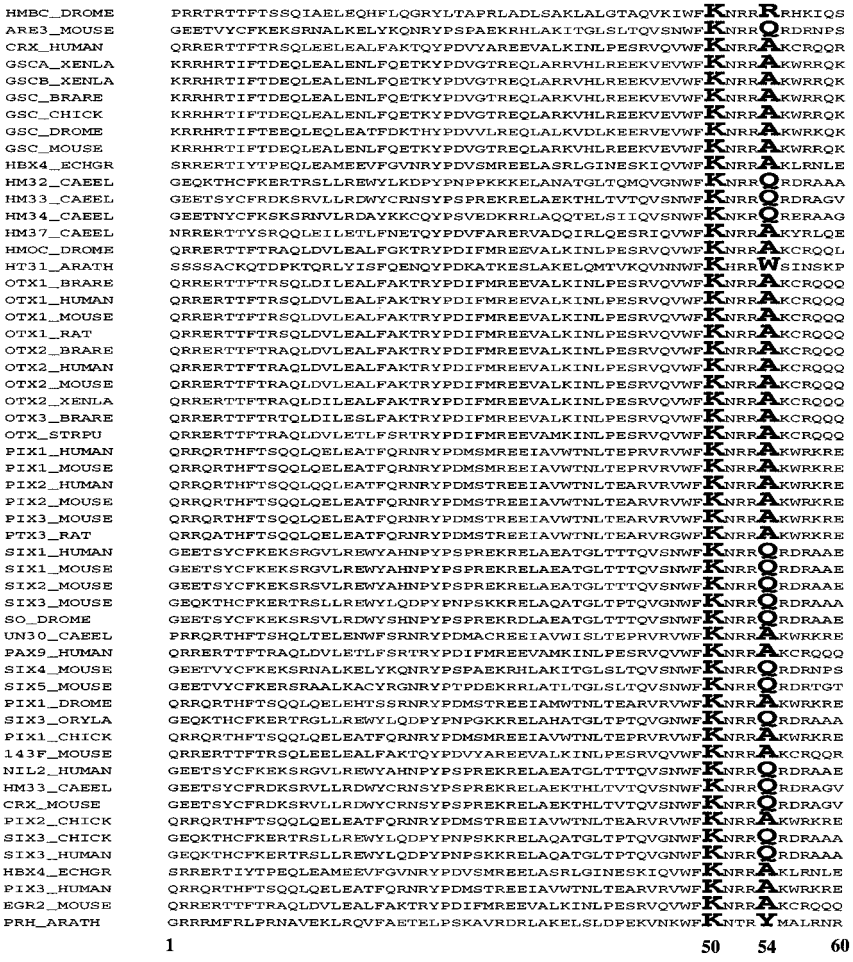


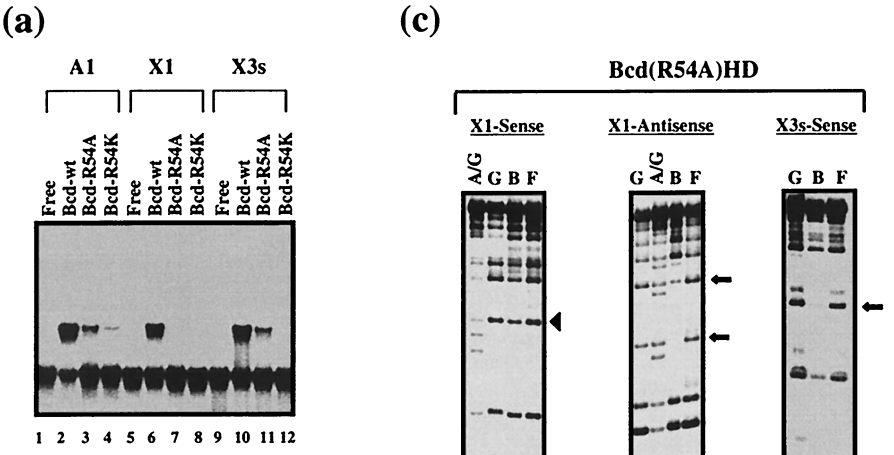
FIG. 5. Alignment of natural K50 homeodomains. Shown are sequences of known natural K50 homeodomains, highlighting amino acids 50 and 54. While alanine (A) and glutamine (Q) are the most frequently found residues at position 54, only the Bcd homeodomain contains an arginine at this position (Arg 54). The sequence data have been extracted from the Homeodomain Resource Database maintained by the Division of Intramural Research, Genome Technology Branch, National Human Genome Research Institute, National Institutes of Health (3). In this figure, the sequence of the Bcd homeodomain (HMBC_DROME) is listed in the first line.

Bcd homeodomain makes base-specific contacts as opposed to phosphate contacts only.

To directly test whether Arg 54 of the Bcd homeodomain makes a specific contact with the fourth-position guanine in X1, we conducted a chemical footprint assay using the Bcd (R54A) mutant homeodomain. Although this protein is defective in interacting with X1 (Fig. 7a), high concentrations of the DNA probe were able to drive the binding reaction to generate sufficient amount of specific complex for our footprint assay. To ensure specific DNA binding, 1 μg of poly(dI-dC) was included as a nonspecific competitor in our preparative gel shift reaction mixture (see Materials and Methods for details). The results of our experiments (Fig. 6c, lane 3) indicate that the homeodomain of Bcd(R54A), unlike the wild-type homeodomain (Fig. 3b, lane 1), fails to protect completely the fourth-position guanine of X1. The loss of protection at this position is a specific effect because the Bcd(R54A) homeodomain, like the wild-type protein, can protect both the fifth-position guanine (3′ ATTCGA 5′) on the antisense strand of X1 and, partially, its −1 position guanine (Fig. 6c, lane 7). In addition, the mutant homeodomain can still protect the second-position guanine on the sense strand of another nonconsensus site, X3s (TGATCC) (Fig. 6c, lane 10). Together, these experiments further support the idea that Arg 54 interacts specifically with the fourth-position guanine in X1, contributing to the ability of the protein to recognize this nonconsensus site.

Contributions of Arg 54 to DNA binding by other homeodomains. To determine whether Arg 54 may also facilitate DNA recognition in other homeodomains, we analyzed the DNA-binding properties of homeodomains isolated from three different proteins. Bozozok (Boz, also called Dharma) is a zebrafish homeodomain protein essential for inducing the gastrula organizer and dorsoanterior structures in the embryo (18, 75). The human pituitary homeobox 2 (Pitx2) protein has been implicated in Reiger syndrome (57), and its mouse counterpart has been shown to participate in the determination of embryonic left-right asymmetry (40, 52, 55). Orthodenticle (Otd) is a *Drosophila* homeodomain protein involved in head formation during early embryonic development (19, 20). All three homeodomains, like the Bcd homeodomain, contain a lysine residue at position 50 but, unlike the Bcd homeodomain, lack an arginine residue position 54 (Fig. 5 and 8d). In addition, we tested an altered-specificity mutant homeodomain with a glutamine-to-lysine change at position 50, Ftz(Q50K), which can bind to a consensus Bcd site efficiently (51, 78).

Figure 8 shows the results of our gel shift experiments using these homeodomains, either with or without an artificial Arg 54. The experiments in Fig. 8a and c were carried out at a DNA



(b)

Site	K_D (nM) Wt	R54A
A1	0.36 ± 0.05	2.86 ± 0.45
X1	0.45 ± 0.06	22.8 ± 1.0

FIG. 6. DNA recognition by the Bcd homeodomain and its derivatives. (a) Gel mobility shift assays using Bcd homeodomain and its mutants on three different DNA sites: A1 (TAATCC), X1 (TAAGCT), and X3s (TGATCC). See Materials and Methods for further details. The DNA probe concentration used in all these experiments was 10^{-9} M, and the estimated active-protein concentration was $\sim 10^{-10}$ M. (b) Dissociation constant (K_D) measurements further confirm that the R54A mutation of the Bcd homeodomain preferentially affects its binding to X1. A Scatchard analysis was performed to determine the dissociation constants of the interactions between the Bcd homeodomain (either wild type or R54A mutant) and two DNA sequences (A1 and X1). For this assay, a quantitative gel shift analysis was performed at increasing concentrations of the DNA probes. The bound and free fractions of the probes were determined with a PhosphorImager and analyzed using Microsoft Excel (linear regression) to determine the K_D value ($-1/K_D$ = slope of the plot of Bound/Free against Bound). The listed values represent three independent assays (mean ± standard deviation). For K_D measurements, the wild-type (Wt) homeodomain was prepared in the same manner as the R54A mutant protein (see Materials and Methods). (c) Methylation protection by the Bcd(R54A) homeodomain on the sense strand of X1 (left), the antisense strand of X1 (middle), or the sense strand of another nonconsensus site X3s (right). A/G and G represent sequencing ladders, while B and F represent bound and free DNA, respectively. The sequence of the X3s sense strand probe is 5′-CTAGATCTGCTCTGATCCAGAATG-3′. The fourth-position guanine of the X1 sense strand that is protected by the wild-type Bcd homeodomain (Fig. 3b) is marked with an arrowhead. The guanines of X1 antisense and X3s sense strands that remain protected by the Bcd(R54A) homeodomain are marked with arrows.

probe concentration of 10^{-9} M; since the Otd and Boz homeodomains bind DNA poorly (Fig. 8a), we also carried out additional experiments (Fig. 8b) for these two homeodomains at a higher probe concentration (10^{-8} M). The active-protein concentration for all homeodomains was kept constant ($\sim 10^{-9}$ M). Our gel shift experiments show that, in general, Arg 54 plays a positive role in DNA recognition. For example, while the wild-type Pitx2 homeodomain failed to recognize X3s in our assay (Fig. 8a, lane 20), the Pitx2(A54R) homeodomain could bind to this site efficiently (lane 23). In addition, when Arg 54 was introduced into the Otd homeodomain, it significantly increased the DNA binding to all three sites: the wild-type Otd homeodomain bound poorly to A1 and undetectably to X1 and X3s in our assay (Fig. 8b, lanes 2, 7, and 12), whereas the Otd(A54R) homeodomain exhibited both improved binding to A1 (lane 4) and an ability to recognize both X1 and X3s, albeit poorly (lanes 9 and 14). Our results also show that, unlike the wild-type Boz homeodomain (lanes 3 and 13), Boz (A54R) appeared to form dissociable complexes on both A1 and X3s, as judged by the smear (Fig. 8b, lanes 5 and 15); taking such smears into account, there might be more binding by the Boz(A54R) homeodomain than by the wild-type Boz homeodomain. Arg 54 did not have any detectable effect on the DNA-binding profile of the Ftz(Q50K) homeodomain (Fig. 8c). Together, our experiments demonstrate that the role that Arg 54 plays in DNA recognition depends on specific homeodomains and recognition sites (also see Discussion).

DISCUSSION

The experiments described in this report probe the molecular interactions between the Bcd homeodomain and its recognition sequences, focusing on two different DNA sites, A1 and X1, both from the *hb* enhancer element. Our chemical-footprint studies suggest that the Bcd homeodomain makes both shared and distinct contacts with the consensus site A1 and the nonconsensus site X1. For example, the fourth-position guanine on the sense strand unique to X1 (TAAGCT) is specifically protected by the Bcd homeodomain (Fig. 3). In addition, the guanine at the −1 position on the antisense strand of X1 is also protected by the Bcd homeodomain, though less efficiently (Fig. 3b); a guanine artificially placed upstream of A1 is not protected (Fig. 3c). Our experiments also show that the protein-DNA complex formed on X1 is more susceptible to external challenges than is that formed on A1, including steric interference outside the recognition sequences (Fig. 1 and 3), ionic strength, and nonspecific DNA (not shown). Despite these and other differences, the Bcd homeodomain makes several conserved contacts with both A1 and X1, such as the fifth-position guanine in the major groove (Fig. 1 and 3) and the second-position thymine in the minor groove (Fig. 2). Taken together, our results suggest that the Bcd homeodomain docks on different DNA sites with a similar overall structure but distinct sets of protein-DNA contacts.

Our analysis of Arg 54 of the Bcd homeodomain provides

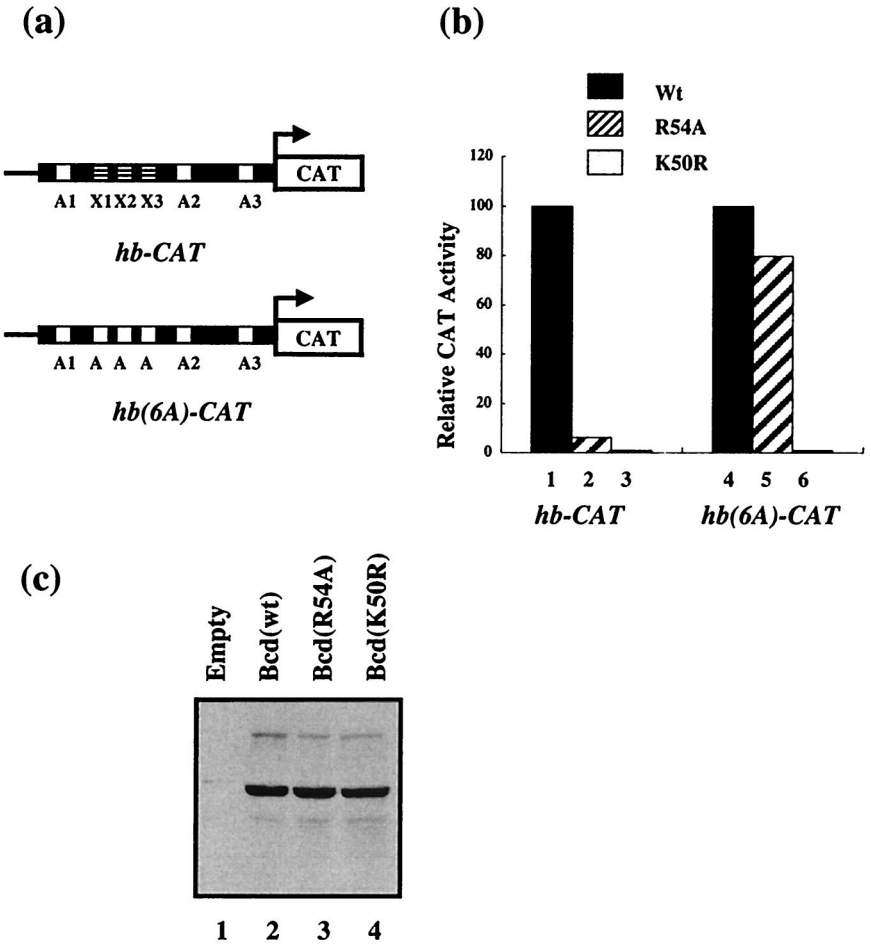


FIG. 7. Mutation of position 54 of the Bcd homeodomain reduces transcription activation in transient-transfection assays. (a) CAT reporter plasmids used in our transient-transfection assays in *Drosophila* Schneider S2 cells. *hb-CAT* contains a wild-type 250-bp enhancer element from *hb*, whereas *hb(6A)-CAT* contains a modified enhancer element with nonconsensus sites converted to consensus sites. (b) Relative CAT activities from four independent experiments, with wild-type (Wt) Bcd activity set at 100% for each reporter (lanes 1 and 4). Mutant Bcd(R54A) is defective in transcription activation from *hb-CAT*, while a modified *hb* enhancer containing all consensus A sites [*hb(6A)-CAT*] restores activation (compare lanes 2 and 5). Bcd(K50R), which contains a lysine-to-arginine change at position 50, does not show any activity from either reporters (lanes 3 and 6). The arbitrary activities are 100, 6.2, <1, 44.5, 35.4, and <1 for lanes 1 to 6, respectively. (c) Western blot analysis using HA-tagged Bcd derivatives. All the Bcd derivatives (lane 2, 3, and 4) accumulate at similar levels in the cell. Lane 1 represents cell lysate transfected with the vector expressing no activator.

important insights into the molecular interactions between the Bcd homeodomain and its DNA sites. Our results suggest that Arg 54 makes a specific contact with the unique fourth-position guanine of X1 (TAAGCT), thus enabling the protein to recognize this nonconsensus site efficiently. These results support an adaptive-recognition model, where the deviating fourth-position guanine of X1 provides an opportunity for the Bcd homeodomain to acquire a new interaction. Our molecular modeling of the complexes of Bcd homeodomain on A1 and X1 suggests that the side chain of Arg 54 makes adjustments to acquire such a new interaction (Fig. 9). Our model suggests that when complexed with the consensus site A1 (Fig. 9a), Arg 54 of the Bcd homeodomain makes a hydrogen bond with the third-position adenine (TAATCC), whereas a translational movement of its side chain enables it to make a unique hydrogen bond with the fourth-position guanine in X1 (TAAGCT) (Fig. 9b). Taken together, our analyses suggest that Bcd uses reprogrammable recognition codes for different DNA sites. Although our present work emphasizes the adaptive role of Arg 54 of the Bcd homeodomain in DNA recognition, other residues are likely to undergo similar adjustments to maximize the contacts with different DNA sites.

Structural studies of other homeodomains also suggest a dynamic nature of their interaction with DNA. For example, two different Even-skipped (Eve) homeodomains (complexed with DNA as a dimer) project their Gln 50 side chains differently to make distinct sets of base interactions (35). In addition, Lys 50 of an altered-specificity Engrailed homeodomain, En(Q50K), is in a dynamic equilibrium between two different conformations, making different sets of contacts with the recognition sequence TAATCC (66). In one conformation, the Lys 50 side chain interacts with the fifth- and sixth-position guanines on the antisense strand (3′ ATTAGG 5′). In the other conformation, it interacts with the fifth-position guanine on the antisense strand (3′ ATTAGG 5′) and the fourth-position thymine on the sense strand (TAATCC). This latter conformation is particularly relevant to the present study. Our methylation footprint experiments show that the fourth-position guanine of X1 remains partially protected by the Bcd (R54A) mutant homeodomain (Fig. 6c, lane 3), a finding con-

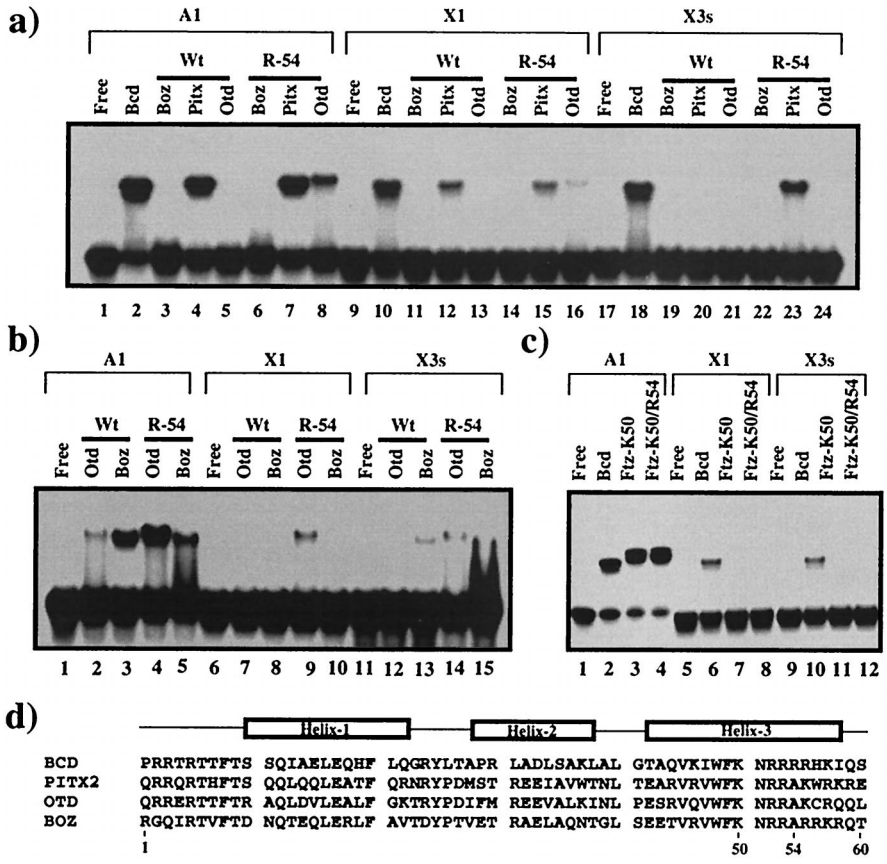


FIG. 8. Effect of R54 on the DNA-binding activity of other homeodomains. (a to c) Gel mobility shift assay results obtained using different homeodomains and different DNA sites. The DNA probe concentration was 10^{-9} M for the experiments in panels a and c and 10^{-8} M for those in panel b. The active-protein concentration was normalized with respect to the Bcd homeodomain and kept at 10^{-9} M. See the text for further details. (d) Sequence alignment of the four natural K50-class homeodomains used in this study. The three helices of the homeodomains are marked. Wt, wild type.

sistent with our molecular modeling in which Lys 50 also interacts with this guanine (C.-S. Tung, unpublished data). We would like to note that although our present results support the notion that Arg 54 makes base-specific contacts (Fig. 9), we cannot formally exclude the possibility that it also makes phosphate contacts.

The experiments described in this report reveal two further findings. First, depending on the specific pairs of homeodomains and DNA sequences, the role of Arg 54 in DNA recognition can vary dramatically. It ranges from increasing the binding to a specific sequence (e.g., Otd binding to A1 [Fig. 8b, compare lanes 2 and 4]) to conferring a previously undetectable binding specificity (e.g., Pitx2 binding to X3s [Fig. 8a, compare lanes 20 and 23] and Otd binding to X1 and X3s [Fig. 8b, compare lanes 7 and 9 and lanes 12 and 14]). In some cases, Arg 54 appears to have no detectable effect on DNA binding [e.g., Boz binding to X1 (Fig. 8b, compare lanes 8 and 10), Pitx2 binding to A1 and X1 (Fig. 8a, compare lanes 4 and 7 and lanes 12 and 15), and Ftz(Q50K) binding to all three sites (Fig. 8c)]. We propose that an overall structural context and specific amino acids of a homeodomain together determine whether the protein can recognize a given nonconsensus site. The role of Arg 54 in the recognition of different DNA sites by different homeodomains represents a magnifying indicator for subtle structural or docking differences of individual homeodomains. In this context, it is interesting that the mobility of protein-DNA complexes containing Otd or Ftz(Q50K) homeodomains appears slightly different from that of the complex containing the Bcd homeodomain (Fig. 8a, compare lanes 2 and 8; Fig. 8c, compare lanes 2 and 3). It is also interesting that the configuration of Arg 54 depicted in Fig. 9 corresponds to one of four families of possible conformations (Tung, unpublished), suggesting that this residue can play a versatile role in DNA recognition depending on specific homeodomain contexts and DNA sites. (Interestingly, Arg 54 has also been proposed recently to play a critical role in interacting with RNA [see below]. In addition, position 54 of homeodomains has been suggested to control the structural stability of the recognition helix [61].) The only available structure of a K50 class homeodomain is the altered-specificity mutant En(Q50K) homeodomain complexed on a consensus Bcd site (66), but residue 54 is not arginine.

Second, Arg 54 is not required for recognition of X1 in some homeodomains (e.g., Pitx2 [Fig. 8a, lane 12]), suggesting that an identical DNA sequence may be recognized by different homeodomains with distinct recognition codes. We do not know exactly how the Pitx2 homeodomain recognizes X1 and, in particular, whether the fourth-position guanine (TAAGCT) is also protected. Regardless, our experiments suggest that the detailed molecular interactions between homeodomains and DNA sequences require an individualized analysis in addition to a generalized approach (34, 62, 66).

The ability of the Bcd homeodomain to bind to nonconsensus sites in natural enhancer elements is critical for proper target selection (see the introduction). Our present study provides the first detailed molecular analysis of the interaction

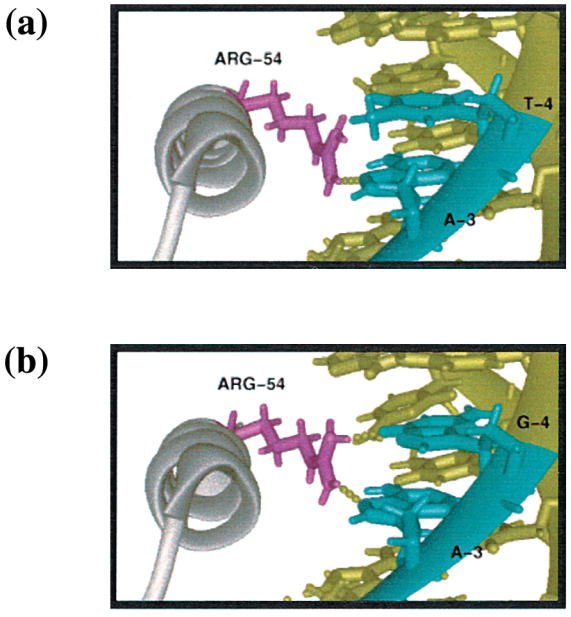


FIG. 9. Adaptive interaction of the Bcd homeodomain on consensus (A1) and nonconsensus (X1) sites mediated by Arg 54. Shown are homology-modeled structures of the Bcd homeodomain complexed with site A1 (a) or X1 (b). Only the third helix (pointing toward the page) and the side chain of Arg 54 are shown in this figure for clarity. For A1, N-7 of the third-base adenine (TAATCC; the corresponding base pair is colored in cyan) is in the vicinity of the NH_2 group emanating from Arg 54 (colored in magenta) and may form a single hydrogen bond (shown in yellow). For the fourth-position thymine, the NH_2 group of Arg 54 is too far away to form a hydrogen bond (>3 Å). For X1, the side chain of Arg 54 swings upward with a vertical translation, allowing the NH_2 groups to move about 1.4 Å in order to form a second hydrogen bond with the fourth-position guanine (TAAGCC). Thus, Arg 54 forms bidentate hydrogen bonds on X1 in this model: one with N-7 of the third-base adenine and the other with O-6 of the fourth base guanine (TAAGCC; the corresponding base pairs are colored in cyan). In this model structure, as seen with other homeodomains (72), Asn 51 (not shown in the figure) is well positioned to contact the third-base adenine, and therefore we propose that the N-7 position of this adenine engages in shared hydrogen bonds: one from Asn 51 and the other from Arg 54.

between the Bcd homeodomain and one of the nonconsensus sites. In addition to X1, which contains a TAAG core, there are at least two other types of naturally occurring nonconsensus sites, with TGAT and AAAT core sequences. We suggest that the deviating nucleobases in the nonconsensus sites are not coincidental but, rather, play specific roles in interactions with the Bcd homeodomain. As argued in this report, the fourth-position guanine of TAAG is contacted by Arg 54 of the Bcd homeodomain. In addition, while X3s (TGATCC) is a naturally occurring sequence recognized by Bcd, TCATCC cannot be recognized by Bcd (date not shown). We propose that the ability of Bcd to recognize the nonconsensus sites helps define its unique biological specificity. In particular, our experiments show that while Bcd can recognize both X1 and X3s, none of the homeodomains tested in this study can recognize both (Boz can bind X3s but not X1, and Pitx2 can bind X1 but not X3s). Our previous experiments show that Bcd can bind DNA cooperatively (41, 77). Although a protein-DNA complex formed on the consensus site TAATCC may be energetically optimal, cooperativity can further strengthen the recognition of Bcd to multiple sites, particularly nonconsensus sites. It will be interesting to determine whether the special docking modes on nonconsensus sites play any role in facilitating cooperativity.

After this manuscript was completed, we became aware of a recent paper addressing the function of Arg 54 of Bcd (48). Niessing et al. suggest that while Arg 54 of Bcd plays a critical role in RNA recognition and translation inhibition, it is not required for DNA recognition. Unlike our study, only a consensus Bcd site was analyzed in that study and, moreover, the data do not permit a quantitative comparison. Interestingly, *hb* is still expressed in R54A embryos, which can develop at a low temperature (18°C) but not at the normal temperature. Our transcriptional activation experiments (Fig. 7a) were carried out at low Bcd concentrations, at which the defect of R54A was readily detectable. At higher concentrations, R54A activated transcription almost as efficiently as the wild-type protein did (data not shown), reflecting nonconsensus site occupancy presumably facilitated by cooperative DNA binding (6, 41, 77). Our experiments suggest that the defect of R54A in the embryos would be restricted to low concentrations of Bcd with subtle differences at the posterior border of the *hb* expression domain. It is therefore not surprising that R54A embryos can develop conditionally—previous studies have demonstrated that embryos with different doses of maternal *bcd* can develop into normal, healthy adults despite shifted expression domains of *hb* and other segmentation genes (8, 20, 23, 24, 47, 59). It is particularly interesting that Arg 54 of Bcd is required for recognition of both an RNA sequence and a nonconsensus DNA site. It remains to be determined whether these two functions have common requirements for other structural features of the Bcd homeodomain.

ACKNOWLEDGMENTS

V. Dave and C. Zhao contributed equally to this work.

We thank E. Semina, J. Murray, W. Driever, R. Finkelstein, and J. Manley for plasmids and S. Burley for discussions.

This work has been supported in part by NIH grants (R01 GM52467 and P30 ES06096). C.-S. Tung is supported by the Los Alamos National Laboratory LDRD program.

REFERENCES

1. **Akman, S. A., J. H. Doroshow, and M. Dizdaroglu.** 1990. Base modifications in plasmid DNA caused by potassium permanganate. Arch. Biochem. Biophys. **282:**202–205.
2. **Ausubel, F., R. Brent, R. Kingston, D. Moore, J. Seidman, J. Smith, and K. Struhl.** 1994. Current protocols in molecular biology. John Wiley & Sons, Inc., New York, N.Y.
3. **Banerjee-Basu, S., E. S. Ferlanti, J. F. Ryan, and A. D. Baxevanis.** 1999. The homeodomain resource: sequences, structures and genomic information. Nucleic Acids Res. **27:**336–367.
4. **Berleth, T., M. Burri, G. Thoma, D. Bopp, S. Richstein, G. Frigerio, M. Noll, and C. Nüsslein-Volhard.** 1988. The role of localization of bicoid RNA in organizing the anterior pattern of the *Drosophila* embryo. EMBO J. **7:**1749–1756.
5. **Billeter, M., Y. Q. Qian, G. Otting, M. Muller, W. Gehring, and K. Wuthrich.** 1993. Determination of the nuclear magnetic resonance solution structure of an Antennapedia homeodomain-DNA complex. J. Mol. Biol. **234:**1084–1093.
6. **Burz, D. S., R. Pivera-Pomar, H. Jackle, and S. D. Hanes.** 1998. Cooperative DNA-binding by Bicoid provides a mechanism for threshold-dependent gene activation in the *Drosophila* embryo. EMBO J. **17:**5998–6009.
7. **Connolly, J. P., J. G. Augustine, and C. Francklyn.** 1999. Mutational analysis of the engrailed homeodomain recognition helix by phage display. Nucleic Acids Res. **27:**1182–1189.
8. **Dalton, D., R. Chadwick, and W. McGinnis.** 1989. Expression and embryonic function of empty spiracles: a *Drosophila* homeo box gene with two patterning functions on the anterior posterior axis of the embryo. Genes Dev. **3:**1940–1956.
9. **Damante, G., L. Pellizzari, G. Esposito, F. Fogolari, P. Viglino, D. Fabbro, G. Tell, S. Formisano, and R. Di Lauro.** 1996. A molecular code dictates sequence-specific DNA recognition by homeodomains. EMBO J. **15:**4992–5000.
10. **Driever, W.** 1992. The Bicoid morphogen: concentration dependent transcriptional activation of zygotic target genes during early *Drosophila* development, p. 1221–1250. *In* S. L. McKnight and K. Yamamoto (ed.), Transcriptional regulation. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.
11. **Driever, W., J. Ma, C. Nusslein-Volhard, and M. Ptashne.** 1989. Rescue of bicoid mutant *Drosophila* embryos by Bicoid fusion proteins containing heterologous activating sequences. Nature **342:**149–154.

12. **Driever, W., and C. Nüsslein-Volhard.** 1988. The bicoid protein determines position in the *Drosophila* embryo in a concentration dependent manner. Cell **54:**95–104.
13. **Driever, W., and C. Nüsslein-Volhard.** 1989. Bicoid protein is a positive regulator of hunchback transcription in the early *Drosophila* embryo. Nature **337:**138–143.
14. **Driever, W., and C. Nüsslein-Volhard.** 1988. A gradient of bicoid protein in *Drosophila* embryos. Cell **54:**83–93.
15. **Driever, W., G. Thoma, and C. Nüsslein-Volhard.** 1989. Determination of spatial domains of zygotic gene expression in the *Drosophila* embryo by the affinity of binding site for the bicoid morphogen. Nature **340:**363–367.
16. **Elrod-Erickson, M., T. E. Benson, and C. O. Pabo.** 1998. High-resolution structures of variant Zif268-DNA complexes: implications for understanding zinc finger-DNA recognition. Structure **6:**451–464.
17. **Elrod-Erickson, M., and C. O. Pabo.** 1999. Binding studies with mutants of Zif268. Contribution of individual side chains to binding affinity and specificity in the Zif268 zinc finger-DNA complex. J. Biol. Chem. **274:**19281–19285.
18. **Fekany, K., Y. Yamanaka, T. Leung, H. I. Sirotkin, J. Topczewski, M. A. Gates, M. Hibi, A. Renucci, D. Stemple, A. Radbill, A. F. Schier, W. Driever, T. Hirano, W. S. Talbot, and L. Solnica-Krezel.** 1999. The zebrafish bozozok locus encodes Dharma, a homeodomain protein essential for induction of gastrula organizer and dorsoanterior embryonic structures. Development **126:**1427–1438.
19. **Finkelstein, R., and N. Perrimon.** 1990. The orthodenticle gene is regulated by bicoid and torso and specifies *Drosophila* head development. Nature **346:** 485–488.
20. **Finkelstein, R., D. Smouse, T. M. Capaci, A. C. Spradling, and N. Perrimon.** 1990. The orthodenticle gene encodes a novel homeo domain protein involved in the development of the *Drosophila* nervous system and ocellar visual structures. Genes Dev. **4:**1516–1527.
21. **Fraenkel, E., and C. O. Pabo.** 1998. Comparison of X-ray and NMR structures for the Antennapedia homeodomain-DNA complex. Nat. Struct. Biol. **5:**692–697.
22. **Fraenkel, E., M. A. Rould, K. A. Chambers, and C. O. Pabo.** 1998. Engrailed homeodomain-DNA complex at 2.2 A resolution: a detailed view of the interface and comparison with other engrailed structures. J. Mol. Biol. **284:** 351–361.
23. **Frohnhöfer, H. G., and C. Nüsslein-Volhard.** 1987. Maternal genes required for the anterior localization of bicoid activity in the embryo of *Drosophila*. Genes Dev. **1:**880–890.
24. **Frohnhöfer, H. G., and C. Nüsslein-Volhard.** 1986. Organization of anterior pattern in the *Drosophila* embryo by the maternal gene bicoid. Nature **324:** 120–125.
25. **Gao, Q., and R. Finkelstein.** 1998. Targeting gene expression to the head: the *Drosophila* orthodenticle gene is a direct target of the Bicoid morphogen. Development **125:**4185–4193.
26. **Gehring, W. J., M. Affolter, and T. Burglin.** 1994. Homeodomain proteins. Annu. Rev. Biochem. **63:**487–526.
27. **Gehring, W. J., Y. Q. Qian, M. Billeter, K. Furukubo-Tokunaga, A. F. Schier, D. Resendez-Perez, M. Affolter, G. Otting, and K. Wuthrich.** 1994. Homeodomain-DNA recognition. Cell **78:**211–223.
28. **Gralla, J. D.** 1993. Footprinting of nucleic acid-protein complexes. Academic Press, Inc., New York, N.Y.
29. **Gruschus, J. M., D. H. Tsao, L. H. Wang, M. Nirenberg, and J. A. Ferretti.** 1997. Interactions of the vnd/NK-2 homeodomain with DNA by nuclear magnetic resonance spectroscopy: basis of binding specificity. Biochemistry **36:**5372–5380.
30. **Gruschus, J. M., D. H. Tsao, L. H. Wang, M. Nirenberg, and J. A. Ferretti.** 1999. The three-dimensional structure of the vnd/NK-2 homeodomain-DNA complex by NMR spectroscopy. J. Mol. Biol. **289:**529–545.
31. **Han, K., M. Levine, and J. Manley.** 1989. Synergistic activation and repression of transcription by *Drosophila* homeobox proteins. Cell **56:**573–583.
32. **Hanes, S., and R. Brent.** 1991. A genetic model for interaction of the homeodomain recognition helix with DNA. Science **251:**426–430.
33. **Hanes, S., G. Riddihough, D. Ish-Horowicz, and R. Brent.** 1994. Specific DNA recognition and intersite spacing are critical for action of the Bicoid morphogen. Mol. Cell. Biol. **14:**3364–3375.
34. **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.
35. **Hirsch, J. A., and A. K. Aggarwal.** 1995. Structure of the even-skipped homeodomain complexed to AT-rich DNA: new perspectives on homeodomain specificity. EMBO J. **14:**6280–6291.
36. **Kissinger, C. R., L. Beishan, E. Martin-Blanco, T. Kornberg, and C. O. Pabo.** 1990. Crystal structure of an engrailed homeodomain-DNA complex at 2.8 Å resolution: a framework for understanding homeodomain-DNA interactions. Cell **63:**579–590.
37. **La Rosee, A., T. Hader, H. Taubert, R. Rivera-Pomar, and H. Jackle.** 1997. Mechanism and Bicoid-dependent control of hairy stripe 7 expression in the posterior region of the *Drosophila* embryo. EMBO J. **16:**4403–4411.
38. **Li, T., M. R. Stark, A. D. Johnson, and C. Wolberger.** 1995. Crystal structure of the Mata1/Matα2 homeodomain heterodimer bound to DNA. Science **270:**262–269.
39. **Lillie, J. W., and M. R. Green.** 1989. Transcription activation by the adenovirus E1a protein. Nature **338:**39–44.
40. **Lin, C. R., C. Kioussi, S. O'Connell, P. Briata, D. Szeto, F. Liu, J. C. Izpisua-Belmonte, and M. G. Rosenfeld.** 1999. Pitx2 regulates lung asymmetry, cardiac positioning and pituitary and tooth morphogenesis. Nature **401:** 279–282.
41. **Ma, X., D. Yuan, K. Diepold, T. Scarborough, and J. Ma.** 1996. The *Drosophila* morphogenetic protein Bicoid binds DNA cooperatively. Development **122:**1195–1206.
42. **Ma, X., D. Yuan, T. Scarborough, and J. Ma.** 1999. Contributions to gene activation by multiple functions of Bicoid. Biochem. J. **338:**447–455.
43. **Mandel-Gutfreund, Y., O. Schueler, and H. Margalit.** 1995. Comprehensive analysis of hydrogen bonds in regulatory protein-DNA complexes: in search of common principles. J. Mol. Biol. **253:**370–382.
44. **Maniatis, T., E. F. Fritsch, and J. Sambrook.** 1982. Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
45. **Maxam, A. M., and W. Gilbert.** 1980. Sequencing end-labeled DNA with base-specific chemical cleavages. Methods Enzymol. **65:**499–560.
46. **Nadassy, K., S. J. Wodak, and J. Janin.** 1999. Structural features of protein-nucleic acid recognition sites. Biochemistry **38:**1999–2017.
47. **Namba, R., T. M. Pazdera, R. L. Cerrone, and J. S. Minden.** 1997. *Drosophila* embryonic pattern repair: how embryos respond to bicoid dosage alteration. Development **124:**1393–1403.
48. **Niessing, D., W. Driever, F. Sprenger, H. Taubert, H. Jackle, and R. Rivera-Pomar.** 2000. Homeodomain position 54 specifies transcriptional versus translational control by Bicoid. Mol. Cell **5:**395–401.
49. **Patikoglou, G. A., J. L. Kim, L. Sun, S. H. Yang, T. Kodadek, and S. K. Burley.** 1999. TATA element recognition by the TATA box-binding protein has been conserved throughout evolution. Genes Dev. **13:**3217–3230.
50. **Pellizzari, L., G. Tell, D. Fabbro, C. Pucillo, and G. Damante.** 1997. Functional interference between contacting amino acids of homeodomains. FEBS Lett. **407:**320–324.
51. **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. (Erratum, **11:**382, 1993.)
52. **Piedra, M. E., J. M. Icardo, M. Albajar, J. C. Rodriguez-Rey, and M. A. Ros.** 1998. Pitx2 participates in the late phase of the pathway controlling left-right asymmetry. Cell **94:**319–324.
53. **Rivera-Pomar, R., X. Lu, H. Taubert, N. Perrimon, and H. Jackle.** 1995. Activation of posterior gap gene expression in the *Drosophila* blastoderm. Nature **376:**253–256.
54. **Rubin, C. M., and C. W. Schmid.** 1980. Pyrimidine-specific chemical reactions useful for DNA sequencing. Nucleic Acids Res. **8:**4613–4619.
55. **Ryan, A. K., B. Blumberg, C. Rodriguez-Esteban, S. Yonei-Tamura, K. Tamura, T. Tsukui, J. de la Pena, W. Sabbagh, J. Greenwald, S. Choe, D. P. Norris, E. J. Robertson, R. M. Evans, M. G. Rosenfeld, and J. C. Izpisua Belmonte.** 1998. Pitx2 determines left-right asymmetry of internal organs in vertebrates. Nature **394:**545–551.
56. **Schwabe, J. W., L. Chapman, and D. Rhodes.** 1995. The oestrogen receptor recognizes an imperfectly palindromic response element through an alternative side-chain conformation. Structure **3:**201–213.
57. **Semina, E. V., R. Reiter, N. J. Leysens, W. L. M. Alward, K. W. Small, N. A. Datson, J. Siegel-Bartelt, D. Bierke-Nelson, P. Bitoun, B. U. Zabel, J. C. Carey, and J. C. Murray.** 1996. Cloning and characterization of a novel bicoid-related homeobox transcription factor gene, RIEG, involved in Rieger syndrome. Nat. Genet. **14:**392–399.
58. **Small, S., A. Blair, and M. Levine.** 1992. Regulation of even-skipped stripe 2 in the *Drosophila* embryo. EMBO J. **11:**4047–4057.
59. **Struhl, G., K. Struhl, and P. Macdonald.** 1989. The gradient morphogen bicoid is a concentration-dependent transcriptional activator. Cell **57:**1259–1273.
60. **Suzuki, M., and N. Yagi.** 1994. DNA recognition code of transcription factors in the helix-turn-helix, probe helix, hormone receptor, and zinc finger families. Proc. Natl. Acad. Sci. USA **91:**12357–12361.
61. **Tell, G., R. Acquaviva, S. Formisano, F. Fogolari, C. Pucillo, and G. Damante.** 1999. Comparative stability analysis of the thyroid transcription factor 1 and Antennapedia homeodomains: evidence for residue 54 in controlling the structural stability of the recognition helix. Int. J. Biochem. Cell Biol. **31:**1339–1353.
62. **Treisman, J., P. Gönczy, 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.
63. **Treisman, J., E. Harris, D. Wilson, and C. Desplan.** 1992. The homeodomain: a new face for the helix-turn-helix? Bioessays **14:**145–150.
64. **Truss, M., G. Chalepakis, and M. Beato.** 1990. Contacts between steroid hormone receptors and thymines in DNA: an interference method. Proc. Natl. Acad. Sci. USA **87:**7180–7184.
65. **Tsai, C., and J. P. Gergen.** 1994. Gap gene properties of the pair-rule gene runt during *Drosophila* segmentation. Development **120:**1671–1683.

66. **Tucker-Kellogg, L., M. A. Rould, K. A. Chambers, S. E. Ades, R. T. Sauer, and C. O. Pabo.** 1997. Engrailed (Gln50→Lys) homeodomain-DNA complex at 1.9 A resolution: structural basis for enhanced affinity and altered specificity. Structure **5:**1047–1054.
67. **Tung, C. S.** 1999. Structural study of homeodomain protein-DNA complexes using a homology modeling approach. J. Biomol. Struct. Dyn. **17:**347–354.
68. **Weiler, S., J. M. Gruschus, D. H. Tsao, L. Yu, L. H. Wang, M. Nirenberg, and J. A. Ferretti.** 1998. Site-directed mutations in the vnd/NK-2 homeodomain. Basis of variations in structure and sequence-specific DNA binding. J. Biol. Chem. **273:**10994–11000.
69. **Weiner, S. J., P. A. Kollman, D. T. Nguyen, and D. A. Case.** 1986. An all atom force field for simulations of proteins and nucleic acids. J. Computational Chem. **7:**230–252.
70. **Wilson, D. S., G. Sheng, S. Jun, and C. Desplan.** 1996. Conservation and diversification in homeodomain-DNA interactions: a comparative genetic analysis. Proc. Natl. Acad. Sci. USA **93:**6886–6891.
71. **Wimmer, E. A., M. Simpson-Brose, S. M. Cohen, C. Desplan, and H. Jackle.** 1995. Trans- and cis-acting requirements for blastodermal expression of the head gap gene buttonhead. Mech. Dev. **53:**235–245.
72. **Wolberger, C.** 1996. Homeodomain interactions. Curr. Opin. Struct. Biol. **6:** 62–68.
73. **Wolberger, C.** 1993. Transcription factor structure and DNA binding. Curr. Opin. Struct. Biol. **3:**3–10.
74. **Wolberger, C., A. K. Vershon, B. Liu, A. D. Johnson, and C. O. Pabo.** 1991. Crystal structure of a MATα2 homeodomain-operator complex suggests a general model for homeodomain-DNA interactions. Cell **67:** 517–528.
75. **Yamanaka, Y., T. Mizuno, Y. Sasai, M. Kishi, H. Takeda, C. H. Kim, M. Hibi, and T. Hirano.** 1998. A novel homeobox gene, dharma, can induce the organizer in a non-cell-autonomous manner. Genes Dev. **12:**2345–2353.
76. **Yuan, D., X. Ma, and J. Ma.** 1999. Recognition of multiple patterns of DNA sites by *Drosophila* homeodomain protein Bicoid. J. Biochem. **125:**809–817.
77. **Yuan, D., X. Ma, and J. Ma.** 1996. Sequences outside the homeodomain of Bicoid are required for protein-protein interaction. J. Biol. Chem. **271:** 21660–21665.
78. **Zhao, C., V. Dave, F. Yang, and J. Ma.** Target selectivity of Bicoid is dependent on noncensensus site recognition and protein-protien interaction. Mol. Cell. Biol., in press.