898–906 Nucleic Acids Research, 1996, Vol. 24, No. 5

Hox homeodomain proteins exhibit selective complex stabilities with Pbx and DNA

Wei-Fang Shen, Ching-Pin Chang, Sofia Rozenfeld, Guy Sauvageau[2], R. Keith Humphries[2], Ming Lu[3], H. Jeffrey Lawrence, Michael L. Cleary[1] and Corey Largman*

Department of Medicine San Francisco Veterans Affairs Medical Center and University of California, San Francisco, CA 94121, USA, [1]Department of Pathology, Stanford University Medical Center, Stanford, CA 94305, USA, [2]Terry Fox Laboratory, British Columbia Cancer Agency and the Department of Medicine, University of British Columbia, Vancouver, Canada and [3]Cancer Center and Department of Medicine, University of California, San Diego, CA, USA

Received November 1, 1995; Revised and Accepted January 7, 1996

ABSTRACT

Eight of the nine homeobox genes of the *Hoxb* locus encode proteins which contain a conserved hexapeptide motif upstream from the homeodomain. All eight proteins (Hoxb-1–Hoxb-8) bind to a target oligonucleotide in the presence of Pbx1a under conditions where minimal or no binding is detected for the Hox or Pbx1a proteins alone. The stabilities of the Hox–Pbx1a–DNA complexes vary >100-fold, with the proteins from the middle of the locus (Hoxb-5 and Hoxb-6) forming very stable complexes, while Hoxb-4, Hoxb-7 and Hoxb-8 form complexes of intermediate stability and proteins at the 3′-side of the locus (Hoxb-1–Hoxb-3) form complexes which are very unstable. Although Hox-b proteins containing longer linker sequences between the hexapeptide and homeodomains formed unstable complexes, shortening the linker did not confer complex stability. Homeodomain swapping experiments revealed that this motif does not independently determine complex stability. Naturally occurring variations within the hexapeptides of specific Hox proteins also do not explain complex stability differences. However, two core amino acids (tryptophan and methionine) which are absolutely conserved within the hexapeptide domains appear to be required for complex formation. Removal of N- and C-terminal flanking regions did not influence complex stability and the members of paralog group 4 (Hoxa-4, b-4, c-4 and d-4), which share highly conserved hexapeptides, linkers and homeodomains but different flanking regions, form complexes of similar stability. These data suggest that the structural features of Hox proteins which determine Hox–Pbx1a–DNA complex stability reside within the precise structural relationships between the homeodomain, hexapeptide and linker regions.

INTRODUCTION

The *Hox* homeobox-containing genes appear to regulate regional specification along the A–P axis and in the limb bud of the developing embryo (1,2). The distinctive morphogenetic role of these genes has been most clearly demonstrated by the different phenotypic changes observed in mice carrying a range of targeted disruptions of different *Hox* genes (3–9). The homeodomain has been shown to be a DNA binding motif (reviewed in 10) and Hox proteins are thought to function as transcription factors (11,12). However, despite markedly divergent biological effects *in vivo*, the homeodomains of many different Hox proteins can bind *in vitro* to similar or identical DNA sequences containing a TAAT core recognition sequence with similar affinities (13,14). Indeed, the highest degree of sequence conservation between different Hox proteins is in helix three of the homeodomain (15), which appears to bind in the major groove of the target DNA and to contribute to binding specificity (16–18). These observations raise the question of how Hox proteins mediate selective gene regulation.

The 38 human or murine Hox homeodomain proteins, arrayed within four genetic loci, can be aligned in so-called paralog groups based on the degree of amino acid homology within the homeodomain (15,19). Outside the homeodomain there is variable sequence conservation among paralog members and relatively little homology between proteins across a single locus. An exception to this general rule is the presence of a hexapeptide motif, containing a relatively conserved YPWM core sequence and variably located 5–53 amino acids N-terminal of the homeodomain, which is shared by all of the Hox proteins in paralog groups 1–8 and is absent in groups 9–13 (15). Although X-ray crystallographic studies which have defined the binding of the homeodomain to DNA have not used proteins containing the hexapeptide (16,17), NMR studies have shown that the hexapeptide and linker sequences are unstructured and do not appear to make direct DNA binding contacts (18). Much of the data on DNA binding by Hox proteins has been obtained using only the 61 amino acid homeodomain (13,14). These experiments have shown that the highly variable N-terminal portion of the homeodomain can contribute incrementally to DNA binding

* To whom correspondence should be addressed at: 151H, VA Medical Center, 4150 Clement Street, San Francisco, CA 94121, USA

selectivity and affinity. However, there is little data on contributions to DNA binding by the flanking regions of the full-length Hox proteins. In one experiment the full-length Ubx-Ib protein and the corresponding homeodomain bound the same consensus DNA binding site, suggesting that under the conditions studied the hexapeptide motif did not directly influence DNA binding (13).

One proposed mechanism for generating binding specificity and increasing affinity is the formation of multi-protein–DNA complexes. The best example of homeodomain protein cooperativity is the interaction of yeast mating type proteins α2 and a1, for which both genetic and biochemical data have been obtained for cooperative binding to the haploid-specific gene operator (20,21). Following genetic data which indicated that the *Drosophila* Ubx homeodomain protein acts in concert with the Exd homeoprotein, several laboratories have demonstrated that these proteins form cooperative complexes with DNA *in vitro* (22,23). *Pbx1* is a non-Hox homeobox-containing proto-oncogene which was initially discovered at the site of t(1,13) chromosomal translocations in pre-B cell leukemias (24,25). The Pbx proteins share a high degree of sequence homology with the Exd and *Caenorhabditis elegans* ceh-20 proteins (26,27). We (28) and others (29,30) have recently shown that Hox homeodomain proteins cooperatively bind with Pbx proteins to DNA targets. In our studies three homeodomain proteins from the human *Hoxb* locus, Hoxb-4, b-6 and b-7, cooperatively bound with Pbx1a protein to an oligonucleotide containing Hox and Pbx DNA recognition sites under conditions where there was minimal DNA binding by either the full-length Hox or Pbx1a proteins alone. These studies also demonstrated that the conserved hexapeptide motif and the homeodomain were required for cooperative binding with Pbx1a. Introduction of the Hoxb-7 hexapeptide together with the 13 amino acid linker preceding the homeodomain into Hoxa-10 protein, which lacks this motif, substantially increased cooperative binding to Pbx1a/DNA.

In the current study the stabilities of complexes formed by the eight Hoxb proteins which contain a hexapeptide motif with Pbx1a and an oligonucleotide target have been determined. There is a >100-fold difference in dissociation rates across the locus, with the Hoxb-5 and Hoxb-6 proteins forming the most stable complexes with Pbx1a on an oligonucleotide containing consensus binding sites for both proteins. Data from mutagenesis and domain swapping studies suggest that within each Hox protein the hexapeptide, the linker sequence and the homeodomain function as a unit to mediate cooperative DNA binding.

MATERIALS AND METHODS

Protein expression

cDNAs encoding individual full-length Hox proteins, tcl-3 and Pbx1 were subcloned into either a derivative of sp65 (Promega, Madison, WI) which was engineered to contain a FLAG sequence (MDYKDDDDK) (Hoxb-2, b-3, b-4, b-6, b-7, tcl-3 and Pbx 1a), pSG5, which contains a T7 promoter (19) (Hoxb-1), Bluescript (Stratagene, La Jolla, CA), which contains a T7 promoter (Hoxb-5) or into a pET vector (Novagen, Madison, WI) containing the T7 promoter and His sequences (MGSSHHHHHH-SSGLVPRGSH-MASMTGGQQMGR) (Hoxa-4, b-8, c-4 and d-4). Full-length Hox proteins fused to the respective peptides were synthesized using the TNT coupled *in vitro* transcription/translation system (Promega, Madison, WI), in both the presence and absence of [^{35}S]methionine. The labeled proteins were electrophoresed separately to demonstrate synthesis of the appropriate full-length product and to ensure that similar amounts of proteins were used for complex formation. For each synthesis reaction a control containing reticulocyte lysate and the appropriate viral polymerase was used to detect possible DNA binding by endogenous factors. As shown in the figures, lysate controls show variable intensity gel shift bands with the DNA target. These bands varied with both the lysate batch and the batch of poly(dI·dC) used as non-competitive inhibitor. Human Hoxb-6 and b-7 have been cloned previously (39,40). Variant forms of Hoxb-3 were cloned from either a human CD34$^+$ bone marrow library or by PCR amplification of mRNA from the human erythroleukemia cell line K562. The other human full-length cDNA clones were kindly provided as follows: Hoxb-1 (15), b-2 (15), b-4 (41) and c-4 (42), Dr E. Boncinelli; Hoxb-5 (43), Dr C. Hauser. Full-length murine clones were provided as follows: Hoxa-4 (44), Dr D. Wolgemuth; Hoxb-8 (45), Dr S. Cory; Hoxd-4 (46), Dr G. Firestone.

DNA constructs

Specific amino acids within the YPWM sequence of Hoxb-4 were altered by insertion of synthetic double-stranded oligonucleotides containing the desired change and flanked by *Acc*I and *Pml*I sites, which were ligated into the *Acc*I (463) and *Pml*I (487) sites in the original clone. The distance between the YPWM motif and the homeodomain in Hoxb3 was reduced from 53 to either 17 amino acids, consisting of the first 12 and last five residues in the linker region (Table 2), or to a linker consisting of the first 12 residues by PCR amplification and conventional cloning methods. Replacement of the Hoxb-3 homeodomain with that of Hoxb-6 and the Hoxb-6 homeodomain with that of Hoxb-1 were accomplished using PCR to generate specific restriction sites allowing precise replacement of each homeodomain using conventional cloning methods. Each mutation was checked by DNA sequence analysis and mutant proteins were shown to have the correct size by gel electrophoresis of the *in vitro* translation products generated in the presence of [^{35}S]methionine.

Electrophoretic mobility shift assays

Complementary oligonucleotides which contained consensus binding sites for Hox proteins (18) and Pbx1 (47) were synthesized (Operon Technologies, Alameda, CA). The orientation of the Hox and Pbx sites to each other within the oligonucleotide sequence was selected on the basis of the orientation of sites in the yeast haploid-specific operator (48). This DNA fragment was previously used to demonstrate cooperative binding of Hoxb-4, b-6 and b-7 to DNA in the presence of Pbx1a (28). The standard conditions used were similar to those previously described. Briefly, double-stranded end-labeled DNA (50 000 c.p.m./binding reaction, ~1 ng) was incubated with 2 µl test Hox protein (<0.5 ng) either in the presence of Pbx1a (2 µl) or with 2 µl lysate control, in 75 mM NaCl, 1 mM EDTA, 1 mM dithiothreitol, 10 mM Tris–HCl, pH 7.5, 6% glycerol, 2 µg bovine serum albumin and 2 µg poly(dI·dC) as non-specific competitor, in a final reaction volume of 15 µl. In reactions performed under reduced stringency conditions the concentration of poly(dI·dC) non-specific inhibitor was reduced to 16 ng/15 µl reaction and the concentration of NaCl was decreased to 50 mM. Reaction mixtures were incubated on ice for 30 min and either applied directly to a 6% polyacrylamide gel (zero time sample) or mixed with a 100-fold excess of unlabeled oligonucleotide followed by incubation for fixed times (1–35 min) prior to application to the running polyacrylamide gel. Gel electrophoresis was performed in 0.25× TBE buffer as described previously. For the Hoxb-6/b-1 box

complex, which migrates to a position very close to non-specific gel shift bands attributed to lysate proteins, a supershift using antiserum to the FLAG-tag protein (IBI, New Haven, CT) was performed to confirm the identity of the complex (Fig. 4, lane 15).

Calculation of complex half-lives

Gels from electrophoretic mobility shift assays were dried and autoradiographed to permit densitometric estimation of complex band intensities by scanning autoradiographs into a MacIntosh Quadra computer, followed by quantitation of individual bands using the NIH Image software program. Each gel was autoradiographed for various times to ensure that the densities measured were within the linear range of the scanner and software program. A dissociation rate was calculated for each Hox–Pbx1a–DNA complex by plotting the log of the complex band intensities versus time. The dissociation rate was estimated as the slope of the regression line generated from this plot. For each dissociation experiment the correlation coefficient for the line was >0.96. For each complex the half-life was calculated using the equation $T_{½} = -\log(0.5)/k_d$.

RESULTS

Hoxb locus proteins containing the hexapeptide motif form cooperative DNA binding complexes with Pbx1a

We have previously shown cooperative binding by several Hoxb proteins with Pbx1a on an oligonucleotide (5′-CGAATTGATTGATGCACTAATTGGAG-3′) containing consensus binding sites for both proteins (28). We and others have shown that both the Hox and Pbx proteins bind to the TGATTGAT sequence within this or similar DNA fragments (28,29,31). The same oligonucleotide was now used to study cooperative binding of each of the full-length Hoxb proteins which contain a conserved hexapeptide motif. As shown in Figure 1, Hoxb proteins representing paralog groups 1–8 were capable of forming a cooperative DNA binding complex with Pbx1a. With the exception of weak complexes formed by Hoxb-4 (Fig. 1) and moderately strong complexes by Hoxb-3 (Fig. 4, lane 6), there was no evidence of interaction of the full-length Hox proteins with the DNA in the absence of Pbx1a under the conditions of our standard electrophoretic mobility shift assays (EMSA), which employ moderately high levels of non-specific competitor DNA as well as low concentrations of *in vitro* synthesized Hox proteins. In this regard, we have observed that in each case tested (Hoxb-4, b-6 and b-7) truncated proteins containing only the homeodomain were able to form complexes with the oligonucleotide under conditions where DNA binding by the full-length Hox proteins was not detected (results not shown).

Hoxb protein–Pbx1a protein–DNA complexes show varying stability

Although each of the Hoxb proteins tested was capable of cooperative DNA binding, we wished to quantify these interactions. Since dissociation rate constants are independent of both protein and DNA concentrations, we investigated the stability of the Hoxb–Pbx1a–DNA complexes. Preformed complexes containing radiolabeled DNA were incubated at 4°C with a large excess of cold oligonucleotide to prevent reformation of labeled complexes following dissociation. At timed intervals aliquots of the reaction mixtures were subjected to EMSA to measure the remaining labeled

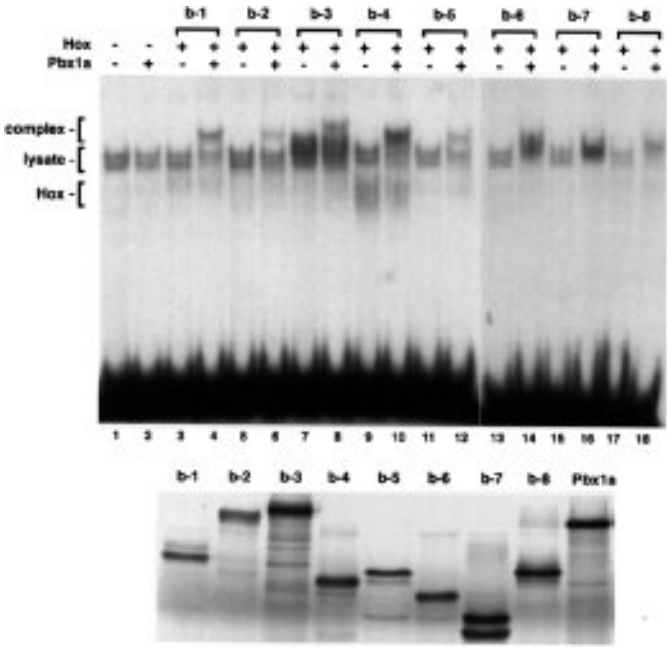

Figure 1. Hoxb proteins form cooperative DNA binding complexes with Pbx1a. Each of the Hoxb proteins formed a complex with a radiolabeled oligonucleotide containing consensus Hox and Pbx1a binding sites in the presence of Pbx1a. No specific complex was visualized in the absence of Pbx1a protein. However, non-specific bands of variable intensity, representing DNA binding by endogenous proteins in the reticulocyte lysate used to synthesize Hoxb and Pbx1a proteins, were observed in control lanes containing DNA and lysate alone (lane 1). Pbx1a protein did not bind to the DNA template in the absence of Hox proteins (lane 2). DNA–protein complexes were visualized by electromobility shift assay of the reaction mixtures. The *in vitro* translated ^{35}S-labeled proteins are shown below the gel shift lanes to demonstrate that approximately equal amounts of protein were used in each complex formation assay.

complex. Figure 2 shows the complete EMSA gels for three representative dissociation experiments, while Figure 3 shows the dissociation time course profiles of the Hox–Pbx1a–DNA complexes for each of the other Hoxb proteins. There is at least a 100-fold variation in the dissociation rates and corresponding half-lives of the Hoxb protein complexes with Pbx1a and the oligonucleotide (Table 1). Hoxb-6 and b-5 formed the most stable complexes, Hoxb-4, b-7 and b-8 showed intermediate complex stabilities, while Hoxb-1 dissociated much more rapidly. Hoxb-2 and Hoxb-3 dissociated so rapidly that it was impossible to visualize any complexes at the earliest time point following addition of competitor DNA. Dissociation rate constants were also determined at 22°C for Hox proteins representing stable and rapidly dissociating complexes (Hoxb-6 and b-1 respectively). These experiments revealed that although each complex was less stable at the higher temperature, the relative stabilities remained similar to those determined at 4°C. Thus the half-life of the complex formed between Hoxb-6 and Pbx1a on the DNA target decreased from 59 to 16 min, while that for Hoxb-1 decreased from 2.5 to 1.0 min (data not shown).

Spacing between the YPWM motif and the homeodomain does not regulate complex stability

To investigate the possible structural features responsible for the differences in complex stability we first examined the effects of the distance between the hexapeptide domain and the homeodomain.

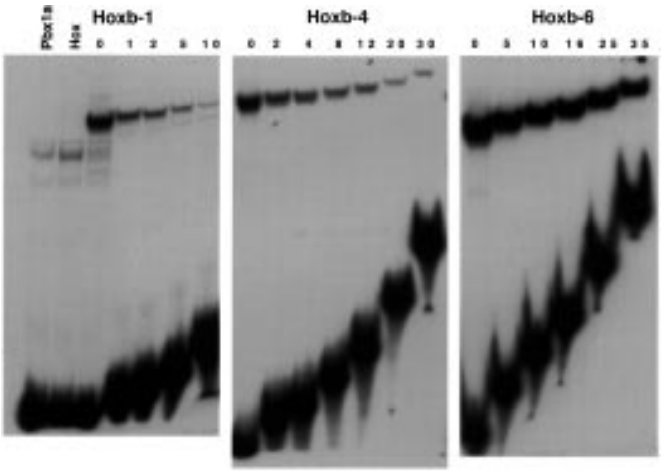

Figure 2. Hox proteins form complexes with Pbx1a and DNA of highly variable stability. Pre-formed Hox protein–Pbx1a protein–DNA complexes were either subjected to immediate gel electrophoresis (time 0) or incubated in the presence of a 100-fold excess of unlabeled DNA for the times indicated (min) prior to electrophoresis. For each Hox protein, samples lacking the Hox or Pbx1a protein appeared similar to those shown for Hoxb-1. Densitometric measurements of each Hox–Pbx1a–DNA band were used to calculate dissociation rate constants and apparent half-lives as described in Materials and Methods and reported in Table I.

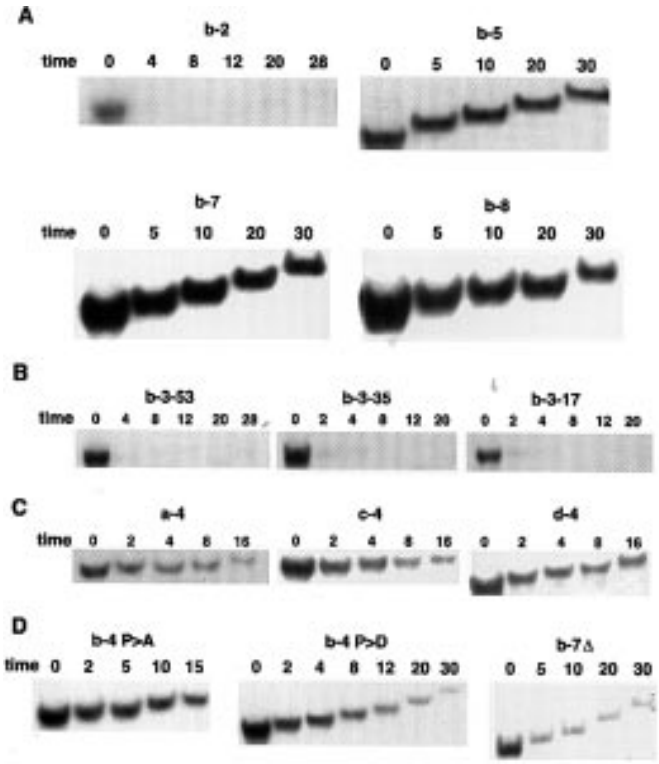

Figure 3. Dissociation of Hox–Pbx1a–DNA complexes. EMSA bands representing the dissociation time courses are shown for naturally occuring and mutant Hoxb protein complexes with Pbx1a and the target oligonucleotide. Dissociation rates calculated from these data and the derived half-lives are reported in Table I. For each dissociation experiment approximately equal amounts of Hox protein was used, as estimated from ^{35}S-labeled protein synthesis as reported in Figure 1. (**A**) Hoxb proteins show variable complex stability. Dissociation data are presented for the Hoxb proteins not shown in Figure 2 or in part (B) below. (**B**) Hoxb-3 proteins form unstable complexes with Pbx1a and DNA. Hoxb-3-53 and Hoxb-3-35 represent alternatively spliced transcripts, while Hoxb-3-17 is a mutant protein containing a truncated linker region. The two alternatively spliced proteins were indistinguishable, since they both form complexes (0 time) but dissociate too rapidly to measure a dissociation rate. The mutant Hoxb-3-17 protein, which contains a shortened linker, did not show increased complex stability (compare also with Hoxb-3-12, Figure 4). (**C**) The Hox proteins from the fourth paralog form complexes with Pbx1a and DNA which show similar dissociation rates. The four members of paralog 4 share highly conserved hexapeptide, linker and homeodomain sequences, but vary considerably in their N- and C-terminal flanking regions. These proteins form complexes with Pbx1a and DNA of similar stability, suggesting that the information required for complex stability resides within the conserved domains. (**D**) Hoxb proteins with changes at position 3 of the hexapeptide motif or lacking the N- and C- terminal flanking regions show similar dissociation rates to unmodified Hox proteins. Hoxb-4 proteins containing point mutations of the proline at position 3 of the hexapeptide (b4-P→A, b4-P→D) form complexes with similar stability to that of Hoxb-4, suggesting that this residue does not influence complex stability. A truncated Hoxb-7 protein lacking the N- and C-termini (b-7Δ) shows a similar dissociation rate to native Hoxb-7, suggesting that the terminal regions do not contribute to complex stability.

This explanation seemed particularly interesting since there appears to be a rough correlation between the distance separating the hexapeptide and homeodomains in the Hoxb proteins and Pbx1a–DNA complex stability (Tables 1 and 2). Hoxb-2 and Hoxb-3, which have long linker sequences (43 and 53 amino acids respectively) showed extremely short half-lives, whereas Hoxb-1 and Hoxb-4, which have smaller linker sequences (18 and 15 respectively), showed intermediate stability. Hoxb-5 and Hoxb-6, which have even smaller separation between the hexapeptide and the homeodomain (12 residues) were the most stable.

To test the influence of interval length on complex stability we took advantage of our isolation of two Hoxb-3 proteins containing variable linkers and assayed these proteins for complex stability as described above. One Hoxb-3 transcript, isolated from bone marrow cells, contains 53 amino acids between the hexapeptide and homeodomains (Hoxb-3-53). An apparent alternatively spliced Hox-b3 transcript was cloned from the K562 leukemia cell line, containing a 35 amino acid linker (Hoxb-3-35). In addition, two variant Hoxb-3 proteins were constructed in which the linker was further shortened to 17 (Hoxb-3-17) or 12 amino acids (Hoxb-3-12) respectively. As shown in Figure 3B and Table 1, neither the Hoxb-3-35 splice variant nor the Hoxb-3-17 protein, with a 17 residue linker, formed tighter complexes with Pbx1a than the Hoxb-3-53 protein. In addition, the Hoxb-3-12 variant, containing a 12 amino acid linker, which represents the linker distance in the naturally occurring Hox proteins with the tightest binding, did not bind to DNA with Pbx1a under normal assay conditions (Fig. 4, lane 3). A very weak complex could be visualized under lower stringency conditions (Fig. 4, compare lanes 20 and 21 or 24 and 25), demonstrating that the homeodomain of the mutant protein was folded correctly. Therefore, at least for the Hoxb-3 mutants tested, the distance between the hexapeptide and homeodomains does not appear to have a dominant effect in controlling complex stability.

The homeodomain alone does not confer complex stability

We next investigated whether the homeodomain alone confers complex stability by performing swapping experiments between homeodomains of Hox proteins which show high versus low complex stability. Recent studies have shown that the capacity of the

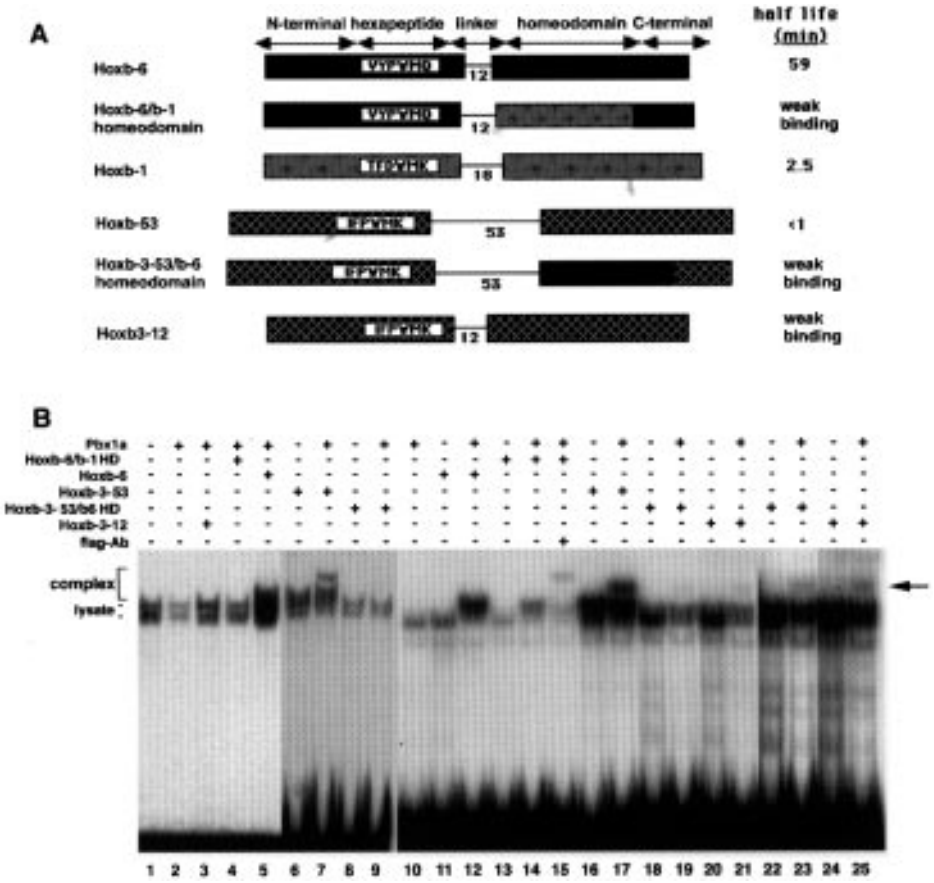

Figure 4. Neither the homeodomain sequence nor linker distance determine complex stability. (**A**) Chimeric Hoxb proteins. Mutant Hox proteins were constructed in which the homeodomain from a protein which forms a strong complex (Hoxb-6) was replaced by the homeodomain from a protein which forms a weak complex (Hoxb-1) to form Hoxb-6/b1 HD or where the homeodomain from a protein which forms an unstable complex (Hoxb-3-53) was replaced by the homeodomain from a protein forming a stable complex (Hoxb6) to form Hoxb-3/b-6 HD. The naturally occuring 53 amino acid linker arm in Hoxb-3-53 was reduced to 12 amino acids in Hoxb-3-12. (**B**) Chimeric Hoxb proteins form weak complexes with Pbx1a and DNA. Complex formation between the mutant Hox proteins with Pbx1a–DNA was compared with that of the unmodified proteins under normal stringency conditions (lanes 1–9) or reduced stringency conditions (lanes 10–25). Neither mutant Hox protein containing a swapped homeodomain forms a complex with Pbx1a and DNA under standard incubation conditions (lanes 4 and 9 respectively). When complex formation was allowed to occur under conditions of reduced stringency, weak complex bands were formed for both Hoxb-6/b1 HD (lanes 13 and 14) and Hoxb-3/b-6 HD (lanes 18 and 19). Since the Hoxb-6/b-1 HD complex migrates to a position very near the shifted bands seen in the lysate controls, an antiserum to a tag sequence fused to the Hox protein was used to supershift this complex (lane 15). The mutant Hoxb-3 protein containing a 12 amino acid spacer (Hoxb-3-12) was unable to form a detectable complex with Pbx1a and DNA under standard conditions (lane 3). However, under reduced stringency conditions a weak complex band can be detected (compare lanes 20 and 21). Lanes 22–25 represent a longer (3×) autoradiographic exposure of lanes 18–21 to more clearly show the complex bands (arrow).

Ubx but not that of the *Drosophila* homeodomain protein Antp to form complexes with Exd is dependent on the respective homeodomain sequence (32). The homeodomain from the Hoxb-6 protein, which forms the most stable complex, was replaced with the homeodomain from the Hoxb-1 protein, which forms an unstable complex (Fig. 4A, Hoxb-6/b-1 HD). Conversely, the homeodomain of Hoxb-3, which forms a very unstable complex, was replaced with the homeodomain from Hoxb-6 (Fig. 4A, Hoxb-3/b-6 HD). The accuracy of each chimeric construction was confirmed by DNA sequencing and by sizing of *in vitro* translation products by gel electrophoresis. In each case the resulting proteins were unable to form complexes with Pbx1a and DNA under our standard assay conditions (Fig. 4, compare lane 4 with 5 or 7 with 9). When complex formation was studied under reduced stringency conditions weak complexes were detected for each chimera (Fig. 4, compare lane 13 with 14 and 15 or 18 with 19), demonstrating in each case that the homeodomains within the chimeric proteins were folded correctly but incapable of forming high affinity complexes with Pbx1a and the target oligonucleotide and suggesting that complex stability is not due solely to properties inherent within the homeodomain.

Certain amino acids within the Hox protein hexapeptide motif are required for complex formation with Pbx1a and DNA

We have previously reported that removal of the entire hexapeptide sequence prevents complex formation between either Hoxb-7 or Hoxb-6 and Pbx1a with DNA (28). We have attempted to dissect the contributions of each of the amino acids within this domain to complex formation and stability. As seen in Table 1, the hexapeptides of the Hox proteins in paralog groups 1–8 show substantial variation around an absolutely conserved tryptophan–methionine core at positions 4 and 5. These invariant amino acids were changed to glycine and isoleucine respectively in Hoxb-4 to test their importance in complex formation. Changing either the

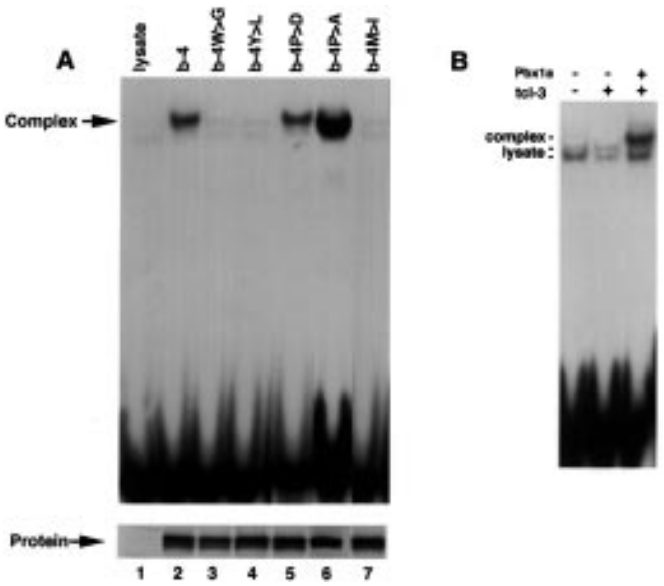

Figure 5. (**A**) Specific amino acids within the hexapeptide domain appear to be required for cooperative binding with Pbx1a and DNA. A series of mutant Hoxb-4 proteins, containing specific point mutations in the hexapeptide domain, were synthesized and tested for cooperative DNA binding with Pbx1a and DNA. ^{35}S-Labeled protein is shown below the gel shift lane to demonstrate that approximately equal amounts of protein were used in each DNA binding assay. Variability in complex intensities reflects differing autoradiographic times, rather than complex stability, as shown by the dissociation data in Figure 3D and reported in Table I. (**B**) tcl-3, a non-Hox homeodomain protein containing a YPWM motif, forms a complex with Pbx1a and DNA. *In vitro* translated tcl-3 and Pbx1a proteins formed a complex with the same oligonucleotide which was used for the Hox-b proteins when incubated under standard assay conditions, under which the tcl-3 protein did not bind alone to the oligonucleotide.

tryptophan (b-4-W→G) or the methionine (b-4-M→I) of the hexapeptide motif resulted in complete abolition of complex formation (Fig. 5A, lanes 3 and 7 respectively). As described below, Hoxb proteins containing either a tyrosine or a phenylalanine at position 2 of the hexapeptide are capable of forming Pbx1a–DNA complexes. However, replacement of an aromatic side chain at this position with a leucine in Hoxb-4 also resulted in complete loss of complex forming capacity (Fig. 5A, lane 4).

Variable amino acids within the hexapeptide motif do not account for differences in complex stability

Examination of the natural variations which occur within the hexapeptide sequence in the Hoxb proteins (Table 1) reveals that differences in the first and sixth positions do not appear to determine complex stability. For example, Hoxb-3 and Hoxb-7, which share an isoleucine at position one of the hexapeptide, formed Pbx1a–DNA complexes of dramatically different stabilities. In a similar fashion, Hoxb-6 and Hoxb-5, which have a glutamine or an arginine respectively at position six of the hexapeptide, showed very similar complex stabilities. Furthermore, Hoxb-1, b-2 and b-3, which formed unstable complexes, possess a basic amino acid (lysine) at position six, while Hoxb-5 and b-7, which also possess a basic residue (arginine) at this position, formed very stable complexes.

Hoxb-1 contains an aspartic acid in place of the proline which is conserved in position 3 of the hexapeptide in all of the other Hoxb proteins. Since proteins which formed both unstable (Hoxb-2 and b-3) and stable (Hoxb-5 and b-6) complexes contain a proline at position three, it was unlikely that this residue conferred complete stability on Pbx1a–DNA interactions. However, since Hoxb-1 formed the least stable complex for which we could obtain meaningful half-life data, we tested the possibility that the aspartic acid in position 3 was destablizing, using a mutant Hoxb-4 containing this amino acid in place of the normal proline in position 3. However, Hoxb-4-P→D formed a complex with Pbx1a and DNA which was at least as stable as that observed for the Hoxb-4 wild-type molecule (Fig. 5A, lane 5, Fig. 3D and Table 1). Since position 3 of the hexapeptide naturally contains either proline or aspartic acid (Table 1), this position was changed to alanine to further test the importance of this residue for complex stability. The Hoxb-4-P→A protein formed a complex with Pbx1a and DNA which had a half-life very similar to that of wild-type Hoxb-4. (Fig. 5A, lane 6, Fig. 3D and Table 1). These data demonstrate that complex stability of the Hox proteins with Pbx1a and DNA is relatively insensitive to changes at position 3. Since proline is relatively conserved within the hexapeptide domains, there may be additional functions for this amino acid which are not revealed by the current assay.

Although the literature refers to a conserved YPWM core, examination of Table 1 shows that the tyrosine at position two of the hexapeptide is replaced by phenylalanine in four of the eight Hoxb proteins studied. This variability does not appear to influence complex stability, since proteins containing either amino acid can form strong Pbx1a–DNA interactions (Table 1, compare Hoxb-5 and Hoxb-6). These data, together with those described above, suggest that within the core YPWM the invariant tryptophan and methionine as well as an aromatic residue at position 2 appear to be required for cooperative binding with Pbx1a to DNA. However, we cannot exclude the possibility that substitution of these positions with other amino acids might permit complex formation.

The hexapeptide domain, linker region and homeodomain appear to be sufficient for stable complex formation

We have previously reported that a Hoxb-7 deletion protein which contains only the hexapeptide, a five amino acid linker and the homeodomain can form a complex with Pbx1a and DNA (28), but the stability of this mutant was not determined. The mutant Hoxb-7, which lacks 119 amino acids preceding the conserved hexapeptide motif and 50 amino acids C-terminal to the homeodomain, formed a complex with Pbx1a and the oligonucleotide which was essentially as stable as that formed with the full-length protein (Fig. 3D and Table 1, b-7Δ). These data, together with the results of linker shortening and homeodomain swapping experiments, suggest that although each component alone does not confer stability, taken together, the hexapeptide, linker and homeodomain are sufficient to determine complex stability. Support for this hypothesis was obtained by examining complex stabilities of the four Hox proteins which comprise the fourth homology group (a-4, b-4, c-4 and d-4). In these proteins the hexapeptide motif is absolutely conserved in the first five amino acids, with a conservative Lys→Arg change in position 6 in Hoxb-4, the 15 amino acid linker preceding the homeodomain is highly conserved (Table 2) and the homeodomains contain only seven conservative amino acid substitutions. There is considerably more sequence variation outside these regions among the four proteins. The observation that these four Hox proteins showed very similar Pbx1a–DNA complex stabilities (Fig. 3C and Table 1) suggests that the hexapeptide, linker and homeodomain confer most of the complex stability.

Table 1. Half-lives of complexes formed between Hox proteins, Pbx1a and DNA

Hox protein	Hexapeptide sequence	Amino acids between hexapeptide and homeodomain	k_d[a]	Half-life (min)
b-1	TFDWMK	18	0.120	2.5
b-2	EFPWMK	43	Binds	<0.5[b]
b-3-53[c]	IFPWMK	53	Binds	<0.5[b]
b-3-35[c]		35	Binds	<0.5[b]
b-3-17		17	Binds	<0.5[b]
b-3-12		12	No binding[d]	
b-3/Hox b-6 HD		53	No binding[d]	
b-4	VYPWMR	15	0.031	9.6
b-4-Y→L	--L---------		No binding	
b-4-P→A	----A-------		0.034	8.8
b-4-W→G	------G-----		No binding	
b-4-M→I	---------I---		No binding	
b-4-P→D	----D-------		0.027	11.0
b-5	IFPWMR	12	0.007	40.1
b-6	VYPWMQ	12	0.005	59.0
b-6/Hoxb-1 HD		12	No binding[d]	
b-7	IYPWMR	5	0.021	14.4
b-7Δ		5	0.018	10.6
b-8	LFPWMR	6	0.013	23.1
a-4	VYPWMK	15	0.029	10.3
b-4	VYPWMK	15	0.031	9.6
c-4	VYPWMK	15	0.024	12.3
d-4	VYPWMK	15	0.033	9.0
tcl-3	TFPWME	22	Binds	<0.5b

[a]The value for k_d represents the slope of the line generated from a dissociation plot of log(complex) versus time. For each determination the correlation coefficient for the best fit line was >0.96.
[b]Dissociation is too rapid to measure a half-life, all complex having dissociated by 1 min following addition of cold competitor DNA.
[c]Alternatively spliced variants, differing by an 18 amino acid glycine-rich region between the hexapeptide motif and the homeodomain.
[d]No binding is observed under standard conditions, while weak binding is detected under reduced stringency.

Table 2. Sequences separating the hexapeptide[a] and homeodomains

Hox protein	Amino acid sequence between hexapeptide and homeodomain	k_d	Relative stability[b]
b-1	TFDWMKVKRNPPKTAKVSEPGLGS	0.120	4.2
b-2	EFPWMKEKKSAKKPSQSATSPSPAASAVPASGVGSPADGLGLPEAGGGG	Binds	<1[c]
b-3-53[d]	IFPWMKESRQTSKLKNNSPGTAEGCGGGGGGGGGGGSGGSGGGGGGGGGGGDKSPPGSAA	Binds	<1[c]
b-3-35 [d]	IFPWMKESRQTSKLKNNSPGTAEGCG----------GGGGGGDKSPPGSAA	Binds	<1[c]
b-3-17	IFPWMKESRQTSKLKNNS----------PGSAA	Binds	<1[c]
b-3-12	IFPWMKESRQTSKLKNNS----------	No binding[e]	
b-4	VYPWMRKVHVSTVNPNYAGGE	0.031	16
b-5	IFPWMRKLHISHDMTGPD	0.008	68
b-6	VYPWMQRMNSCNSSSFGP	0.005	100
b-7	IYPWMRSSGTD	0.021	24
b-8	LFPWMRPQAAAG	0.013	39
a-4[f]	VYPWMKKIHVSAVNSSYNGGE	0.029	17
b-4[f]	VYPWMRKVHVSTVNPNYAGGE	0.031	16
c-4[f]	VYPWMKKIHVSTVNPNYNGGE	0.024	21
d-4[f]	VYPWMKKVHANSVNPNYTGGE	0.033	15
tcl-3	TFPWMESNRRYTKDRFTGHPYQNRTPPK	Binds	<1[c]

[a]The hexapeptide sequences are underlined.
[b]The relative stability is reported as the ratio of the half-life compared with that for Hoxb-6, which was normalized to 100.
[c]Dissociation is too rapid to measure, a maximum relative stability is estimated.
[d]The alternatively spliced variant (b3-35) is missing a glycine-rich domain.
[e]No binding is observed under standard conditions; weak binding is detected under reduced stringency.
[f]Identical amino acids within the paralogous proteins are double underlined.

A non-Hox homeodomain protein containing a YPWM motif binds cooperatively to DNA with Pbx1a

Although the majority of homeodomain-containing proteins which contain a conserved hexapeptide sequence are found within the Antennapedia or class I homeoproteins, there are several proteins with divergent homeodomains which contain a hexapeptide motif. The tcl-3 (Hox11) protein, which was isolated as a break-point mutation in T cell leukemias (33,34), contains a TFPWME sequence which is located 22 amino acids upstream from the homeodomain (Tables 1 and 2). tcl-3 protein bound cooperatively to the target oligonucleotide with Pbx1a, but exhibited no capacity to bind in the absence of Pbx1a (Fig. 5B). The tcl-3–Pbx–DNA complex was so unstable that a half-life could not be measured (data not shown).

DISCUSSION

Although all of the Hoxb proteins containing a recognizable hexapeptide sequence (Hoxb-1–Hoxb-8) display cooperative binding with Pbx1a, the stability of the complexes formed with a specific oligonucleotide sequence vary over 100-fold. We have examined the contributions of three structural features of Hox proteins to complex stability: (i) variability in the linker length between the hexapeptide and homeodomain; (ii) differences in homeodomain sequences; (iii) variability of amino acids within the hexapeptide. Our findings suggest that these determinants function in restricted combinations to modulate complex stabilities.

Within the proteins of the *Hoxb* locus the conserved hexapeptide motif is separated from the homeodomain by a variable linker. Computer analysis of the linkers for the Hoxb proteins revealed no conserved amino acids or apparent structural motifs. Most *Hox* genes contain an intron which interrupts the linker region, presenting the possibility of multiple splicing variants which would place the hexapeptide and homeodomains at different distances from each other, such as are found in the protein variants of the *Drosophila Ubx* gene (35). The *Hoxb-3* cDNA clones isolated from normal bone marrow $CD34^+$ cells and a leukemic cell line contain linker regions of 53 and 35 amino acids respectively. However, in this case both variants formed unstable complexes and it was not possible to detect differences between them. There appears to be a rough inverse correlation between linker length and complex stability for the Hoxb proteins, as well as for tcl-3, which contains a long linker arm between the two domains. However, mutant Hoxb-3 proteins in which the linker was reduced to 17 or 12 residues showed either no change (Hoxb-3-17) or decreased complex stability (Hoxb-3-12), suggesting that linker length by itself is not sufficient to determine complex stability. Although stability data were not reported, Ubx proteins containing different linkers show differences in their abilities to interact with Exd in a yeast two hybrid system, suggesting that linker length and composition may, under other conditions, contribute to overall DNA–protein complex stability (32). Since many *Hox* genes express multiple cDNAs, most of which have not yet been characterized, it remains to be seen if differences between linker distances represents a mechanism for altering Hox protein function. In addition, individual linkers may be involved in additional protein–protein interactions.

In some respects the Hox proteins resemble other transcription factors which appear to have distinct modular structures which confer DNA binding or transcriptional activation (36). Thus the homeodomains of Hox proteins can bind to DNA in the absence of flanking N- or C-terminal regions (13,14). We have obtained contradictory data concerning the modularity of Hox proteins with regard to cooperative DNA binding with Pbx1a. Replacement of either the Hoxb-3 homeodomain with that of Hoxb-6 or the Hoxb-6 homeodomain with that of Hoxb-1 yields mutant proteins which form barely detectable complexes with Pbx1a and DNA. These results support the concept that the unique flanking regions of each Hox protein interact with its homeodomain in a manner which is sensitive to subtle changes induced by domain swapping experiments and that the capacity to bind cooperatively to DNA with Pbx1a reflects specific protein–protein contact residues within each particular Hox protein. However, we have previously demonstrated substantial enhancement of the cooperative DNA binding of Hoxa-10 with Pbx1a by mutational insertion of the Hoxb-7 hexapeptide and linker domains into the Hoxa-10 structure (28). Taken together, these results suggest that the capacity to form a cooperative DNA binding complex with Pbx1a appears to result from the combined interactions of the hexapeptide and homeodomain motifs which, at least in some cases, do not appear to function as autonomous modules. Furthermore, the structural information required for stable complex formation appears to reside largely within the hexapeptide, linker and homeodomain, since removal of amino acids outside these regions did not affect the stability of complexes formed with Hoxb-7. To exemplify this point further, the four proteins from the fourth paralog, which have almost identical hexapeptide, linker and homeodomain sequences, were shown to form Pbx1a–DNA complexes with very similar half-lives.

In addition to our previous report demonstrating the importance of the hexapeptide motif for cooperative Hox protein binding with Pbx1a and DNA (28), several other groups have also recently identified the YPWM sequence as being important for mediation of complex formation (30,32,37). While this work was under submission, Neuteboom *et al.* have reported that within the Hoxb-8 hexapeptide the phenylalanine, tryptophan and methinone at positions 2, 4 and 5 respectively are required for interaction with Pbx proteins (37). These results agree completely with those which we have observed for Hoxb-4 interactions with Pbx1a. However, other data suggest that a conserved YPWM or similar motif may not be required for cooperative binding. The Engrailed protein, which lacks a recognizable hexapeptide motif, cooperatively binds to DNA in the presence of Exd (22). At much higher protein concentrations a Ubx construct lacking the N-terminal region, including the hexapeptide, forms a complex with Exd in the presence of DNA (23). Ubx appears to react with Exd through the C-terminal portion of the homeodomain itself, suggesting that the analogous region of the mammalian homeodomains may interact with Pbx and DNA. We have previously demonstrated that the Hoxa-10 homeodomain protein, which lacks a YPWM motif, can weakly cooperate with Pbx1 to bind DNA, but that cooperativity is greatly enhanced by insertion of a YPWM domain into the Hoxa-10 protein (28). These data suggest that although elements of the hexapeptide appear to be required for cooperative binding of certain Hox proteins to DNA through Pbx1a, there are additional mechanisms which stabilize interactions between Hox and Pbx proteins.

In addition to effects of protein–protein interactions, a likely explanation for the observed differences in complex stabilities is that the various Pbx–Hox complexes might have different optimal DNA sequence preferences. Thus there may exist endogenous sites in which either the sequence and/or the orientation between Hox and Pbx1a recognition sites favors formation of strong complexes between the proteins which form weak complexes on the particular

synthetic site we studied. In this regard, the Engrailed protein cooperatively binds with Exd to a DNA sequence which differs from that to which Ubx and Exd cooperatively bind (22). Furthermore, a different consensus DNA binding site for Hoxb-1 with Pbx/Exd has recently been identified (38). Thus, there may be a 'code' of particular Hox proteins and Pbx proteins which will bind to specific DNA sites which result in differing arrays of complex stabilities across the *Hox* loci. Although we have initiated studies to identify optimal synthetic and biological binding sites for each of the different Hox proteins with Pbx, these experiments will require substantial efforts beyond the scope of the present investigation.

ACKNOWLEDGEMENTS

This work was supported by the Research Service of the Department of Veterans Affairs (CL and HJL), by grants from the National Institutes of Health (N44DK-3-2219, CL and CA42971, MLC) and by the National Cancer Institute of Canada with funds from the Canadian Cancer Society and the Medical Research Council of Canada (GS and RKH). HJL is the recipient of a Department of Veterans Affairs Career Development Award. MLC is a fellow of the Leukemia Society of America.

REFERENCES

1 Dolle,P., Izpisua-Belmonte,J.-C., Falkenstein,H., Renucci,A. and Duboule,D. (1989) *Nature*, **342**, 767–772.
2 Wilkinson,D.G., Bhatt,S., Cook,M., Boncinelli,E. and Krumlauf,R. (1989) *Nature*, **341**, 405–409.
3 Chisaka,O. and Capecchi,M.R. (1991) *Nature*, **350**, 473–479.
4 Lufkin,T., Dierich,A., LeMur,M., Mark,M. and Chambon,P. (1991) *Cell*, **66**, 1105–1119.
5 Chisaka,O., Musci,T.S. and Capecchi,M.R. (1992) *Nature*, **355**, 516–520.
6 Rijii,F.M., Mark,M., Lakkaraju,S., Dierich,A., Dolle,P. and Chambon,P. (1993) *Cell*, **75**, 1333–1349.
7 Ramirez-Solis,R., Zheng,H., Whiting,J., Krumlauf,R. and Bradley,A. (1993) *Cell*, **73**, 279–294.
8 Gendron-Maguire,M., Mallo,M., Zhang,M. and Gridley,T. (1993) *Cell*, **75**, 1317–1331.
9 Condie,B.G. and Capecchi,M.R. (1994) *Nature*, **370**, 304–306.
10 Gehring,W.J., Qian,Y.Q., Billeter,M., Furukubo-Tokunaga,K., Affolter,M., Otting,G. and Wuthrich,K. (1994) *Cell*, **69**, 211–223.
11 Levine,M. and Hoey,T. (1988) *Cell*, **55**, 537–540.
12 Sasaki,H., Yamamoto,M. and Kuroiwa,A. (1992) *Mech. Dev.*, **37**, 25–36.
13 Ekker,S.C., von Kessler,D.P. and Beachy,P.A. (1992) *EMBO J.*, **11**, 4059–4072.
14 Pellerin,I., Schnabel,C., Catron,K.M. and Abate,C. (1994) *Mol. Cell. Biol.*, **14**, 4532–4545.
15 Acampora,D., D'Esposito,M., Faiella,A., Pannese,M., Migliaccio,E., Morelli,F., Stornaiuolo,A., Nitro,V., Simeone,A. and Boncinelli,E. (1989) *Nucleic Acids Res.*, **17**, 10385–10402.
16 Kissinger,C.R., Liu,B., Martin-Blanco,E., Kornberg,T.B. and Pabo,C.O. (1990) *Cell*, **63**, 579–590.
17 Wolberger,C., Vershon,A.K., Liu,B., Johnson,A.D. and Pabo,C.O. (1991) *Cell*, **67**, 517–528.
18 Qian,Y.Q., G., O., Furukubo-Tokunaga,K., Affolter,M., Gehring,W.J. and Wuthrich,K. (1992) *Proc. Natl. Acad. Sci. USA*, **89**, 10738–10742.
19 Kessel,M. and Gruss,P. (1990) *Science*, **249**, 374–379.
20 Stark,M.R. and Johnson,A.D. (1994) *Nature*, **371**, 429–432.
21 Mak,A. and Johnson,A.D. (1993) *Genes Dev.*, **7**, 1862–1870.
22 van Dijk,M.A. and Murre,C. (1994) *Cell*, **78**, 617–624.
23 Chan,S.-K., Jaffe,L., Capovilla,M., Botas,J. and Mann,R.S. (1994) *Cell*, **78**, 603–615.
24 Nourse,J., Mellentin,J.D., Galili,N., Wilkinson,J., Stanbridge,E., Smith,S.D. and Cleary,M.C. (1990) *Cell*, **60**, 535–545.
25 Kamps,M.P., Murre,C., Sun,X.-H. and Baltimore,D. (1990) *Cell*, **60**, 547–555.
26 Monica,K., Galili,N., Nourse,J., Saltman,D. and Cleary,M.J. (1991) *Mol. Cell. Biol.*, **11**, 6149–6157.
27 Rauskolb,C., Peifer,M. and Wieschaus,E. (1993) *Cell*, **74**, 1101–1112.
28 Chang,C.-P., Shen,W.-F., Rozenfeld,S., Lawrence,H.J., Largman,C. and Cleary,M.L. (1995) *Genes Dev.*, **9**, 663–674.
29 Lu,Q., Knoepfler,P.S., Scheele,J., Wright,D.D. and Kamps,M.P. (1995) *Mol. Cell. Biol.*, **15**, 3786–3795.
30 Phelan,M.L., Rambaldi,I. and Featherstone,M.S. (1995) *Mol. Cell. Biol.*, **15**, 3989–3997.
31 van Djik,M.A., Peltenburg,L.T.C. and Murre,C. (1995) *Mech. Dev.*, **52**, 99–108.
32 Johnson,F.B., Parker,E. and Krasnow,M.A. (1995) *Proc. Natl. Acad. Sci. USA*, **92**, 739–743.
33 Lu,M., Gong,Z.Y., Shen,W.-F. and Ho,A.D. (1991) *EMBO J.*, **10**, 2905–2910.
34 Hatano,M., Roberts,C.W.M., Minden,M., Crist,W.M. and Korsmeyer,S.J. (1991) *Science*, **253**, 79–82.
35 Kornfeld,K., Saint,R.B., Beachy,P.A., Harte,P.J., Peattie,D.A. and Hogness,D.S. (1989) *Genes Dev.*, **3**, 243–258.
36 Ptashne,M. (1988) *Nature*, **335**, 683–689.
37 Neuteboom,S.T.C., Peltenburg,L.T., van Dijk,M.A. and Murre,C. (1995) *Proc. Natl. Acad. Sci. USA*, **92**, 9166–9170.
38 Popperl,H., Bienz,M., Studer,M., Chan,S.-K., Aparicio,S., Brenner,S., Mann,R.S. and Krumlauf,R. (1995) *Cell*, **81**, 1031–1042.
39 Shen,W.-F., Largman,C., Lowney,P., Corral,J., Detmer,K., Hauser,C.,A., Simonitch,T.A., Hack,F.M. and Lawrence,H.J. (1989) *Proc. Natl. Acad. Sci. USA*, **86**, 8536–8540.
40 Shen,W.F., Detmer,K., Simonitch-Eason,T.S., Lawrence,H.J. and Largman,C. (1991) *Nucleic Acids Res.*, **19**, 539–545.
41 Peverali,F., D'Esposito,M., Acampora,D., Bunone,G., Negri,M., Faicella,A., Stornaiuolo,A., Pannese,M., Migliaccio,E., Simeone,A., Della Valle,G. and Boncinelli,E. (1990) *Differentiation*, **45**, 61–69.
42 Simeone,A., Pannese,M., Acampora,D., D'Esposito,M. and Boncinelli,E. (1988) *Nucleic Acids Res.*, **16**, 5379–5390.
43 Hauser,C.A. and Galang,C.K. (1992) *New. Biol.*, **4**, 558–568.
44 Galliot,B., Dolle,P., Vigneron,M., Featherstone,M.S., Baron,A. and Duboule,D. (1989) *Development*, **107**, 343–359.
45 Kongsuwan,K., Allen,J. and Adams,J.M. (1989) *Nucleic Acids Res.*, **17**, 1881–1892.
46 Featherstone,M.S., Baron,A., Gaunt,S.J., Mattei,M.-G. and Duboule,D. (1988) *Proc. Natl. Acad. Sci. USA*, **85**, 4760–4764.
47 LeBrun,D.L. and Cleary,M.L. (1994) *Oncogene*, **9**, 1641–1647.
48 Goutte,C. and Johnson,A.D. (1994) *EMBO J.*, **13**, 1434–1442