Multiple Amino Acids Determine the DNA Binding Specificity of the Msx-1 Homeodomain†

Veronica Ebu Isaac,‡,§ Peter Sciavolino,‡,|| and Cory Abate*,‡,||

Center for Advanced Biotechnology and Medicine, Piscataway, New Jersey, Department of Biology, Rutgers University, New Brunswick, New Jersey, and Department of Neuroscience and Cell Biology, UMDNJ-Robert Wood Johnson Medical School, Piscataway, New Jersey

Received January 23, 1995; Revised Manuscript Received March 29, 1995⊗

ABSTRACT: This study investigates the sequence features that contribute to the differential DNA binding properties of two divergent homeodomains, Msx-1 and HoxA3. We show that these homeodomains have overlapping, but nonidentical, DNA binding site preferences. We defined the amino acid residues that contribute to the observed differences in DNA binding specificity by producing a series of mutated polypeptides in which selected residues in Msx-1 were replaced with the corresponding ones in HoxA3. These analyses show that the DNA binding specificity of Msx-1 versus HoxA3 results from the cumulative action of multiple residues in all segments of the homeodomain (i.e., the N-terminal arm and helices I, II, and III). Therefore, substitutions of residues in both helix III and the N-terminal arm (but not in either segment alone) produced an Msx-1 polypeptide whose binding site preference was indistinguishable from that of HoxA3. Residues in helices I and II also influence DNA binding activity; these oppositely charged residues (e.g., lysine 19 and glutamate 30) may mediate ionic interactions between helices I and II which stabilize DNA binding by Msx-1. These findings demonstrate a critical interplay between residues in each homeodomain segment for appropriate conformation of the protein−DNA complex.

Selective gene expression in eukaryotic cells is achieved by the action of transcriptional regulatory proteins which function by interacting with sequence-specific DNA elements in the transcriptional control regions of their respective target genes. A hallmark of transcription factors is that they utilize a limited repertoire of conserved domains to interact with DNA; thus, many such proteins are encoded by members of related gene families which share conserved sequences corresponding to these domains (Mitchell & Tjian, 1989). Although the various family members are likely to exhibit related DNA binding specificities *in vitro*, typically they have highly specific functions *in vivo* that are mediated by selective interactions with DNA elements of specific target genes. The homeodomain provides an excellent example of a conserved DNA binding domain and exhibits limited DNA binding specificity *in vitro* that also contributes to target gene selection *in vivo* (Kuziora & McGinnis, 1989; Gibson *et al.*, 1990; Mann & Hogness, 1990; Lin & McGinnis, 1992; Zeng *et al.*, 1993). The homeodomain is found in numerous proteins that regulate gene expression during embryogenesis, and it has been highly conserved throughout evolution with respect to primary sequence, tertiary structure, and mode of interaction with DNA (Gehring, 1987; Kessel & Gruss, 1990; McGinnis & Krumlauf, 1992; Scott *et al.*, 1989; Laughon, 1991). The sequence corresponds to 60 amino acid residues that comprise three α-helices and an extended N-terminal arm (Figure 1). A subset of these residues are invariant among all homeodomains and constitute a consensus sequence (Figure 1A) that is critical for appropriate structure and for mediating DNA binding activity (Scott *et al.*, 1989; Laughon, 1991). Accordingly, the structures of many homeodomain proteins are highly conserved (Kissinger *et al.*, 1990; Billeter *et al.*, 1993), as is their interaction with DNA sites that share the common nucleotide sequence TAAT (Scott *et al.*, 1989; Laughon, 1991).

Despite its many conserved features, the homeodomain is also a major determinant of functional specificity *in vivo*. Therefore, numerous homeodomain proteins are expressed during development, and (of those examined) their characteristic functions are directed by specific residues within their homeodomains (Kuziora & McGinnis, 1989; Gibson *et al.*, 1990; Mann & Hogness, 1990; Lin & McGinnis, 1992; Zeng *et al.*, 1993). It is presumed that these residues mediate target gene selection, through a combination of selective protein−protein and protein−DNA interactions. Therefore, while all homeodomains bind to DNA sites that contain a common TAAT core, divergent homeodomains exhibit subtle, but distinct, preferences for DNA sites that differ at the nucleotides flanking this core (Treisman *et al.*, 1989; Ekker *et al.*, 1991, 1992; Dessain *et al.*, 1992; Corsetti *et al.*, 1992; Catron *et al.*, 1993; Pellerin *et al.*, 1994). Moreover, it has been shown that DNA binding specificity *in vitro* is correlated with functional specificity *in vivo* (Dessain *et al.*, 1992; Ekker et al., 1992), which indicates that their differential binding properties may provide a significant contribution to the selective functions of homeodomain-containing proteins.

We have compared the DNA binding properties of two divergent murine homeodomain proteins, Msx-1 and HoxA3

† This work was supported by NIH Grant HD29226 03 to C.A.

* To whom correspondence should be addressed: Center for Advanced Biotechnology and Medicine, 679 Hoes Lane, Piscataway, NJ 08854-5638. (908) 235-5161. Fax: (908) 235-4850.

‡ Center for Advanced Biotechnology and Medicine.

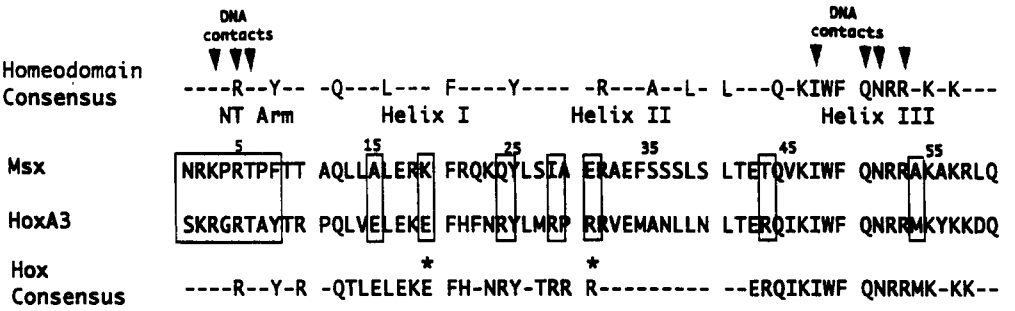
§ Rutgers University.

|| UMDNJ-Robert Wood Johnson Medical School.

⊗ Abstract published in *Advance ACS Abstracts,* May 15, 1995.

0006-2960/95/0434-7127$09.00/0

A



B

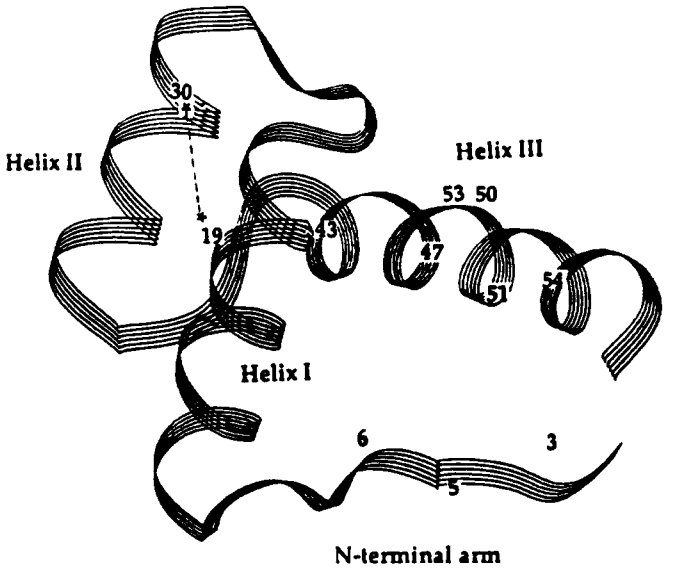


FIGURE 1: Comparison of Msx with the homeodomain and Hox consensus sequences and modeling of the Msx-1 homeodomain. (A) Homeodomain consensus sequence (Scott *et al.*, 1989) indicating residues that contact DNA and the positions of the N-terminal arm (NT arm) and helices I, II, and III. Also shown is the position of the homeodomain (numerically) and the sequences of Msx, HoxA3, and the Hox consensus. The boxed residues are those that were replaced in the mutated polypeptides (see Table 1). Asterisks (*) indicate residues 19 and 30 that are predicted by our mutagenesis studies and homology modeling to interact ionically. (B) Model of the structure of the Msx-1 homeodomain [described in Shang *et al.* (1994a)] based on the crystallographic structure of the engrailed–DNA complex. The homeodomain is shown as a ribbon representation indicating the residues in helix III (47, 50, 51, 53) and the N-terminal arm (3, 5, 6) that contact DNA; note that residues 43 and 54 lie on the same face of helix III as the above mentioned residues. Also shown are residues 19 and 30, which are predicted to interact ionically as indicated by the dashed line.

(Catron *et al.*, 1993). Msx-1 [formerly Hox 7.1 (Scott, 1992)] belongs to a relatively small family, comprising three members whose homeodomain sequences are nearly identical (Holland, 1991). HoxA3 [formerly Hox 1.5 (Scott, 1992)] belongs to a relatively large family comprising 38 members (McGinnis & Krumlauf, 1992) that share a high degree of sequence identity within their homeodomains, including certain residues that are conserved among all Hox proteins (e.g., Hox consensus, Figure 1A). Previously we showed that Msx and HoxA3 exhibit differential DNA binding site preferences *in vitro*. In this study we define the amino acid residues that contribute to these differences, and we show that the DNA binding specificity of Msx-1 results from the cumulative action of multiple residues distributed in each segment of the homeodomain.

EXPERIMENTAL PROCEDURES

Bacterial expression plasmids encoding the homeodomains of Msx-1 (Msx) and HoxA3 have been described (Catron *et al.*, 1993). Mutations were introduced into the Msx bacterial expression plasmid [p*msx-1*(157–233)] by PCR mutagenesis using overlapping oligonucleotides that contained the appropriate nucleotide substitutions. Each of the mutations was verified by DNA sequence analysis. Table 1 lists the specific amino acid substitutions in each of the mutated Msx polypeptides. The 5′ and 3′ oligonucleotides used for PCR mutagenesis also contained *Bam*HI and *Hin*dIII restriction sites to facilitate cloning into the bacterial expression plasmid pDS56 [as in Catron *et al.* (1993) and Pellerin *et al.* (1994)]. This plasmid contains six histidine codons, and the Msx polypeptides were expressed as hexahistidine fusion proteins to facilitate purification by nickel affinity chromatography [as in Catron *et al.* (1993) and Pellerin *et al.* (1993)]. Proteins were purified in the presence of 6 M guanidine-HCl (GuHCl) and renatured by dialysis against 25 mM potassium phosphate, pH 7.5, 50 mM potassium chloride, 10% glycerol (v/v), and 1 mM dithiothreitol.

For *in vitro* transcription and translation, *Bam*HI–*Hin*dIII fragments containing sequences encoding the *msx* or *hoxA3* homeodomains (Catron et al., 1993) were cloned into plasmid pGEM-7zf (+) (Promega) which was engineered to delete the downstream *Bam*HI site in the polylinker. The *msx*- or *hoxA3*-pGEM plasmids were linearized at the *Hin*dIII site, and RNA was transcribed *in vitro* from the T7 promoter as

Table 1[a]

protein	
Msx	Msx-1(157−233)
Msx(III, 1)	Msx-1(157−233) (A220M)
Msx(III, 2)	Msx-1(157−233) (T209R)
Msx(III)	Msx-1(157−233) (T209R; A220M)
Msx(II)	Msx-1(157−233) (Q190R; I194R; E196R)
Msx(I)	Msx-1(157−233) (A181E; K185E)
Msx(NT)	Msx-1(157−233) (N166S; R167K; K168R; P169G; P172A; F173Y)
Msx(I + II)	Msx-1(157−233) (A181E; K185E; Q190R; I194R; E196R)
Msx(II + III)	Msx-1(157−233) (Q190R; I194R; E196R; T209R; A220M)
Msx(I + III)	Msx-1(157−233) (A181E; K185E; T209R; A220M)
Msx(I + II + III)	Msx-1(157−233) (A181E; K185E; Q190R; I194R; E196R; T209R; A220M)
Msx(III + NT)	Msx-1(157−233) (T209R; A220M; N166S; R167K; K168R; P169G; P172A; F173Y)
Msx(II + NT)	Msx-1(157−233) (Q190R; I194R; E196R; N166S; R167K; K168R; P169G; P172A; F173Y)
Msx(I + II + NT)	Msx-1(157−233) (A181E; K185E; Q190R; I194R; E196R; N166S; R167K; K168R; P169G; P172A; F173Y)
Msx(I + II + III + NT)	Msx-1(157−233) (A181E; K185E; Q190R; I194R; E196R; T209R; A220M; N166S; R167K; K168R; P169G; P172A; F173Y)

[a] Msx refers to the homeodomain of Msx-1 (amino acids 157−233). The roman numerals refer to the helices containing the indicated amino acid substitutions; NT refers to the residues in the N-terminal arm. Specific substitutions in each polypeptide are shown in parentheses. The numbers of amino acids refer to the context of the full-length protein (Hill *et al.*, 1989).

per the manufacturer (Promega). Proteins were translated *in vitro* using a rabbit reticulocyte lysate system (Promega). In some experiments, Msx also contained an epitope from the myc protein which did not affect DNA binding specificity or affinity. The representative assay shown in Figure 2 used the epitope-tagged version of Msx; identical results were obtained with the Msx protein that had not been epitope tagged.

Electrophoretic mobility shift assays were performed as in Catron *et al.* (1993). Briefly, oligonucleotides containing the Msx binding sites were radiolabeled using T4 polynucleotide kinase in the presence of [γ-P^{32}]ATP. Oligonucleotides were annealed at equimolar concentrations for 1 h at 37 °C. Binding reaction mixtures contained 10 mM Tris-HCl, pH 7.5, 50 mM sodium chloride, 1 mM EDTA, 7.5 mM $MgCl_2$, 5% (v/v) glycerol, 5% (w/v) sucrose, 0.1% NP40, 0.5 mg/mL bovine serum albumin, 10 mM dithiothreitol, and 0.025 mg/mL dIdC (Pharmacia). Conditions for the *in vitro* translated proteins were identical to those described for the bacterial proteins except for the omission of bovine serum albumin in the reaction mixtures. Protein−DNA complexes were formed at room temperature for 10 min and resolved by polyacrylamide gel electrophoresis at 4 °C. Where indicated, DNA binding activity was quantitated using a PhosphorImager (Molecular Dynamics) and was calculated as the percentage of bound DNA divided by total DNA [bound/(bound + free)]. The DNA binding activity of Msx to site TAATTG was in the linear range (30−50%); this value was designated as 1. Values of other proteins were expressed relative to this value by using the following formula: [bound/(bound + free)] (mutated Msx)/[bound/(bound + free)](Msx). To account for the variability observed in DNA binding activity (see, e.g., Figure 6B), experiments were repeated a minimum of five times and DNA binding activity was calculated as an average of these mutliple assays.

Circular dichroism (CD) analysis was performed using an Aviv Model 6LD spectrophotometer at 25 °C in buffer containing 10 mM potassium phosphate, pH 7.0. Spectra were recorded in a 1-mm jacketed cuvette, and data were collected in triplicate from 260 to 198 nm with a step size of 0.25 nm. Protein concentrations were determined by absorbance at 280 nm (ϵ_{280} = 7000 cm^{-1} M^{-1}) in the presence of 6 M GuHCl. CD experiments were repeated a minimum of three times.

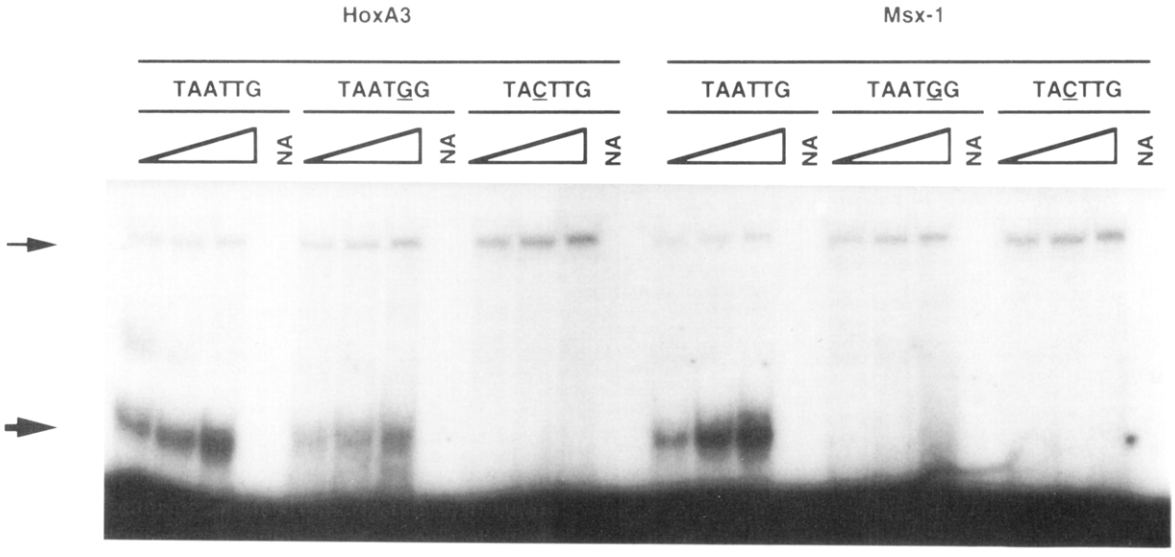


FIGURE 2: DNA binding properties of *in vitro* translated Msx and HoxA3. Electrophoretic mobility shift assays were performed using the consensus DNA site (TAATTG), a 3′-substituted DNA site (TAAT*G*G), or a TAAT core substituted DNA site (TA*C*TTG) (Catron *et al.*, 1993). Binding assays contained 1, 2, or 3 μL of the *in vitro* translated HoxA3 or Msx (as indicated by the triangles above the lanes). Protein−DNA complexes were formed at room temperature in DNA binding buffer, resolved from free DNA by non-denaturing gel electrophoresis, and visualized by autoradiography. The boldface arrow denotes the homeodomain−DNA complex; the upper arrow indicates a nonspecific protein−DNA complex formed in the reticulocyte lysates. NA indicates no added protein. Binding assays were performed a minimum of three times; a representative assay is shown.

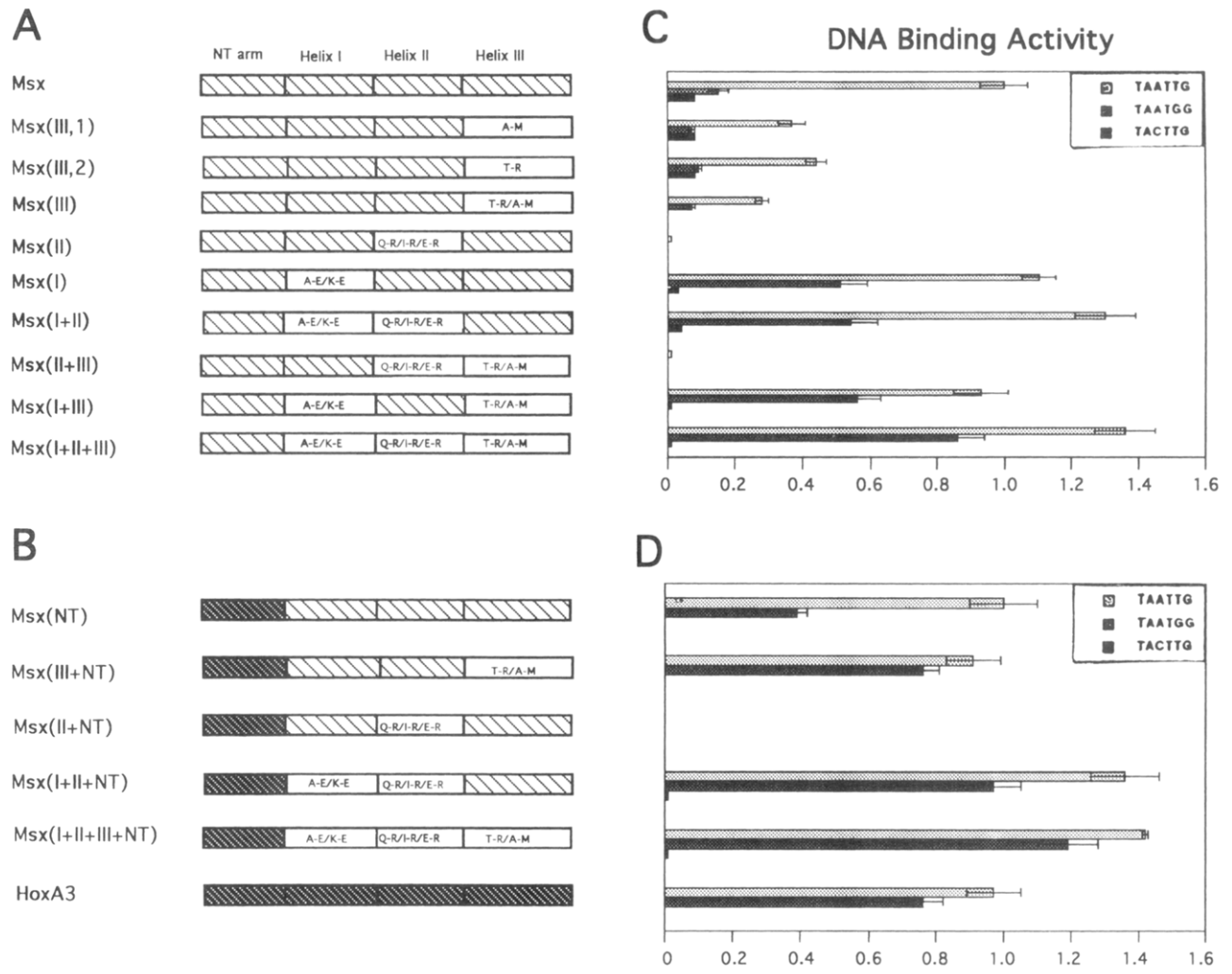


FIGURE 3: Design of the Msx mutated polypeptides. (A and B) Msx is shown by the light hatched boxes; HoxA3 is shown by the dark hatched boxes. The relative positions of the NT arm and helices I, II, and III are indicated. Amino acid substitutions in helices I, II, and III are shown using the single-letter amino acid codes. Table 1 lists the amino acid substitutions in the mutated Msx polypeptides. (C and D) DNA binding activity was assessed by electrophoretic mobility shift assays performed (as in Figure 6) using the DNA sites shown. Activity was quantitated as described in Experimental Procedures. To compare directly the binding activities of each protein, binding by Msx to the consensus DNA site (TAATTG) was designated 1.0, and the activities of the mutated polypeptides were expressed relative to this value as indicated by the bars. Data represent the average of five independent experiments using 0.06 μM protein (as in Figure 6), and the standard errors are shown.

RESULTS AND DISCUSSION

Comparison of the Homeodomain Sequences of Msx and HoxA3. In a previous study, we compared the DNA binding properties of Msx and HoxA3 using proteins obtained by *Escherichia coli* expression, and we showed that they bound with similar apparent affinity to a common consensus DNA site (e.g., TAATTG) (Catron *et al.*, 1993). However, HoxA3, but not Msx, also bound efficiently with DNA sites that differ at the nucleotide position 3′ of the TAAT core (e.g., TAAT*GG*), and neither interacted with DNA sites that contained substitutions within the TAAT core (e.g., TA*C*T-TG) (Catron *et al.*, 1993). We extended this observation to show that the binding specificity of HoxA3 was similar to other members of the Hox family with respect to preferences for 3′-substituted DNA sites (Pellerin *et al.*, 1994). As shown in Figure 2, these differences in DNA binding site preference were greatly enhanced by using Msx and HoxA3 proteins obtained by *in vitro* transcription/translation; this is presumably due to the limiting concentrations of HoxA3 and Msx in the rabbit reticulate lysate system. In fact, under these conditions both Msx and HoxA3 interact with the same apparent affinity with the consensus site (TAATTG); however, only HoxA3 interacts with the 3′-substituted site (e.g., TAAT*GG*).

To investigate the basis for these observed differences, we sought to identify amino acid residues that contribute to the differential DNA binding specificity of Msx and HoxA3 (Catron *et al.*, 1993). Since these proteins share limited sequence identity (i.e., 50%) within their homeodomains (Figure 1A; Catron *et al.* 1993), we selected as potentially critical residues several that are conserved within the Hox family and that differ in Msx (Figure 1A). The rationale for choosing these Hox consensus residues is based on our observation that the ability of HoxA3 to interact with 3′-substituted sites is a general feature of Hox homeodomains (Pellerin *et al.*, 1994). As shown in Figure 1A, residues that distinguish Msx and the Hox consensus are contained within each segment of the homeodomain, namely, the N-terminal arm and helices I, II, and III. Of particular interest are two residues at positions 43 and 54 in helix III, the DNA recognition helix (Kissinger *et al.*, 1990; Laughon, 1991; Billeter *et al.*, 1993). Molecular modeling predicts that these residues lie on the same face of helix III as those that directly contact DNA in the major groove (Figure 1B; Shang *et al.*,

1994b; H. Li and G. Montelione, personal communication). This suggests that they are good candidates for determining the DNA binding specificity of Msx. Residues in the N-terminal arm contact DNA in the minor groove (Kissinger *et al.*, 1990; Billeter *et al.*, 1993). Although the residues in the N-terminal arm are not well conserved (Figure 1A), they are notable since N-terminal arm residues mediate functional specificity of certain homeodomains *in vivo* (Lin & McGinnis, 1992; Zeng *et al.*, 1993).

Helices I and II lie perpendicular to helix III and may influence the way in which helix III docks with the major groove (Kissinger *et al.*, 1990; Laughon, 1991; Billeter *et al.*, 1993). Several residues in helices I and II differ between Msx and the Hox consensus (Figure 1A). It is noteworthy that some of these residues in helix I are oppositely charged from those in helix II (Figure 1A), suggesting the potential for ionic interactions between these helices. For example, position 19 in helix I is occupied by a lysine (K) in Msx and a glutamic acid (E) in the Hox consensus (Figure 1A). Conversely, position 30 in helix II is occupied by a glutamic acid (E) in Msx and an arginine (R) in the Hox consensus (Figure 1A).

Design and Characterization of Mutated Msx Polypeptides. To investigate the contribution of residues which distinguish Msx and HoxA3 (Figure 1A), we constructed a series of mutated polypeptides that contained substitutions of residues in Msx with the corresponding ones from HoxA3 (Figure 3A,B). A majority of residues were selected for this analysis, although those corresponding to conservative substitutions were not changed (e.g., positions 18 and 23, Figure 1A). The mutated polypeptides contained substitutions within the individual homeodomain segments [Figure 3A,B; e.g., Msx-(III), Msx(II), Msx(I), and Msx(NT)] or within combinations of these segments [Figure 3A,B; e.g., Msx(I + II), Msx(II + III), Msx(I + II + III), Msx(III + NT), Msx(I + II + NT), Msx(I + II + III + NT)]. To facilitate rapid and comparative analyses, polypeptides were expressed in *E. coli* and were purified from bacterial cell lysates to near homogeneity as evidenced by Coomasie blue staining of an SDS−polyacrylamide gel (Figure 4). Each mutated polypeptide was expressed at similar levels in *E. coli* and had similar solubilities subsequent to renaturation, which indicated that the substitutions did not adversely affect protein expression or stability.

To confirm that the mutated polypeptides were appropriately folded, CD analysis was performed (Figure 5). The CD spectra indicated that each polypeptide (with the exceptions noted below) formed α-helical structures similar to Msx (Figure 5A,C,D). Moreover, their helical content, as determined by the nonconstrained method of Greenfield and Fasman (1969), was in the range of 70%, similar to Msx and to the predicted value based on the homology-modeled structure (67%) [Figure 1B and Shang *et al.* (1994b); H. Li and G. Montelione, personal communication]. Polypeptides that contained substitutions of residues in helix II alone [e.g., Msx(II)] or in combination with substitutions in the N-terminal arm or helix III [e.g., Msx(II + NT), Msx(II + III)] did not form α-helical structures; rather, their CD spectra were reminiscent of a denatured or aggregated protein (Figure 5B). In contrast, polypeptides that contained substitutions of residues in helix II in combination with substitutions of residues in helix I had appropriate α-helical structure (Figure 5D). Therefore, these substitutions in helices I and II are complementary. Since these residues are oppositely charged

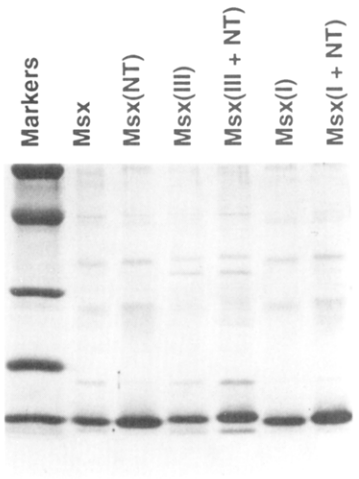


FIGURE 4: SDS−polyacrylamide gel Msx polypeptides (as in Figure 3; Table 1) were produced in *E. coli* and purified by nickel affinity chromatography (Catron *et al.*, 1993). Purified proteins (2.5 μg) were resolved on a 13.5% SDS−polyacrylamide gel and visualized by staining with Coomassie brilliant blue. Markers (from top to bottom) correspond to the molecular weight standards (Bio-Rad): bovine serum albumin, 68 kDa; ovalbumin, 46 kDa; carbonic anhydrase, 31 kDa; soybean trypsin inhibitor, 20 kDa; and lysozyme, 14 kDa.

(Figures 1A and 3A), their complementarity is likely due to ionic interactions between helices I and II which stabilize the structure of the homeodomain. In fact, inspection of the homology-modeled structure of Msx indicates that residues 19 (helix I) and 30 (helix II) are in close enough proximity (~4.5 Å) to form a salt bridge (H. Li and G. Montelione, personal communication). It is noteworthy that all classes of homeodomains contain oppositely charged residues at positions 19 and 30 (Scott *et al.*, 1989).

Analysis of the DNA Binding Site Preference of Mutated Msx Polypeptides. DNA binding analyses were performed using the three DNA sites that were previously tested to compare the binding properties of Msx and HoxA3 (Figure 2). To facilitate reliable comparison, DNA sites were radiolabeled in parallel to equivalent specific activity, and equimolar concentrations of each site were included in the binding assays. Electrophoretic mobility shift assays were performed using a range of protein concentrations (0.06 or 0.25 μM) and two independent preparations of the purified Msx polypeptides. Our strategy to compare the relative binding specificities of the various Msx polypeptides was to measure their binding activity and to compare the relative activity of each protein to each DNA site with those of Msx. It should be noted that this analysis was not intended to provide a quantitative comparison of DNA affinities but rather a comparison of relative binding activities. The results shown in Figure 3 represent the average of several experiments; Figure 6 shows a representative electrophoretic mobility shift assay.

As discussed above, it was anticipated that the residues in helix III that differ between Msx and the Hox consensus (e.g., residues 43 and 54) would be primary determinants for directing DNA binding specificity because of their position within the recognition helix (Figure 1B). Therefore, we examined the binding properties of Msx polypeptides that contained substitutions of the individual residues in helix III [e.g., Msx(III,1) or Msx(III,2)] or of both residues [e.g., Msx(III); Figures 3 and 6]. Surprisingly, these substitutions did not grossly alter DNA binding specificity since there was no increase in the DNA binding activity to site TAAT*GG*.

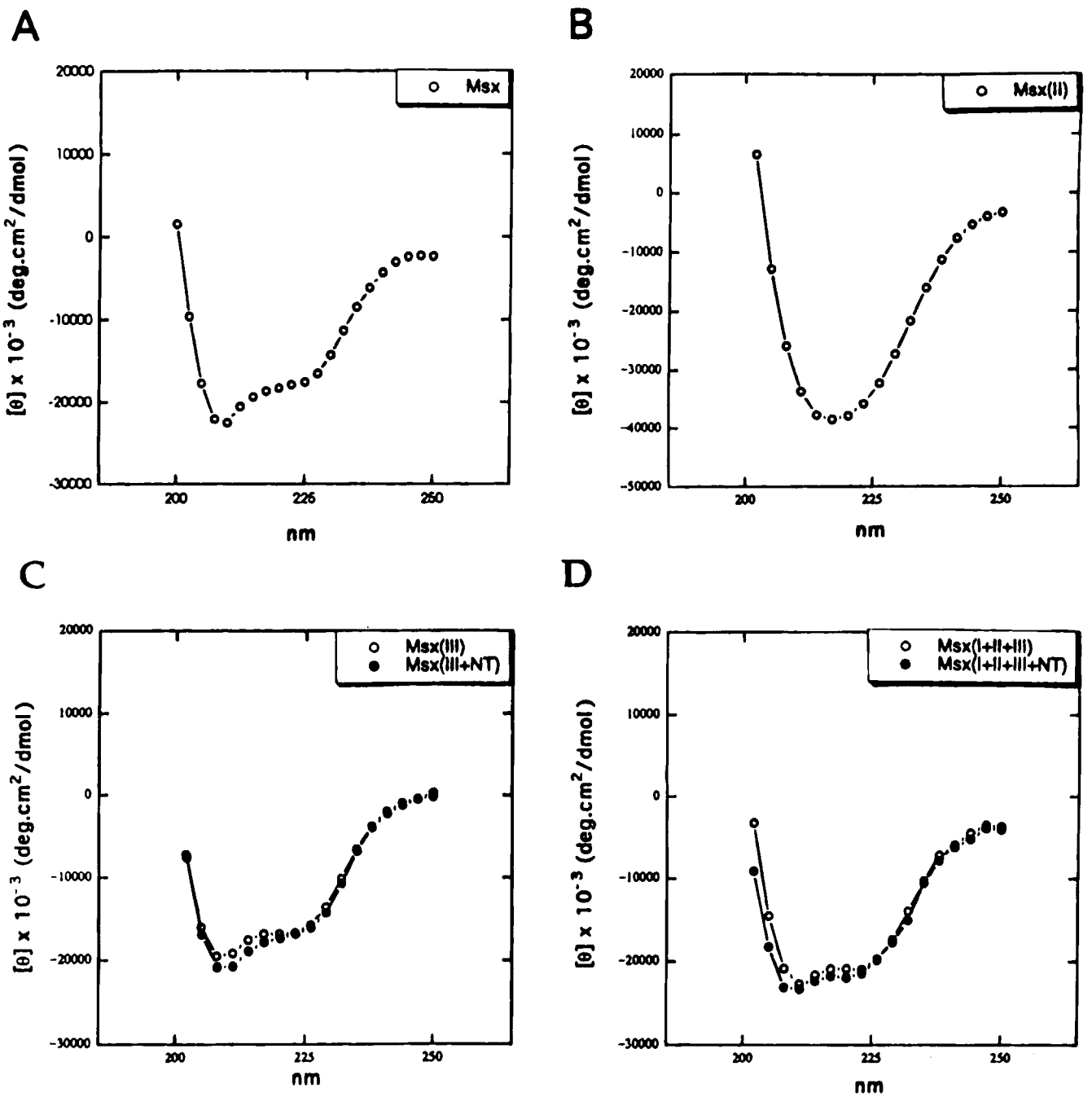


FIGURE 5: Circular dichroisim analysis (A–D) Far-UV CD spectra were measured for the indicated proteins (0.1 mg/mL) in buffer containing 10 mM potassium phosphate, pH 7.0.

In fact, binding activity was reduced by 3-fold relative to Msx (Figures 3 and 6). Their CD spectra show that Msx-(III) and Msx have similar α-helical contents (Figure 5A,C), and their thermal denaturation profiles are indistinguishable (V. Ebu Isaac and C. Abate, unpublished observations). Therefore, the reduced binding activity of Msx(III) relative to Msx is not due to a gross alteration in protein structure or stability.

In contrast to substitutions of residues in helix III alone, a protein that contained substitutions in the N-terminal arm combined with those in helix III [e.g., Msx(III + NT)] exhibited binding properties that were indistinguishable from Hox A3 (Figures 3 and 6). Therefore, Msx(III + NT) interacted with the 3′-substituted DNA site (TAATGG) with similar efficiency as HoxA3 (Figure 3). The apparent alteration in DNA binding specificity of Msx(III + NT) relative to Msx was due to substitutions of residues in both helix III and the N-terminal arm, since the binding properties of a protein that contained substitutions in only the N-terminal arm [e.g., Msx(NT)] were intermediate between Msx and HoxA3 (Figures 3 and 6). It is noteworthy that the CD spectra and thermal denaturation profiles of Msx(III + NT) are similar to those of Msx (Figure 5 and V. Ebu Isaac and C. Abate, unpublished observations), which further indicates that differences in binding specificities were not due to alterations in protein structure or stability. These findings identify a critical interplay between the residues in the N-terminal arm and helix III for directing DNA binding specificity. Therefore, it seems likely that residues in both the N-terminal arm and helix III need to be compatible to achieve efficient binding activity and appropriate binding specificity.

Although helices I and II do not make primary contacts with DNA, they are thought to stabilize the interaction of

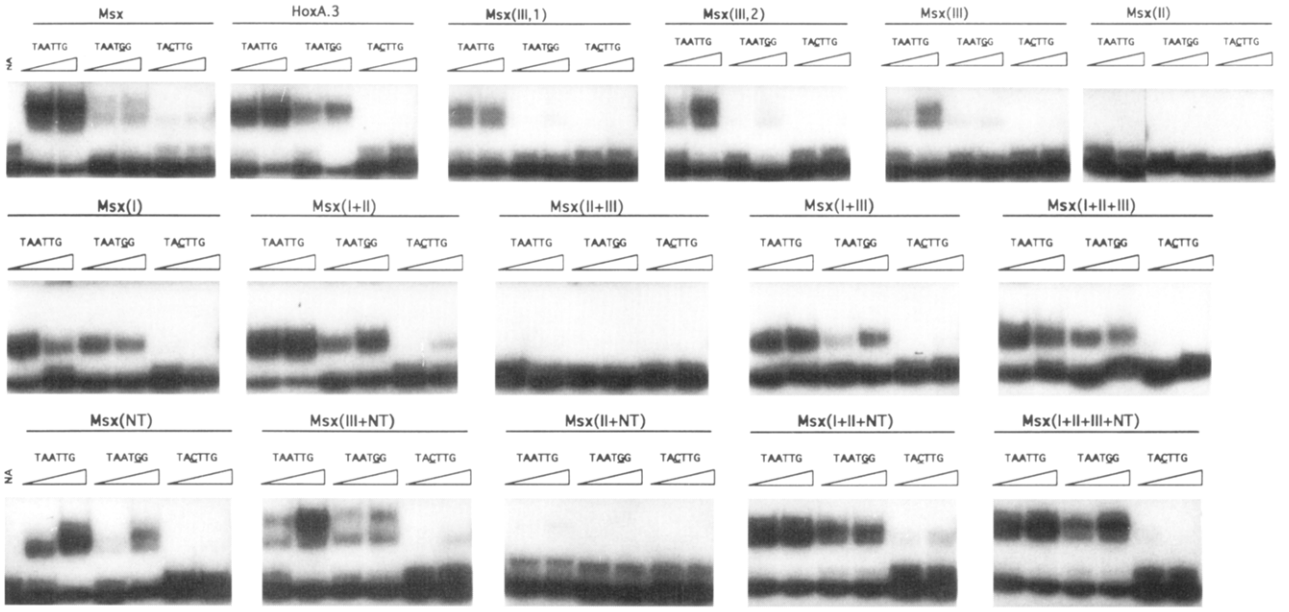


FIGURE 6: DNA binding analysis of the mutated Msx polypeptides. Electrophoretic mobility shift assays were performed as in Figure 2. Binding assays contained 0.06 or 0.25 μM protein (as indicated by the triangles above the lanes) and 5 nM DNA. Protein−DNA complexes were formed at room temperature in DNA binding buffer, resolved from free DNA by non-denaturing gel electrophoresis, and visualized by autoradiography. NA indicates no added protein. Binding assays were performed a minimum of five times; a representative assay is shown.

helix III with DNA and may affect the positioning of the homeodomain with respect to DNA (Kissinger *et al.*, 1990; Billeter *et al.*, 1993). The results obtained by comparing the binding properties of polypeptides containing substitutions within helices I or II alone [e.g., Msx(I) and Msx(II), respectively; Figure 3] or a combination of these [e.g. Msx-(I+II), Figure 3] are analogous to those obtained from the CD analyses. Specifically, Msx(II) [as well as Msx(II + III) and Msx (II + NT)], which lacks appropriate secondary structure (Figure 5B), also lacks DNA binding activity (Figures 3 and 6), whereas Msx(I) and Msx(I + II), which form α-helical structures (Figure 5D), exhibit binding activity (Figures 3 and 6). In fact, Msx(I + II) exhibits a modest (1.2-fold), but reproducible, increase in DNA binding activity relative to Msx (Figure 3). The finding that Msx(I + II) interacts with DNA further emphasizes the complementary nature of substitutions of the residues in helices I and II and further indicates that ionic interactions between residues in these helices may contribute to appropriate structure and DNA binding activity.

The preceding results show that residues in helix III combined with those in the N-terminal arm influence DNA binding specificity and that residues in helices I and II influence structure and binding activity. The implication of these findings is that all segments of the homeodomain contribute to the DNA binding properties of Msx. To test this idea, polypeptides were examined that contained substitutions of residues in various combinations of the homeodomain segments [e.g., Msx(I + II + III), Msx(I + II + NT), and Msx (I + II + III + NT)]. These polypeptides exhibited altered DNA binding specificity and enhanced binding activity relative to Msx (Figures 3 and 6). Therefore, Msx(I + II + III), Msx(I + II + NT), and Msx(I + II + III + NT) exhibited a 1.4-fold increase in binding activity relative to Msx for the consensus DNA site (TAATTG) (Figures 3 and 6). Moreover, each protein interacted with the 3′-substituted site (TAAT*GG*) with enhanced efficiency relative to Msx and bound with similar or greater efficiency than HoxA3 (Figures 3 and 6). In fact, a polypeptide that contained substitutions in all homeodomain segments [e.g., Msx(I + II + III + NT)] interacted most efficiently with this DNA site (Figures 3 and 6). Overall, the appropriate interaction of Msx with DNA is conferred by a combination of residues in each homeodomain segment.

CONCLUSIONS

The major finding of the present study is that the differential DNA binding specificities of two divergent homeodomains, Msx and HoxA3, result from the cumulative interaction of multiple residues in all segments of the homeodomain. Thus, residues in helix III are not sufficient, and additional residues in the N-terminal arm are required for appropriate DNA binding specificity. These observations, which highlight the role of the N-terminal residues in combination with residues in helix III, are provocative in light of numerous studies that have shown the importance of the N-terminal residues for directing functional specificity *in vivo* (Kuziora & McGinnis, 1989; Gibson *et al.*, 1990; Mann & Hogness, 1990; Lin & McGinnis, 1992; Zeng *et al.*, 1993). One possibility is that the N-terminal residues, which are divergent even among closely related homeodomains (Scott et al., 1989), allow for subtle differences in the conformation of the protein−DNA complex, which contributes to the ability of particular homeodomains to interact with distinct DNA sites. Moreover, these differences are likely to be amplified *in vivo* due to competition among related proteins, interactions with associated proteins, and limiting concentrations of other protein factors.

In addition to the residues that contact DNA, residues in helices I and II also contribute to appropriate binding activity through complementary ionic interactions which presumably affect the overall positioning of the homeodomain with DNA. Therefore, the conformation of the Msx−DNA complex is dictated by multiple residues in all segments of the homeodomain. This concept contrasts a previous notion which suggested that homeodomain binding specificity could be explained by a single residue (position 50) within helix III

which was thought to confer DNA binding specificity (Triesman *et al.*, 1989). It has become increasingly evident that the differential DNA binding properties of homeodomains cannot be explained by a single residue. In fact, both Msx and HoxA3 contain the same residue at position 50 (see Figure 1). Moreover, two Antennapedia-type homeodomains that share the same residue at position 50 have distinct DNA binding specificities which, in fact, correlate with their distinct functions *in vivo* (Dessain *et al.*, 1991). The observations presented herein suggest that it is more suitable to consider the homeodomain as an intact structural unit in which multiple segments contribute to the overall protein−DNA interaction. Indeed, this is the case for POU domains, which have been shown to be dynamic structures in which multiple segments contribute interdependently to DNA binding activity (Aurora & Herr, 1992).

ACKNOWLEDGMENT

The CD experiments were performed with the expert assistance of Norma Greenfield. We are indebted to G. Montelione and H. Li for pointing out the close proximity of residues 19 and 30 in Msx, for sharing results of molecular modeling prior to publication, and for many discussions. We are also grateful to Katrina Catron for engineering the pGEM vector. The homology modeling was performed with the assistance of Shahid Imran. We thank Norma Greenfield, Ann Stock, Nancy Iler, Katrina Catron, Hailan Zhang, and Isabelle Pellerin for critical reading of the manuscript, and Rita Sweeney for preparation. We would also like to thank two anonymous reviewers for their useful comments that helped in clarification of the manuscript.

REFERENCES

Aurora, R., & Herr, W. (1992) *Mol. Cell. Biol. 12*, 445−467.
Billeter, M., Qian, Y. Q., Otting, G., Müller, M., Gehring, W., & Wüthrich, K. (1993) *J. Mol. Biol. 234*, 1084−1097.
Catron, K. M., Iler, N., & Abate, C. (1993) *Mol. Cell. Biol. 13*, 2354−2365.
Corsetti, M. T., Briata, P., Sanseverino, L., Daga, A., Airoldi, I., Simeone, A., Palmisano, G., Angelini, C., Boncinelli, E., & Corte, G. (1992) *Nucleic Acids Res. 20*, 4465−4472.
Dessain, S., Gross, C. T., Kuziora, M. A., & McGinnis, W. (1992) *EMBO J. 11*, 991−1002.
Ekker, S. C., Young, K. E., von Kessler, D. P., & Beachy, P. A. (1991) *EMBO J. 10*, 1179−1186.
Ekker, S. C., von Kessler, D. P., & Beachy, P. A. (1992) *EMBO J. 11*, 4059−4072.
Florence, B., Handrow, R., & Laughon, A. (1991) *Mol. Cell. Biol. 11*, 3613−3623.
Gehring, W. J. (1987) *Science 235*, 1245−1252.
Gibson, G., Schier, A., LeMotte, P., & Gehring, W. J. (1990) *Cell 62*, 1087−1103.
Greenfield, N., & Fasman, G. D. (1969) *Biochemistry 8*, 4108−4115.
Hill, R. E., Jones, P. F., Rees, A. R., Sime, C. M., Justice, M. J., Copeland, N. G., Jenkins, N. A., Graham, E., & Davidson, D. (1989) *Genes Dev. 3*, 26−37.
Holland, P. W. H. (1991) *Gene 98*, 253−257.
Jain, J., McCaffrey, P. G., Miner. Z., Kerppola, T. K., Lambert, J. N., Verdine, G. L., Curran, T., & Rao, A. (1993) *Nature 365*, 352−355.
Kessel, M., & Gruss, P. (1990) *Science 249*, 374−379.
Kissinger, C. R., Liu, B., Martin-Blanco, E., Kornberg, T. B., & Pabo, C. O. (1990) *Cell 63*, 579−590.
Kuziora, M. A., & McGinnis, W. (1989) *Cell 59*, 563−571.
Laughon, A. (1991) *Biochemistry 30*, 11357−11367.
Lin, L., & McGinnis, W. (1992) *Genes Dev. 6*, 1071−1081.
Mann, R. S., & Hogness, D. S. (1990) *Cell 60*, 597−610.
McGinnis, W., & Krumlauf, R. (1992) *Cell 68*, 283−302.
Mitchell, P. J., & Tjian, R. (1989) *Science 245*, 371−378.
Pellerin, I., Schnabel, C., Catron, K. M., & Abate, C. (1994) *Mol. Cell. Biol. 14*, 4532−4545.
Percival-Smith, A., Müller, M., Affolter, M., & Gehring, W. J. (1990) *EMBO J. 9*, 3967−3974.
Scott, M. P. (1992) *Cell 71*, 551−553.
Scott, M. P., Tamkun, J. W., & Hartzell, G. W., III (1989) *Biochim. Biophys. Acta 989*, 25−48.
Shang, Z., Ebright, Y. W., Pendergrast, P. S., Echelard, Y., McMahon, A. P., Ebright, R. H., & Abate, C. (1994a) *Proc. Natl. Acad. Sci. U.S.A. 91*, 118−122.
Shang, Z., Ebu Isaac, V., Li, H., Patel, L., Catron, K. M., Curran, T., Montelione, G. T., & Abate, C. (1994b) *Proc. Natl. Acad. Sci. U.S.A. 91*, 8373−8377.
Treisman, J., Gönczy, P., Vashishtha, M., Harris, E., & Desplan, C. (1989) *Cell 59*, 553−562.
Zeng, W., Andrew, D. J., Mathies, L. D., Horner, M. A., & Scott, M. P. (1993) *Development 118*, 339−352.

BI950152K