Nucleotides Flanking a Conserved TAAT Core Dictate the DNA Binding Specificity of Three Murine Homeodomain Proteins

KATRINA M. CATRON,¹ NANCY ILER,^{1,2} AND CORY ABATE^{1,3*}

Center for Advanced Biotechnology and Medicine, 679 Hoes Lane,¹ Graduate Program in Microbiology, Rutgers University,² and Department of Neuroscience and Cell Biology, Robert Wood Johnson Medical School,³ Piscataway, New Jersey, 08854-5638

Received 17 September 1992/Returned for modification 9 November 1992/Accepted 7 January 1993

Murine homeobox genes play a fundamental role in directing embryogenesis by controlling gene expression during development. The homeobox encodes a DNA binding domain (the homeodomain) which presumably mediates interactions of homeodomain proteins with specific DNA sites in the control regions of target genes. However, the bases for these selective DNA-protein interactions are not well defined. In this report, we have characterized the DNA binding specificities of three murine homeodomain proteins, Hox 7.1, Hox 1.5, and En-1. We have identified optimal DNA binding sites for each of these proteins by using a random oligonucleotide selection strategy. Comparison of the sequences of the selected binding sites predicted a common consensus site that contained the motif (C/G)TAATTG. The TAAT core was essential for DNA binding activity, and the nucleotides flanking this core directed binding specificity. Whereas variations in the nucleotides flanking the 5' side of the TAAT core produced modest alterations in binding activity for all three proteins, perturbations of the nucleotides directly 3' of the core distinguished the binding specificity of Hox 1.5 from those of Hox 7.1 and En-1. These differences in binding activity reflected differences in the dissociation rates rather than the equilibrium constants of the protein-DNA complexes. Differences in DNA binding specificities observed *in vitro* may contribute to selective interactions of homeodomain proteins with potential binding sites in the control regions of target genes.

Studies aimed at elucidating the molecular processes that regulate murine embryonic development have led to the identification of numerous genes whose protein products function to control gene expression during embryogenesis. These transcriptional regulatory proteins play a fundamental role in directing embryogenesis by establishing and maintaining the appropriate patterns of spatial and temporal gene expression. At least 30 such genes have been isolated in mice (29). Many of these share a conserved motif termed the homeobox which encodes a sequence-specific DNA binding domain (the homeodomain) (32, 47). The homeobox was first identified as a common feature of *Drosophila* genes that control pattern formation during embryogenesis and was subsequently shown to be characteristic of numerous genes from a variety of species that regulate transcription during development (5, 16, 29, 47, 51). Of the 60 amino acids that compose the homeodomain, some of the most highly conserved are those that contact DNA directly (30, 32, 41). Some of these conserved amino acids contact nucleotides within a TAAT motif, a critical component of many homeodomain DNA binding sites. Other amino acids make additional contacts with nucleotides flanking the TAAT motif, and such contacts are presumed to influence the specificity of the protein-DNA interaction (20, 21, 48).

Several distinct families of murine homeobox genes are expressed in overlapping spatial and temporal patterns in the developing embryo (25, 29). Members of these gene families share sequence similarity with the homeobox regions of distinct *Drosophila* developmental regulatory genes. The most prevalent group, the *hox* gene family, contains at least

20 members that are clustered on four chromosomes, and these are related to the *Drosophila* homeotic gene, *antennapedia* (5, 37, 47). The *hox* genes are expressed early in the developing murine embryo (beginning from days 7.5 to 8.5) in overlapping patterns throughout the neural tube, the somites, and the sclerotomes (25, 37). These genes have the intriguing property that their anterior boundaries of expression correlate well with their linear organization along the chromosome cluster (13, 18, 37, 50). This colinearity of chromosome organization and gene expression has been conserved from *Drosophila* to *Homo sapiens* and is proposed to represent a molecular code that provides positional information during development. Indeed, targeted disruptions of *hox 1.5* or *hox 1.6* produce gross developmental abnormalities, highlighting a critical role for these genes in pattern formation during development (9, 33).

In addition to the *hox* genes, members of several other homeobox gene families are also expressed in overlapping spatial and temporal patterns during development. The murine genes *engrailed-1* and *hox 7.1*, for example, are members of relatively small gene families (two or three members), and their homeobox sequences are more closely related to those of the *Drosophila* genes *engrailed* and *msh* (for muscle-specific homeobox), respectively, than to that of the *antennapedia* gene (22, 28). Both *engrailed-1* and *hox 7.1*, similar to the *antennapedia*-like *hox* genes, are expressed throughout the developing neural tube, but unlike the *hox* genes, their anterior boundaries of expression extend to the presumptive midbrain and diencephalon (10, 22, 28, 34, 45). Their gene products, En-1 and Hox 7.1, have the hallmark features of transcriptional regulatory proteins and are presumed to function in the control of gene expression during embryogenesis.

* Corresponding author.

As a consequence of their overlapping, but distinct, patterns of expression, many different combinations of homeodomain proteins may be present in developing cells. Presumably, the interactions of these various proteins with specific target genes are essential for maintaining appropriate gene expression, thereby dictating proper cell development. The most conserved feature among these proteins, the homeodomain, is also a major contributor to their functional specificity, at least for certain *Drosophila* proteins (17, 31, 35). Therefore, binding specificity and function may be intimately linked. A comprehensive analysis of their binding properties is likely to provide insight as to how these proteins interact selectively with specific target genes, and this is particularly relevant in situations in which numerous homeodomain proteins are coexpressed. In this paper, we have investigated the DNA binding specificities of three murine homeodomain proteins, Hox 7.1, Hox 1.5, and En-1. Using a random oligonucleotide selection strategy, we have identified optimal DNA binding sites for each of these proteins. We show that the homeodomain regions of Hox 7.1, Hox 1.5, and En-1 bind to similar DNA sites that contain a TAAT core flanked on either side by G or C residues [e.g., (C/G)TAATTG]. Both the core and flanking nucleotides are required for high-affinity DNA binding. We also show that the specific nucleotides directly flanking the 3' side of the core distinguish the binding specificities of Hox 7.1, Hox 1.5, and En-1. Subtle differences in DNA binding specificities observed *in vitro* may contribute to selective interactions of homeodomain proteins with potential binding sites in the control regions of target genes.

MATERIALS AND METHODS

Expression and purification of Hox 7.1, Hox 1.5, and *Engrailed-1*. DNA sequences corresponding to the homeobox regions of *hox 7.1* and *engrailed-1* were obtained by polymerase chain reaction (PCR) with cDNA derived from mouse 9.5-day embryonic RNA (a generous gift of J. McMahon and A. McMahon, Roche Institute of Molecular Biology). The homeobox sequence of *hox 1.5* was isolated by PCR amplification from an 11.5-day mouse embryonic cDNA library (Clontech). Oligonucleotides used for amplification were complementary to sequences encoding amino acids 157 to 233 of Hox 7.1 (22), the homeodomain region of Hox 1.5 (74 amino acids) (36), and the C-terminal region of *Engrailed-1* (En-1) including the homeodomain (130 amino acids) (28). The oligonucleotides also contained *Bam*HI and *Hind*III restriction sites to facilitate cloning. The amplified fragments of *hox 7.1*, *hox 1.5*, and *engrailed-1* were cloned into the bacterial expression vector pDS56 (3), and their sequences were confirmed. This vector contains six histidine codons, and the proteins were expressed as hexahistidine fusion polypeptides; in our experience, the hexahistidine fusion does not interfere with DNA binding activity (1–3). Bacterial cells containing pDS56-*hox 7.1*, pDS56-*hox 1.5*, or pDS56-*engrailed-1* were grown to a density of 0.6 optical density units, and protein expression was induced by the addition of 1 mM isopropyl- β -D-thiogalactopyranoside. After 4 h, cells were harvested, protein lysates were prepared by extraction with 6 M guanidine-HCl, and the proteins were purified from the bacterial cell lysates by nickel-affinity chromatography in the presence of 6 M guanidine-HCl as previously described (2, 3). The purified polypeptides were renatured by extensive dialysis against 25 mM sodium phosphate (pH 7.4), 50 mM potassium chloride, 5 mM magnesium chloride, 10% (vol/vol) glycerol, and 1 mM dithiothreitol.

For expression in mammalian cells, *hox 7.1* was cloned into the vector pC6+ (42) and transfected into Cos-1 cells by a standard DEAE-dextran procedure. Details of these procedures will be published elsewhere (26). Cell lysates were prepared from Cos-1 cells expressing Hox 7.1, and the protein was partially purified by nickel-affinity chromatography with an imidazole gradient for elution as previously described (27).

Oligonucleotide selection. Oligonucleotide selection was performed with a DNA fragment that contained a 14-bp random sequence flanked on either side by 15 bases of nonrandom sequence [5'-AGACGGATCCATTGCA(N14)CTGTAGGAATTCGGA-3'] (a generous gift of J. Morris and F. J. Rauscher III, Wistar Institute). The single-stranded oligonucleotide containing the random sequence was made double stranded by filling in with the Klenow fragment of DNA polymerase I with a template oligonucleotide that was complementary to the 3' nonrandom sequence [5'-TCCGAATTCCTACAG-3' (oligonucleotide SMB)]. Maxam-Gilbert sequencing was performed to confirm that the probe contained a random sequence. PCR amplification of the double-stranded fragment was performed with the template oligonucleotide (SMB) and a second oligonucleotide that corresponded to the 5' nonrandom sequence [5'-AGACGGA TCCATTGCA-3' (oligonucleotide SMA)]. The double-stranded fragment was radiolabeled with T4 polynucleotide kinase in the presence of [γ -³²P]ATP and used as a probe in gel retardation assays. Binding reactions were performed by addition of protein (1 to 2 μ M) to buffer containing 10 mM Tris-HCl (pH 7.5), 50 mM sodium chloride, 7.5 mM magnesium chloride, 1 mM EDTA, 5% (vol/vol) glycerol, 5% (wt/vol) sucrose, 0.1% Nonidet P-40, 0.5 μ g of bovine serum albumin per μ l, 0.5 μ g of poly(dI-dC), and 5 mM dithiothreitol. After incubation for 5 min at room temperature, the radiolabeled DNA probe was added, and the incubation was continued for an additional 5 min. The high levels of poly(dI-dC) and short incubation time were required to minimize aggregation of DNA-protein complexes. The protein-DNA complexes were resolved from free DNA on 6.5% polyacrylamide gels containing 0.5 $\times$ Tris-borate-EDTA. The DNA in the bound complexes was identified by autoradiography and extracted from the gel. PCR amplification was performed with a standard reaction mixture (Perkin-Elmer Cetus) and the following conditions: 93°C, 30 s; 45°C, 2 min; 45 to 67°C, 2 min; and 67°C, 2 min (25 cycles) and 45°C, 10 min; 45 to 67°C, 10 min; and 67°C, 10 min (1 cycle). The amplified DNA was gel purified, radiolabeled with T4 polynucleotide kinase, and used as a probe in gel retardation assays. The selection procedure was repeated for a total of four rounds. In the initial rounds, less than 1% of the probe was shifted and by the fourth round of selection approximately 50% of the probe was shifted. After the last round of PCR amplification, the gel-purified DNA fragments were phosphorylated and cloned into the *Eco*RV site of pBlue-script (Stratagene). Clones containing single inserts were identified, and the sequences of the inserts were determined. To investigate the binding potential of the selected sites, DNA probes were prepared by PCR amplification of plasmids containing the selected binding sites with radiolabeled oligonucleotide primers (SMA and SMB, as above). A total of 1 μ l of the PCR product was used as a probe in the binding assay performed above.

DNA binding assays. For binding assays with oligonucleotide probes, oligonucleotides were synthesized on an Applied Biosystems DNA synthesizer model 391. The oligonucleotides were radiolabeled with T4 polynucleotide kinase

FIG. 1. Cloning and expression of the homeodomain regions of Hox 7.1, Hox 1.5, and En-1. (A) Predicted amino acid sequences corresponding to the homeodomain regions of Hox 7.1 (22), Hox 1.5 (36), and En-1 (28) are indicated with the single-letter code; amino acids conserved among the proteins are underlined. The consensus amino acids are those residues that are highly conserved among all homeodomain proteins (47). The position of the N-terminal arm and helices I, II, and III are indicated, and positions of amino acids within the homeodomain are indicated numerically. (B) DNA sequences corresponding to the homeobox regions of *hox 7.1*, *hox 1.5*, and *en-1* were amplified by PCR with murine embryonic 9.5- or 11.5-day cDNA. Oligonucleotide primers were specific for the 5' or 3' sequences and contained *Bam*HI and *Hind*III restriction sites. The amplified DNA fragments were cloned into the *E. coli* expression vector pDS56 in frame with the initiator methionine and six histidine codons. (C) Polypeptides corresponding to the homeodomain regions of Hox 7.1, Hox 1.5, and En-1 were expressed as hexahistidine fusion proteins in *E. coli* and purified by nickel-affinity chromatography. The purified proteins (2 μ g) were resolved on an SDS-13.5% polyacrylamide gel and visualized by staining with Coomassie brilliant blue. Markers correspond to molecular mass standards (Bio-Rad) in kilodaltons (bovine serum albumin, 68 kDa; ovalbumin, 46 kDa; carbonic anhydrase, 31 kDa; soybean trypsin inhibitor, 20 kDa; and lysozyme, 14 kDa).

and annealed at 37°C for 1 h at equimolar concentrations. The binding reactions were performed as described above, with the exception that the protein-DNA complexes were formed for 15 min at room temperature. DNA binding activity was quantitated with a Phosphor Imager (Molecular Dynamics) and was calculated as the percentage of DNA bound divided by total DNA (bound/bound + free). Each DNA binding experiment was repeated a minimum of four times, and representative gel shift assays are presented.

For determination of equilibrium dissociation constants (K_d), a constant amount of labeled DNA (1×10^{-10} M) was incubated with various amounts of protein (5×10^{-10} to 10×10^{-7} M for Hox 7.1 and Hox 1.5 and 5×10^{-9} to 5×10^{-6} M for En-1) for 30 min at room temperature (to allow protein-DNA complexes to reach equilibrium) in reaction buffer minus poly(dI-dC). The reaction mixtures were electrophoresed, under the conditions described above, in the absence of loading dye. Bound and free DNAs were quantitated with a Phosphor Imager. The K_d was calculated with the equation $K_d = [D][P]/[DP]$, where [D] is the concentration of free DNA, [P] is the concentration of free protein, and [DP] is the concentration of the DNA-protein complex. The DNA concentration was limiting relative to the protein concentration to allow the approximation $[DP] \approx [DP_{total}]$.

The dissociation rate constants (k_d) and complex half-lives ($t_{1/2}$) were determined by incubating the proteins with limit-

ing concentrations of radiolabeled DNA in binding reaction mixture [without poly(dI-dC)]. After incubation at room temperature for 15 min, a 50-fold molar excess of cold competitor DNA was added and the reaction mixtures were incubated for the additional times indicated prior to electrophoresis. The bound and free DNAs were quantitated with a PhosphorImager, and the values for k_d were determined with the formula $\ln(\text{fraction of DNA bound}) = -k_d t$. $t_{1/2}$ (the time required for half of the protein-DNA complexes to dissociate) was calculated with the equation $t_{1/2} = -\ln(0.5)/k_d$.

RESULTS

Expression and purification of the murine homeodomain proteins Hox 7.1, Hox 1.5, and Engrailed-1. The homeodomain regions of Hox 7.1, Hox 1.5, and En-1 share many amino acids that are highly conserved among all homeodomain proteins (Fig. 1A). These similarities are primarily concentrated within helix III, and some of these conserved amino acids have been shown by structural studies to contact DNA directly (30, 41). The conservation is more limited throughout the other helices, and overall, the homeodomains of Hox 7.1, Hox 1.5, and En-1 share approximately 50% homology at the amino acid level (Fig. 1A). Therefore, we reasoned that these proteins share sufficient sequence homology to facilitate a comparative analysis of their DNA

binding specificities. To investigate their binding properties in vitro, purified polypeptides were obtained by overexpression of the homeobox sequences of *hox 7.1*, *hox 1.5*, and *en-1* in *Escherichia coli*. The appropriate sequences were isolated from mouse embryonic cDNA and cloned into the bacterial expression vector pDS56 (Fig. 1B). This vector contained six histidine codons after the initiator methionine codon, such that the recombinant polypeptides were expressed as hexahistidine fusion proteins (2, 3). This histidine fusion facilitated subsequent purification by nickel-affinity chromatography. Protein lysates were prepared from bacterial cells expressing the recombinant Hox 7.1, Hox 1.5, or En-1 polypeptides by extraction with buffer containing 6 M guanidine. The lysates were chromatographed in the presence of guanidine on a nickel-affinity column, and the purified proteins were renatured by extensive dialysis. The proteins were soluble at relatively high concentrations (Hox 7.1, 800 μ M; Hox 1.5, 70 μ M; and En-1, 350 μ M) and were purified to apparent homogeneity as evident by Coomassie blue staining of a sodium dodecyl sulfate (SDS)-polyacrylamide gel (Fig. 1C).

Selection of DNA binding sites for Hox 7.1, Hox 1.5, and En-1. The random oligonucleotide selection procedure has been used successfully to identify optimal DNA binding sites for a variety of transcriptional regulatory proteins (6, 14, 15, 40, 43). This strategy relies on the ability of a protein to interact with specific DNA sites selected from among a random population of potential binding sites. In this report, we have used this procedure to characterize the DNA binding specificities of Hox 7.1, Hox 1.5, and En-1 (Fig. 2A). Binding selection was performed with a double-stranded DNA fragment that contained a 14-bp random sequence. The DNA fragment was used as a probe in gel retardation assays, and those fragments that interacted with the purified Hox 7.1, Hox 1.5, or En-1 polypeptides were separated from the population of random sites by gel electrophoresis (Fig. 2A). In the initial round of selection, a 1 to 3 μ M concentration of each protein was used in the gel retardation assays and less than 1% of the probe was retained in the bound complex. In subsequent rounds of selection, the protein concentration was reduced to 0.5 μ M, and by the fourth round approximately 50% of the probe was shifted by the proteins. After the fourth round, the DNA sites obtained by selection with Hox 7.1, Hox 1.5, and En-1 were cloned and sequenced. The sequences of selected sites and the relative binding affinities of Hox 7.1, Hox 1.5, and En-1 for each of these sites are summarized in Fig. 2B to D.

Hox 7.1 binding specificity. Each of 13 sites selected by Hox 7.1 from the random pool was capable of binding the cognate protein (Fig. 2B). Although all of these sites interacted with Hox 7.1, they exhibited marked differences in the degree of binding, and these differences were most evident when binding assays were performed at lower protein concentrations (Fig. 2B). Therefore, using the selection procedure, we obtained a series of potential DNA sites for Hox 7.1 that included relatively high-affinity (e.g., sites 6 and 35), intermediate-affinity (e.g., 14A, 12A, 34A, 34B, 10, and 14B), and low-affinity (e.g., 24A, 26, 7, 2, and 5) binding sites (Fig. 2B). The presence of such low-affinity sites, even after four rounds of selection, is presumably due to the inherent nonspecific binding activity of Hox 7.1 (7). Many of the selected sites contained AT-rich sequences that were flanked by G or C nucleotides (e.g., Fig. 2B, sites 6, 35, and 34B). Alignment of these sequences (with a combination of computer-generated and visual alignment) identified the consensus binding site, CXGTAAT(A/T)G (Fig. 2B). Interestingly,

the sequence of the consensus site was more similar to that of the highest-affinity binding sites (e.g., sites 6 and 35) than the lower-affinity sites (e.g., sites 7, 2, and 5) (Fig. 2B).

Hox 1.5 binding specificity. Each of the binding sites recovered with Hox 1.5 as a substrate in the selection procedure interacted with the cognate protein in vitro (Fig. 2C). However, Hox 1.5 differed markedly in its apparent affinity for each of these binding sites, analogous to the situation with Hox 7.1 (Fig. 2C). Thus, the population of Hox 1.5 binding sites included relatively high-affinity (e.g., sites 32A and 14), intermediate-affinity (e.g., sites 2, 27B, 13, 16A, 16C, 27A, 31A, 35, and 34), and low-affinity (e.g., 20A, 3, and 12) binding sites (Fig. 2C). As with the Hox 7.1 sites, the Hox 1.5 sites contained AT-rich sequences flanked by G or C nucleotides (compare Fig. 2B and C). However, the predicted consensus sequence for Hox 1.5 [C(C/G)TAATX(G/T)(G/T)] was somewhat different from the consensus site for Hox 7.1 and had more variability for G or T nucleotides on the 3' side of the TAAT core (compare Fig. 2B and C). This may reflect a greater flexibility of Hox 1.5 to interact with DNA sites that contain variations of the 3' flanking nucleotides. The sequence of the Hox 1.5 consensus site was more similar to that of the highest-affinity binding sites (e.g., sites 32A and 14) than the lower-affinity sites (e.g., 20A, 3, and 12) (Fig. 2C).

En-1 binding specificity. By using En-1 as a substrate in the binding selection procedure, several DNA binding sites were identified (Fig. 2D). However, although selection with En-1 yielded some relatively high-affinity sites (e.g., 12B, 2C, and 30C), a majority of the selected sites were of relatively low affinity (Fig. 2D). We have observed that twofold-higher concentrations of En-1 (2 μ M) are required for efficient DNA binding compared with either Hox 7.1 (0.9 μ M) or Hox 1.5 (0.7 μ M) (7). It is likely that the reduced efficiency of the selection procedure was due to an overall lower affinity of En-1 for DNA. The sequences of the binding sites selected for En-1 were similar to those selected for Hox 7.1 and Hox 1.5 in that they contained AT-rich sequences flanked by G or C nucleotides (Fig. 2D). Although fewer sites were obtained with En-1, comparison of these sequences indicated a tentative consensus binding site, GTAATXG (Fig. 2D).

Hox 7.1, Hox 1.5, and En-1 have related binding specificities in vitro. The highest-affinity binding sites selected for Hox 7.1, Hox 1.5, and En-1 were closely related (compare Fig. 2B, C, and D). In particular, these sites each contained a TAAT core flanked on the 5' and 3' sides by C or G nucleotides. These findings suggested that the homeodomains of Hox 7.1, Hox 1.5, and En-1 bind to similar DNA sites in vitro. Indeed, it has been shown that rather divergent *Drosophila* homeodomains interact with the same sites in vitro (11, 24). Hox 7.1, Hox 1.5, and En-1 each interacted with sites that had been selected for binding to the other proteins and bound to these other sites with approximately the same relative affinity as the cognate protein (Fig. 3A to D). Therefore, all three proteins interacted most efficiently with sites 6, 35, 32A, and 12B and less efficiently with sites 12A, 34A, 13, and 30A (Fig. 3A to D). In contrast, other DNA binding proteins, such as Fos and Jun, did not bind to these sites, although these other proteins also contained hexahistidine fusion polypeptides (7). Thus, the homeodomain regions of Hox 7.1, Hox 1.5, and En-1 exhibit similar DNA binding specificities in vitro.

Nucleotides flanking the TAAT core contribute to binding affinity of Hox 7.1, Hox 1.5, and En-1. The selected binding sites for Hox 7.1, Hox 1.5, and En-1 were biased toward those containing TAAT core nucleotides flanked on the 5'

FIG. 2. Selection of optimal DNA binding sites for Hox 7.1, Hox 1.5, and En-1. (A) Strategy for random oligonucleotide selection of DNA binding sites. A double-stranded DNA fragment containing 14 bp of random sequence (variable N14) was radiolabeled and used as probe in gel retardation assays as detailed in Materials and Methods. Protein-DNA complexes (bound DNA) were separated from free DNA by polyacrylamide gel electrophoresis. The DNA contained in the bound complexes was identified by autoradiography, isolated, amplified by PCR, and used as a probe in subsequent gel retardation assays. After four rounds of selection, the DNA sites were cloned and sequenced. (B to D) Characterization of selected DNA binding sites. The nucleotide sequences of the selected DNA binding sites (indicated numerically) are shown. Nucleotides in capital letters are those contributed from the random sequence of the DNA site (panel A); nucleotides in lowercase are those contributed from the nonrandom sequence of the DNA probe. The sites were aligned with a combination of computer-generated and visual alignment, and the consensus site is indicated in the shaded box. Selected DNA binding sites were amplified by PCR with radiolabeled oligonucleotides and used as a probe in gel retardation assays with increasing concentrations of protein (as indicated by the triangles). The concentration of Hox 7.1 was 0, 0.30, or 0.90 μ M (B); the concentration of Hox 1.5 was 0, 0.20, or 0.70 μ M (C); and the concentration of En-1 was 0, 0.60, or 2.0 μ M (D). The DNA-protein complexes were resolved from free DNA on 6.5% nondenaturing polyacrylamide gels and visualized by autoradiography. The relative degree of binding activity to each site is indicated. ++++ denotes a high degree of binding, and + denotes a low degree of binding; +/- denotes minimal binding activity.

and 3' sides by C or G nucleotides (Fig. 2). In fact, the consensus sites for all three proteins contained the motif (G/C)TAATTG. To determine the relative contribution of nucleotides flanking and within the TAAT core towards binding activity of Hox 7.1, Hox 1.5, and En-1, we examined the interaction of these proteins with DNA sites that contained permutations of the (G/C)TAATTG motif (Fig. 4A). The prototypic sequence for these studies was site 6 (Fig.

2B) since it contained the (G/C)TAATTG motif and since all three proteins bound to this site with relatively high affinities (Fig. 3A). Hox 7.1, Hox 1.5, and En-1 were tested for their abilities to bind to a series of DNA sites that contained substitutions within the core or substitutions of nucleotides 5' or 3' of the core (Fig. 4A). To facilitate direct comparisons for preferences among the various sites, the DNA sites were tested in parallel with equimolar concentrations labeled to

FIG. 3. Hox 7.1, Hox 1.5, and En-1 bind with similar specificities in vitro. (A to C) DNA binding sites for Hox 7.1 (A), Hox 1.5 (B), or En-1 (C) were obtained by selection from a population of random DNA sites (as in Fig. 2). The sites were radiolabeled and used as probes in gel retardation assays containing various concentrations (indicated by triangles) of Hox 7.1 (0, 0.30, or 0.9 μ M), Hox 1.5 (0, 0.2, or 0.7 μ M), or En-1 (0, 0.60, or 2.0 μ M). Protein-DNA complexes were resolved from free DNA on 6.5% nondenaturing polyacrylamide gels, and the complexes were visualized by autoradiography. (D) Nucleotide sequences of the DNA binding sites (as in Fig. 2) are shown. The relative binding activities of Hox 7.1, Hox 1.5, and En-1 (as in Fig. 3A to C) for each of the sites are indicated; ++++ denotes a high degree of binding activity, and + indicates relatively low binding activity.

approximately the same specific activity (Fig. 4B to D). Binding activity was linear over the range of protein concentrations tested, and at the highest concentration of protein, maximal binding to the prototypic site was not greater than 50% (Fig. 4B to D).

Contributions of nucleotides 5' of the TAAT core. Substitutions of nucleotides 5' of the TAAT core produced modest, but reproducible, alterations in the binding activity of Hox 7.1, Hox 1.5, and En-1 (Fig. 4A, compare 6 with 6-1 to 6-6). All three proteins bound preferentially to DNA sites that contained a purine nucleotide (A or G) at position 3 and bound less efficiently to sites that contained a T or C in this position (Fig. 4A, compare 6 with 6-1 to 6-3). In general, substitutions at position 4 were less well tolerated than substitutions at position 3 (Fig. 4A, compare 6 with 6-4 to 6-6). All three proteins bound preferentially to DNA sites that contained C⁴, while DNA sites that contained G⁴ or T⁴ were bound better than those that contained an A at this position (Fig. 4A, compare 6 with 6-4 to 6-6). Surprisingly, a DNA site that contained C³T⁴ was bound more efficiently by all three proteins than a DNA site that contained A³T⁴ (Fig. 4A, compare 6-6 with 6-7). This suggests that the nucleotide at position 3 may influence the ability of the proteins to bind to DNA sites that contain certain differences at position 4.

Contributions of nucleotides 3' of the TAAT core. Substitutions of nucleotides directly flanking the 3' side of the TAAT core significantly reduced binding of Hox 7.1 and En-1 (Fig. 4A, compare 6 with 6-9 to 6-14). In contrast, most substitutions of these nucleotides were well tolerated by Hox 1.5 (Fig. 4A, compare 6 with 6-9 to 6-14). In particular, the binding activity of Hox 7.1 and En-1 was markedly

reduced for DNA sites that contained substitutions of T⁹ with any other nucleotide (Fig. 4A, compare 6 with 6-9 to 6-11). However, Hox 1.5 bound very efficiently to a DNA site that contained a G at position 9, and sites that contained A or C were also bound by Hox 1.5, although to a lesser extent (Fig. 4A, compare 6 with 6-9 to 6-11). Although all three proteins bound preferentially to DNA sites that contained a purine nucleotide (A or G) at position 10, other substitutions in this position were better tolerated by Hox 1.5 than by either Hox 7.1 or En-1 (Fig. 4A, compare 6 with 6-12 to 6-14). Nucleotides further 3' of the TAAT core (e.g., position 11) may also contribute to binding specificity. En-1, but not Hox 7.1 or Hox 1.5, bound less efficiently to a DNA site that contained a substitution of G¹¹ with C (Fig. 4A, compare 6 with 6-15). In contrast, substitution of nucleotides at positions 13 and 14 alone did not affect binding activity (Fig. 4A, compare 6 to 6-16). These findings demonstrate that Hox 7.1, Hox 1.5, and En-1 have distinct preferences for nucleotides 3' of the TAAT core. These preferences are likely to contribute to the binding specificity among the three proteins.

Contribution of nucleotides within the TAAT core. Substitutions of nucleotides within the TAAT core were most deleterious for binding of Hox 7.1, Hox 1.5, and En-1 (Fig. 4A, compare 6 with 6-19 to 6-22). Substitution of A⁷ with a C or T nucleotide or insertion of C at this position essentially abolished the binding activity of all three proteins (Fig. 4A, compare 6 with 6-19 to 6-22). Therefore, A⁷ is critical for DNA binding activity of Hox 7.1, Hox 1.5, and En-1. Substitution of T⁸ with A reduced significantly, but did not totally abolish, binding activity (Fig. 4A, compare 6 with 6-22). These data demonstrate that, in contrast to substitutions for flanking nucleotides which reduce binding affinity,

A		FOLD BINDING ACTIVITY		
		Hox 7.1	Hox 1.5	En-1
6	ACCTAATTGGAGGGC	1.00	1.00	1.00
6-1	ACCTAATTGGAGGGC	0.75+/-0.03	0.67+/-0.02	0.64+/-0.02
6-2	ACCTAATTGGAGGGC	0.89+/-0.01	0.89+/-0.03	0.93+/-0.01
6-3	ACCTAATTGGAGGGC	0.76+/-0.03	0.82+/-0.02	0.66+/-0.06
6-4	ACCTAATTGGAGGGC	0.70+/-0.03	0.66+/-0.06	0.68+/-0.09
6-5	ACCTAATTGGAGGGC	0.59+/-0.02	0.49+/-0.04	0.44+/-0.09
6-6	ACCTAATTGGAGGGC	0.67+/-0.03	0.68+/-0.01	0.61+/-0.10
6-7	ACCTAATTGGAGGGC	0.92+/-0.07	0.85+/-0.04	0.90+/-0.07
6-8	ACCTAATTGGAGGGC	0.86+/-0.05	0.70+/-0.06	0.73+/-0.03
6-9	ACCTAATTGGAGGGC	0.28+/-0.03	0.70+/-0.04	0.34+/-0.07
6-10	ACCTAATTGGAGGGC	0.21+/-0.01	0.52+/-0.03	0.20+/-0.05
6-11	ACCTAATTGGAGGGC	0.15+/-0.02	0.85+/-0.06	0.16+/-0.09
6-12	ACCTAATTGGAGGGC	0.81+/-0.02	1.00+/-0.05	1.05+/-0.15
6-13	ACCTAATTGGAGGGC	0.57+/-0.03	0.71+/-0.04	0.48+/-0.04
6-14	ACCTAATTGGAGGGC	0.30+/-0.02	0.58+/-0.01	0.18+/-0.06
6-15	ACCTAATTGGAGGGC	0.84+/-0.05	1.05+/-0.06	0.57+/-0.13
6-16	ACCTAATTGGAGGGC	0.93+/-0.07	1.06+/-0.06	0.94+/-0.10
6-17	ACCTAATTGGAGGGC	0.17+/-0.02	0.38+/-0.02	0.23+/-0.05
6-18	ACCTAATTGGAGGGC	0.19+/-0.03	0.33+/-0.01	0.22+/-0.05
6-19	ACCTAATTGGAGGGC	<0.03	<0.03	<0.03
6-20	ACCTAATTGGAGGGC	<0.03	0.31+/-0.02	0.06+/-0.03
6-21	ACCTAATTGGAGGGC	<0.03	<0.03	<0.03
6-22	ACCTAATTGGAGGGC	0.08+/-0.02	0.15+/-0.04	0.09+/-0.05

B. Hox 7.1

C. Hox 1.5

D. En-1

substitutions within the TAAT core greatly diminish binding activity of Hox 7.1, Hox 1.5, and En-1. Therefore, nucleotides within the TAAT core are essential for DNA binding activity of Hox 7.1, Hox 1.5, and En-1, whereas nucleotides flanking this core contribute to DNA binding affinity and specificity.

Differences in binding activity reflect differences in dissociation rates rather than equilibrium constants. The previous data demonstrate that nucleotides flanking the TAAT core contribute to the binding specificity of Hox 7.1, Hox 1.5, and En-1. To determine the basis for the observed differences in binding activity, the kinetic parameters for binding to the consensus site (site 6) and sites that contained substitutions of the nucleotides 5' or 3' of the TAAT core (sites 6-8, 6-11, and 6-18) were determined (Table 1). The equilibrium dissociation constant (K_d) of Hox 7.1 and Hox 1.5 for the

FIG. 4. DNA binding activity of Hox 7.1, Hox 1.5, and En-1 requires specific nucleotides flanking a TAAT core. (A to D) The binding specificities of Hox 7.1, Hox 1.5, and En-1 were tested with the consensus binding site (site 6) or variations in this site (sites 6-1 to 6-22). Gel retardation assays were performed with an equimolar concentration (5 nM) of each DNA site and were incubated with increasing concentrations of protein (indicated by triangles). The concentration of Hox 7.1 was 0.2 or 0.6 μ M (B), the concentration of Hox 1.5 was 0.2 or 0.6 μ M (C), and En-1 concentration was 0.4 or 0.8 μ M (D). The binding activity was quantitated (from the binding activity of lower protein concentration) with a Phosphor Imager (Molecular Dynamics) and was calculated as the percentage of bound DNA relative to total DNA (bound/bound + free). Binding activity is expressed as the fold difference in binding relative to that of the consensus site (site 6) (A); the data represent the average of four independent experiments, and the standard deviation is indicated. A number of <0.03 indicates that the binding activity was negligible. NA, no protein was added (B to D).

consensus site (site 6) was 2×10^{-9} M (Table 1). In contrast, the K_d of En-1 for site 6 was 10-fold higher (2×10^{-8} M), which is consistent with our observations that En-1 routinely exhibited lower DNA binding activity. For all three proteins, the K_d values for the DNA sites that contained variations in the flanking nucleotides did not differ from that of the consensus site (Table 1). This finding suggested that the observed differences in binding activity (Fig. 4A) were due to differences in the rate of complex formation or dissociation. In fact, the dissociation rates of the protein-DNA complexes (as indicated by half-life [$t_{1/2}$]) differed markedly for each of the DNA sites, in contrast to the similarities observed for the equilibrium dissociation constants (K_d). In particular, a DNA site (site 6-18) that was bound less efficiently by all three proteins (Fig. 4A) exhibited an approximately twofold-lower $t_{1/2}$ compared with that of site 6

TABLE 1. Summary of kinetic parameters^a

Site	K_d (M)	k_d (min ⁻¹)	$t_{1/2}$ (min)	Relative $t_{1/2}$
Hox 7.1				
6	2×10^{-9}	0.55	1.25	1.0
6-18	2×10^{-9}	0.94	0.74	0.59
6-11	1.5×10^{-9}	1.07	0.65	0.52
6-8	1.75×10^{-9}	0.94	0.74	0.59
Hox 1.5				
6	2×10^{-9}	0.70	0.99	1.0
6-18	2×10^{-9}	1.22	0.57	0.58
6-11	2×10^{-9}	0.51	1.36	1.37
6-8	2.5×10^{-9}	0.75	0.92	0.93
En-1				
6	2×10^{-8}	2.22	0.30	1.0
6-18	2×10^{-8}	3.57	0.19	0.63
6-11	2×10^{-8}	3.45	0.20	0.66
6-8	ND ^b	2.70	0.26	0.87

^a Data were derived as described in Materials and Methods.^b ND, not determined.

(Fig. 5 and Table 1). Similarly, site 6-11, which was bound poorly by Hox 7.1 and En-1 (Fig. 4A), also exhibited an approximately twofold-lower $t_{1/2}$ compared with that of site 6 (Fig. 5 and Table 1). Hox 1.5, which bound site 6-11 efficiently (Fig. 4A), had a similar, somewhat longer $t_{1/2}$ compared with that of site 6 (Fig. 5 and Table 1). A site (6-8) that contained substitutions of the 5' flanking nucleotides and was bound relatively well by all three proteins had a modest reduction in the $t_{1/2}$ compared with that of site 6 (Fig. 5 and Table 1). These kinetics studies demonstrate that the differences in binding activity observed with the various DNA sites (Fig. 4A) were due to differences in the dissociation rates of the various DNA-protein complexes rather than differences in the equilibrium dissociation rates. This is somewhat of a paradox, since the equilibrium dissociation rates should parallel the dissociation constants. However, these findings are in good agreement with a recent report by Ekker et al. (13a) in which they showed that the differences in the binding of homeodomain proteins for DNA sites that contained substitutions of flanking nucleotides were mainly due to differences in the dissociation rates rather than the equilibrium constants.

Proteins expressed in mammalian cells bind to DNA sites with specificities similar to that of the bacterially expressed protein. One possible concern when using bacterially expressed proteins to delineate DNA binding specificities is that these may not reflect the binding properties of proteins expressed in mammalian cells. Proteins produced in *E. coli* may be improperly folded or may lack critical posttranslational modifications. To address this possibility, we examined the DNA binding specificities of homeodomain proteins expressed in mammalian cells. For these studies, Hox 7.1 was expressed in Cos-1 cells and was partially purified from cell extracts by nickel-affinity chromatography (Fig. 6A). A gradient of imidazole, rather than guanidine, was used for elution, and therefore the Hox 7.1 protein obtained by this procedure was not subject to a denaturation-renaturation step. By using this procedure, Hox 7.1 binding activity, but not a majority of other cellular proteins, bound to the column at low imidazole concentrations and was eluted with buffer containing 40 mM imidazole (Fig. 6A). The elution profile was also confirmed by Western blot (immunoblot) analysis with polyclonal antisera directed against Hox 7.1 (26). The

binding specificity of Hox 7.1 expressed in mammalian cells (m-Hox 7.1) was similar to that of the bacterially expressed protein (compare Fig. 6B with Fig. 4B). Specifically, m-Hox 7.1 interacted efficiently with the prototypic binding site and also with sites that were bound by the bacterially expressed protein (Fig. 6B).

DISCUSSION

These findings demonstrate that optimal DNA binding activity of Hox 7.1, Hox 1.5, and En-1 requires an obligatory TAAT core and specific nucleotides flanking the 5' and 3' sides of the core. Moreover, the specific 3' flanking nucleotides distinguish the binding specificities among the three proteins. Optimal DNA binding sites for Hox 7.1, Hox 1.5, and En-1 were selected from among a population of random sites by using purified polypeptides. By comparison of the selected binding sites, a common consensus site for Hox 7.1, Hox 1.5, and En-1 that contained the motif (C/G)TAATTG (Table 2) was identified. The consensus site was bound with high affinity by the polypeptides expressed in both bacterial and mammalian cells, whereas other DNA binding proteins, such as Fos and Jun, did not interact with this site. Although there is no direct evidence that the consensus motif functions as a regulatory element in vivo, a computer search of sequences represented in GenBank indicated that this motif was present in the control regions of numerous murine genes, including the complement receptor, the immunoglobulin germline heavy chain V1 gene, the V region of the immunoglobulin active heavy chain, and also in the human GABA receptor. In addition, related DNA sites are present in the control regions of *Drosophila* genes that are developmentally regulated (31, 44). The TAAT core was obligatory for DNA binding, since substitutions of these nucleotides abolished binding activity. In contrast, substitutions of flanking nucleotides reduced binding affinity but did not abolish binding activity. Interestingly, Hox 7.1, Hox 1.5, and En-1 interacted selectively with sites that contained variations in the flanking nucleotides (Table 2). In particular, Hox 7.1 and En-1 exhibited a more restricted preference, compared with that of Hox 1.5, for nucleotides flanking the 3' side of the TAAT core (Table 2). Therefore, the configuration of DNA sites may dictate binding preferences among different classes of homeodomain proteins. These differences in binding affinity and specificity observed in vitro may contribute to selective interactions of these homeodomain proteins with potential regulatory sites in the control regions of target genes.

The TAAT core is a critical feature of many homeodomain DNA binding sites, and structural studies have shown that amino acids highly conserved among all homeodomain proteins contact nucleotides within this core (30, 41). Indeed, these conserved amino acids were also present in Hox 7.1, Hox 1.5, and En-1 (Fig. 1A, Ile-47 and Asn-51), and substitutions of nucleotides that are presumably contacted by these residues were most deleterious for binding. Although it was an essential component of all binding sites, an intact TAAT core was not sufficient for high-affinity binding of Hox 7.1, Hox 1.5, or En-1. Optimal DNA binding required specific flanking nucleotides. In fact, all three proteins interacted with a variety of sites containing AT-rich sequences; however, interaction with these sites was of relatively low affinity compared with binding to the consensus site (7). It is likely that the propensity of these proteins to bind fairly promiscuously, albeit with low affinity, to a range

FIG. 5. Measurement of the dissociation rates of Hox 7.1, Hox 1.5, and En-1. (A to C) The dissociation rates of Hox 7.1, Hox 1.5, and En-1 for the consensus site (site 6) and sites containing substitutions of nucleotides 5' and 3' of the TAAT core (sites 6-8, 6-11, and 6-18) were determined from gel retardation assays as described in Materials and Methods. Proteins were incubated with a limiting concentration of DNA at room temperature for 15 min. Hox 7.1 was included at a concentration of 10 nM (A), the concentration of Hox 1.5 was 10 nM (B), and the concentration of En-1 was 100 nM (C). Subsequent to incubation, a 50-fold excess of unlabeled DNA was added to the reaction mixture, and aliquots were loaded onto the gel at the times indicated. The bound and free DNA were visualized by autoradiography and quantitated on a Phosphor Imager. NCA, no cold competitor added.

FIG. 6. Hox 7.1 expressed in mammalian cells binds to DNA with specificity similar to that of the bacterially expressed protein. (A) Cell extracts were prepared from Cos-1 cells expressing Hox 7.1, and the proteins were partially purified by nickel-affinity chromatography. The load fraction contained Hox 7.1 (indicated by the arrow) and other binding proteins. Cell lysates were resolved on a nickel-affinity column, and the proteins were eluted with buffer containing increasing concentrations (0.8 to 40 mM) of imidazole. Fractions (indicated numerically) were assayed for Hox 7.1 DNA binding activity by gel retardation assay with the consensus binding site (site 6) [Fig. 4]). Protein-DNA complexes were resolved on 6.5% nondenaturing polyacrylamide gels and visualized by autoradiography. (B) DNA binding sites corresponding to the consensus binding site (site 6) or to variations of this site as described in Fig. 4A were incubated in the presence of protein fractions that contained Hox 7.1 binding activity (panel A; fractions 18 and 19). The protein-DNA complexes were resolved from free DNA on a 6.5% polyacrylamide gel and visualized by autoradiography.

of AT-rich DNA sites has hindered attempts to identify optimal binding sites.

The data presented herein demonstrate that Hox 7.1, Hox 1.5, and En-1 bind preferentially to a DNA site that contains the motif (C/G)TAATTG. This sequence is similar to a binding site that was identified for Hox 1.3 with mammalian cell extracts expressing the full-length protein (39). The optimal DNA binding site selected for a fairly divergent *Drosophila* homeodomain, fushi tarazu, also contained similar flanking nucleotides that contributed to high-affinity

binding (e.g., CTTAATTG) (15). Additionally, specific nucleotides flanking a TAAT core also direct binding specificity in vivo (21). In fact, a single amino acid at position 50 distinguished binding to sites that had either a C or TG nucleotide on the 3' side of the core (e.g., TAATC or TAATTG). Specifically, homeodomain proteins that contained glutamine at position 50 interacted with the TAATTG motif, whereas homeodomain proteins that contained a lysine at position 50 preferred the TAATC motif (21). Hox 7.1, Hox 1.5, and En-1, which contain a glutamine at position 50 (Fig. 1A), each bound preferentially to the TAATTG motif compared with the TAATC motif. In combination, in vitro and in vivo studies implicate a role for specific nucleotides flanking the TAAT core in directing the binding affinity of homeodomain proteins.

The most significant differences in the DNA binding specificities of Hox 7.1, Hox 1.5, and En-1 were observed with sites that contained variations in the nucleotides directly 3' of the TAAT core (e.g., TAATTG). For example, whereas all three proteins interacted preferentially with sites containing the TAATTG motif, only Hox 1.5 interacted efficiently with sites containing the TAATG motif. The flexibility of Hox 1.5 to interact with DNA sites that contain variations in the 3' flanking nucleotide to the TAAT core was also reflected by the consensus site selected for Hox 1.5. Other purified Hox polypeptides, including Hox 3.1, Hox 1.1, Hox 1.3, and Hox 1.4, also interacted with DNA sites that contained variations in the 3' nucleotides (7). Moreover, the *Drosophila* Antennapedia-like homeodomains, Antennapedia, Ultrabithorax (UBX), and Deformed, to which the Hox proteins are related, have also been shown to interact with sites that contain a TAATG motif (4, 13a, 14, 44). Therefore, a feature of Antennapedia-like homeodomain proteins which is likely to distinguish their DNA binding specificities from those of other homeodomain proteins is that these interact more promiscuously with sites that contain variations in the 3' flanking nucleotides. Members of the Antennapedia family, including Hox 1.5, contain two amino acids differences within the recognition helix relative to Hox 7.1 and En-1 (Fig. 1). These amino acids may contribute to the differences observed in the site specificity.

Our findings revealed that three murine homeodomains, Hox 7.1, Hox 1.5, and En-1 bind to DNA sites that contain the consensus motif (C/G)TAATTG and that the configuration of these sites distinguished the binding specificities of these proteins. Indeed, the closely related homeodomains, Deformed and UBX, prefer distinct DNA sites in vitro (12, 13a). It is likely that subtle differences in binding affinities observed in vitro have a significant effect in directing binding specificities in vivo. In fact, the differences in binding specificity observed with Deformed and UBX, although modest in vitro, correlated well with their functional specificity in vivo (12, 31). In the context of the cellular environment in which many related homeodomain proteins may be competing for potential binding sites, subtle differences in binding affinities may be greatly exaggerated. We have shown that Hox 7.1, Hox 1.5, and En-1 each interact with a range of DNA binding sites with various degrees of affinity. Therefore, their interaction with potential sites may be influenced by the relative concentrations of these proteins, in addition to coexpression with other homeodomain proteins. Clearly, however, the specificities of homeodomain-DNA interactions are likely to involve additional mechanisms. For example, protein domains other than the homeodomain may also influence DNA binding specificities (8, 46, 49). Moreover, interactions of homeodomain proteins

TABLE 2. Summary of preferences for nucleotides flanking the TAAT core^a

Protein	Preference for nucleotides flanking TAAT core				
Hox 7.1	A,G>T,C	C>G,T>A	T>>A,C,G	G>A>T>C	
Hox 1.5	A,G,C>T	C>G,T>A	T,G>A>C	G,A>T>C	
En-1	A,G>T,C	C>G,T>A	T>>A>C,G	G,A>>T>C	
Consensus	A	C	TAAT	T	G

^a Data are derived from results in Fig. 4 and are described in the text.

with other proteins (19, 38) and competition among transcriptional regulatory proteins for overlapping or adjacent binding sites (23) are likely to play a significant role in directing DNA binding specificity. The studies presented herein establish the appropriate foundation in which to examine these possibilities.

ACKNOWLEDGMENTS

Preliminary experiments were performed in Tom Curran's laboratory, Roche Institute of Molecular Biology, and we are indebted to him for his generous support, encouragement, and many helpful discussions. We thank Jill McMahon and Andy McMahon, Roche Institute of Molecular Biology, for the gift of the mouse embryonic cDNA and for helpful discussions, and Jennifer Morris and Frank J. Rauscher III, Wistar Institute, for the gift of the random oligonucleotide probe and for helpful discussions. We thank Zhigang Shang, Scott Holmes, and Peter Lobel for assistance in analyzing the kinetic data. We acknowledge Aaron Shatkin, Drew Vershon, Mike Tocci, Guy Montelione, and Danny Reinberg for reading the manuscript and Janet Hansen and Alyson Broads for preparation.

This work was supported, in part, by funds to C.A. from the New Jersey Commission on Science and Technology and by a grant (HD29446-01) from the NIH; K.M.C. was supported by a postdoctoral training grant from Hoffmann-La Roche, and N.I. was supported by a predoctoral fellowship from Merck. C.A. is a recipient of a Sinsheimer Scholar award.

REFERENCES

1. Abate, C. Unpublished data.
2. Abate, C., D. Luk, and T. Curran. 1991. Transcriptional regulation by Fos and Jun interaction among multiple activator and regulatory domains. *Mol. Cell. Biol.* **11**:3624-3632.
3. Abate, C., D. Luk, R. Gentz, F. J. Rauscher III, and T. Curran. 1990. Expression and purification of the leucine zipper and DNA-binding domains of Fos and Jun: both Fos and Jun contact DNA directly. *Proc. Natl. Acad. Sci. USA* **87**:1032-1036.
4. 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.
5. Akam, M. 1989. *Hox* and *Hom*: homologous gene clusters in insects and vertebrates. *Cell* **57**:347-349.
6. Blackwell, T. K., and H. Weintraub. 1990. Differences and similarities in DNA-binding preferences of MyoD and E2A protein complexes revealed by binding site selection. *Science* **250**:1104-1110.
7. Catron, K. M., and C. Abate. Unpublished data.
8. Chalepakis, G., R. Fritsch, H. Fickenscher, U. Deutsch, M. Goulding, and P. Gruss. 1991. The molecular basis of the *undulated/Pax-1* mutation. *Cell* **66**:873-884.
9. Chisaka, O., and M. R. Capecchi. 1991. Regionally restricted developmental defects resulting from targeted disruption of the mouse homeobox gene *Hox-1.5*. *Nature (London)* **350**:473-479.
10. Davis, C. A., and A. L. Joyner. 1988. Expression patterns of the homeobox-containing genes *en-1* and *en-2* and the proto-oncogene *int-1* diverge during mouse development. *Genes Dev.* **2**:1736-1744.
11. Desplan, C., J. Theis, and P. H. O'Farrell. 1988. The sequence specificity of homeodomain-DNA interaction. *Cell* **54**:1081-1090.
12. Dessain, S., C. T. Gross, M. A. Kuziora, and W. McGinnis. 1992. Anti-type homeodomains have distinct DNA binding specificities that correlate with their different regulatory functions in embryos. *EMBO J.* **11**:991-1002.
13. Dolle, P., J.-C. Izpisua-Belmonte, H. Falkenstein, A. Renucci, and D. Duboule. 1989. Coordinate expression of the murine *hox-5* complex homeobox-containing genes during limb pattern formation. *Nature (London)* **342**:767-772.
- 13a. Ekker, S. C., D. P. von Kessler, and P. A. Beachy. 1992. Differential sequence recognition is a determinant of specificity in homeotic gene action. *EMBO J.* **11**:4059-4072.
14. Ekker, S. C., K. E. Young, D. P. von Kessler, and P. A. Beachy. 1991. Optimal DNA sequence recognition by the Ultrabithorax homeodomain of *Drosophila*. *EMBO J.* **10**:1179-1186.
15. Florence, B., R. Handrow, and A. Laughon. 1991. DNA-binding specificity of the fushi tarazu homeodomain. *Mol. Cell. Biol.* **11**:3613-3623.
16. Gehring, W. J. 1987. Homeo boxes in the study of development. *Science* **235**:1245-1252.
17. Gibson, G., A. Schier, P. LeMotte, and W. J. Gehring. 1990. The specificities of Sexcombs Reduced and Antennapedia are defined by a distinct portion of each protein that includes the homeodomain. *Cell* **62**:1087-1103.
18. Graham, A., N. Papalopulu, and R. Krumlauf. 1989. The murine and *Drosophila* homeobox gene complexes have common features of organization and expression. *Cell* **57**:367-378.
19. Grueneberg, D. A., S. Natesan, C. Alexandre, and M. Z. Gilman. 1992. Human and *Drosophila* homeodomain proteins that enhance the DNA-binding activity of serum response factor. *Science* **257**:1089-1095.
20. 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.
21. Hanes, S. D., and R. Brent. 1991. A genetic model for interaction of the homeodomain recognition helix with DNA. *Science* **251**:426-430.
22. Hill, R. E., P. F. Jones, A. R. Rees, C. M. Sime, M. J. Justice, N. G. Copeland, N. A. Jenkins, E. Graham, and D. R. Davidson. 1989. A new family of mouse homeo box-containing genes: molecular structure, chromosomal location, and developmental expression of *Hox-7.1*. *Genes Dev.* **3**:26-37.
23. Hoch, M., N. Gerwin, H. Taubert, and H. Jackle. 1992. Competition for overlapping sites in the regulatory region of the *Drosophila* gene *Kruppel*. *Science* **256**:94-97.
24. Hoey, T., and M. Levine. 1988. Divergent homeo box proteins recognize similar DNA sequences in *Drosophila*. *Nature (London)* **332**:858-861.
25. Holland, P. W. H., and B. L. M. Hogan. 1988. Expression of homeo box genes during mouse development: a review. *Genes Dev.* **2**:773-782.
26. Iler, N., and C. Abate. Unpublished data.
27. Janknecht, R., G. de Martynoff, J. Lou, R. A. Hipkind, A. Nordheim, and H. G. Stunnenberg. 1991. Rapid and efficient purification of native histidine-tagged protein expressed by recombinant vaccinia virus. *Proc. Natl. Acad. Sci. USA* **88**:8972-8976.
28. Joyner, A. L., and G. R. Martin. 1987. *en-1* and *en-2*, two mouse genes with sequence homology to the *Drosophila engrailed* gene: expression during embryogenesis. *Genes Dev.* **1**:29-38.
29. Kessel, M., and P. Gruss. 1990. Murine developmental control genes. *Science* **249**:374-379.
30. 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 2.8 Å resolution: a framework for understanding homeodomain-DNA interactions. *Cell* **63**:579-590.
31. 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.
32. Laughon, A. 1991. DNA binding specificity of homeodomains. *Biochemistry* **30**:11357-11367.
33. Luffin, T., A. Dierich, M. LeMeur, M. Mark, and P. Chambon. 1991. Disruption of the *hox-1.6* homeobox gene results in defects in a region corresponding to its rostral domain of expression. *Cell* **66**:1105-1119.
34. Mackenzie, A., G. L. Leeming, A. K. Jowett, M. W. J. Ferguson, and P. T. Sharpe. 1991. The homeobox gene *hox-7.1* has specific regional and temporal expression patterns during early murine craniofacial embryogenesis, especially tooth development *in vivo* and *in vitro*. *Development* **111**:269-285.
35. Mann, R. S., and D. S. Hogness. 1990. Functional dissection of Ultrabithorax proteins in *D. melanogaster*. *Cell* **60**:597-610.
36. McGinnis, W., R. L. Garber, J. Wirz, A. Kuroiwa, and W. J. Gehring. 1984. A homologous protein-coding sequence in *Dros-*

- ophila* homeotic genes and its conservation in other metazoans. *Cell* 37:403–408.
37. McGinnis, W., and R. Krumlauf. 1992. Homeobox genes and axial patterning. *Cell* 68:283–302.
 38. Mendel, D. B., P. A. Khavari, P. B. Conley, M. K. Graves, L. P. Hansen, A. Admon, and G. R. Crabtree. 1991. Characterization of a cofactor that regulates dimerization of a mammalian homeodomain protein. *Science* 254:1762–1767.
 39. Odenwald, W. F., J. Garbern, H. Arnheiter, E. Tournier-Lasserre, and R. A. Lazzarini. 1989. The Hox-1.3 homeo box protein is a sequence-specific DNA-binding phosphoprotein. *Genes Dev.* 3:158–172.
 40. Oliphant, A. R., C. J. Brandl, and K. Struhl. 1989. Defining the sequence specificity of DNA-binding proteins by selecting binding sites from random-sequence oligonucleotides: analysis of yeast GCN4 protein. *Mol. Cell. Biol.* 9:2944–2949.
 41. 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.
 42. Patwardhan, S., A. Gashler, M. G. Siegel, L. C. Chang, L. G. Joseph, T. B. Shows, and M. M. Lebeau. 1991. *egr3*, a novel member of the *egr* family of genes encoding immediate-early transcription factors. *Oncogene* 6:917–928.
 43. Rauscher, F. J., III, J. F. Morris, O. E. Tournay, D. M. Cook, and T. Curran. 1990. Binding of the Wilms' tumor locus zinc finger protein to the EGR-1 consensus sequence. *Science* 250:1259–1262.
 44. Regulski, M., S. Dessain, N. McGinnis, and W. McGinnis. 1991. High-affinity binding sites for the Deformed protein are required for the function of an autoregulatory enhancer of the *deformed* gene. *Genes Dev.* 5:278–286.
 45. Robert, B., D. Sassoon, B. Jacq, W. Gehring, and M. Buckingham. 1989. *hox-7*, a mouse homeobox gene with a novel pattern of expression during embryogenesis. *EMBO J.* 8:91–100.
 46. Rosenfeld, M. G. 1991. POU-domain transcription factors: Powerful developmental regulators. *Genes Dev.* 5:897–907.
 47. 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.
 48. Treisman, J., P. Gonczy, 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.
 49. Treisman, J., E. Harris, and C. Desplan. 1991. The *paired* box encodes a second DNA-binding domain in the Paired homeo domain protein. *Genes Dev.* 5:594–604.
 50. Wilkinson, D. G., S. Bhatt, M. Cook, E. Bonicelli, and R. Krumlauf. 1989. Segmental expression of *hox-2* homeobox-containing genes in the developing mouse hindbrain. *Nature (London)* 341:405–409.
 51. Wright, C. V. E., K. W. Y. Cho, G. Oliver, and E. M. DeRobertis. 1989. Vertebrate homeodomain proteins: families of region-specific transcription factors. *Trends Biochem. Sci.* 14:52–56.