Oncogene (1997) 14, 75–83

Heterodimerization of Hox proteins with Pbx1 and oncoprotein E2a-Pbx1 generates unique DNA-binding specificities at nucleotides predicted to contact the N-terminal arm of the Hox homeodomain – demonstration of Hox-dependent targeting of E2a-Pbx1 *in vivo*

Qiang Lu[1] and Mark P Kamps[2]

[1]*Harvard Medical School, Department of Cell Biology, 240 Longwood Avenue, Boston, Massachussetts 02115, USA;* [2]*Department of Pathology, University of California, San Diego, School of Medicine, La Jolla, California 92093-0612, USA*

Hox proteins control genetic programs that orchestrate development, and a large subset of Hox proteins can bind DNA elements as heterodimers with the Pbx family of homeodomain proteins. A transcriptionally activated version of Pbx1, E2a-Pbx1, is an oncoprotein in human pre-B cell leukemia that strongly suppresses differentiation and retains its ability to heterodimerize with Hox proteins. Because monomeric Hox proteins bind very similar DNA motifs, it is unclear how they activate diverse developmental programs. Here we demonstrate that heterodimers containing different Hox proteins and a common Pbx1 or E2a-Pbx1 partner bind different DNA motifs. Structural models suggest that the specificity of the Hox protein is altered by a conformation change involving residues in the N-terminal arm of the Hox homeodomain. Mutational analysis also supported the hypothesis that unique sequences in the N-terminal arm of the Hox homeodomain are at least partially responsible for mediating this specificity. *In vivo*, Hox proteins directed E2a-Pbx1-mediated transactivation with moderate specificity to cognate Hox-Pbx motifs. Thus, the development specificity of individual Hox proteins may be mediated, in part, by differential targeting of cellular genes by Pbx1-Hox complexes. Likewise, through its function as a common heterodimer partner, oncoprotein E2a-Pbx1 may be able to interfere with multiple programs of development that are induced by the sequential or simultaneous expression of Hox proteins during hematopoiesis.

Keywords: E2a-Pbx1; oncoprotein; homeodomain; Hox

Introduction

The t(1;19) chromosomal translocation is found in 25% of human pediatric pre-B cell lymphoblastic leukemias (Carroll *et al*., 1984). This translocation produces a chimeric transcription factor containing the N-terminal transcriptional activation domain of E2a and the majority of the homeodomain (HD) protein, Pbx1 (Kamps *et al*., 1991; Nourse *et al*., 1990). E2a-Pbx1, but not Pbx1, functions as a persistent activator of transcription (Lu *et al*., 1994; Lebrun and Cleary, 1994; Van Dijk *et al*., 1993). E2a-Pbx1 exhibits a mitogenic activity in NIH3T3 fibroblasts (Kamps *et al*., 1991) and a proven ability to transform T cells (Dedera *et al*., 1993) and block differentiation in the myeloid lineage in mice. *In vivo*, mice reconstituted with marrow expressing E2a-Pbx1 develop myeloblastic leukemias (Kamps and Baltimore, 1993) and marrow-derived myeloid progenitors are immortalized after infection with E2a-Pbx1 retrovirus, exhibiting indefinite growth in the presence of the myeloid lymphokine, granulocyte-macrophage colony-stimulating factor (Kamps and Wright, 1994). Therefore, the molecular mechanism of transformation by E2a-Pbx1 in pre-B cell lymphoblastic leukemia may also involve both stimulation of proliferation and disruption of differentiation.

PBX1, and its related genes, PBX2 and PBX3 (Monica *et al*., 1991), are homologues of Drosophila Extradentical (EXD). The Pbx1 HD binds TGAT, and both Pbx1/Exd and E2a-Pbx1 bind DNA elements (e.g. TGATTGAT and TGATTAAT) as heterodimers with HD proteins encoded by homeotic selector genes residing in the Drosophila HOM-C loci (Van Dijk and Murre, 1994; Chan *et al*., 1994) and by Hox proteins of higher eucaryotes that reside on the structurally similar HoxA-HoxD loci (Lu *et al*., 1995; Chang *et al*., 1995; Neuteboom *et al*., 1995; Figure 1a). While heterodimer formation with Hox proteins is not essential for the ability of E2a-Pbx1 to induce foci in NIH3T3 (Kamps *et al*., 1996) or T cell leukemia in mice (Dedera *et al*., 1993) it remains important for its ability to block myeloid differentiation (Kamps *et al*., 1996). Like their positional cognates of the Drosophila Antennapedia and Bithorax complexes (ANT-C and BX-C, respectively, Figure 1a), Hox genes exhibit a temporally- and spatially-restricted pattern of expression that orchestrates normal differentiation of structures along the anterior-posterior axis of the skeletal and central nervous systems (Krumlauf, 1994; McGinnis *et al*., 1990). Hox genes are also expressed during organogenesis and hematopoiesis (Vielle-Grosjean *et al*, 1992; Mathews *et al*., 1991; Petrini *et al*., 1992; Lawrence *et al*., 1993). Variants of the Pbx1-Hox recognition element are important for tissue-specific expression of the HoxB1 gene in rhombomere 4 of the developing mouse hindbrain (TGATGGAT and AGATTGAT, Popperl *et al*., 1995) and in pancreatic expression of somatostatin (TGATTAAT, Peers *et al*., 1995). Formation of Pbx1-Hox and E2a-Pbx1-Hox heterodimers requires interaction of the Pbx1 protein with a highly conserved Hox pentapeptide, usually YPWMR, which is positioned 4 to 56 residues N-

Correspondence: MP Kamps,
Received 5 April 1996; revised 2 September 1996; accepted 3 September 1996

terminal to the Hox HD and is encoded by all Hox genes numerically designated 1–8 (Peers *et al*., 1995; Knoepfler and Kamps, 1995; Johnson *et al*., 1995; Chang *et al*., 1995). Although different Hox proteins induce different genetic programs and morphologic structures *in vivo*, the DNA-binding specificity of their HD's *in vitro* is similar, preferring a TAAT core followed by GG, GA, TA or TG (Gehring *et al*., 1994a,b; Laughon, 1991). Therefore, one puzzle in developmental biology is how diverse genetic and morphologic events are mediated by factors that bind very similar or identical DNA elements.

Based on our biochemical characterization of a Pbx1-Hox-DNA complex formed on TGATTAATGG (Figure 1b; Lu *et al*., 1995; Lu and Kamps, 1996), as well as the work of others (Chan and Mann, 1996), the Pbx1 HD binds the 5′ TGAT core (positions 1–4) and the Hox HD binds the adjacent 3′ TAAT core (positions 5–10), maintaining a strong preference for GG, GA, TA, or TG 3′ of this core. As predicted from the crystal structure of Drosophila Eng (Kissinger *et al*., 1990), this orientation positions highly conserved Arg5 and invariant Arg3 of the *N-t*erminal *a*rm (NTA) of the Hox HD in the minor groove of DNA binding T5 and A6, respectively, and Asn51 and Gln50 of helix 3 in the major groove, binding A7 and the 3′ GG, GA, TA, or TG dinucleotide, respectively (Figure 1c). The orientation of Hox proteins on this core was proven biochemically by showing that conversion of Gln50 to Lys50 of the Hox protein, which is known to alter 3′ dinucleotide specifically from GG to CC, also converts a Pbx1-Hox recognition motif from TGATTAATGG to TGATTAATCC (Knoepfler *et al*., 1996).

Two observations suggest that Hox proteins may exhibit a different DNA-binding specificity as heterodimers with Pbx1 and E2a-Pbx1 than they do as monomers. First, HoxA5, HoxB7 and HoxB8 bind core motifs containing either TAAT or TGAT with apparently equal affinities when binding as heterodimers with Pbx1 (Lu *et al*., 1995). Second, the Pbx-HoxB1 element in the HoxB1 promoter contains GGAT as the Hox core (Popperl *et al*., 1995); thus, in both instances, DNA sequences proposed to bind the NTA of the Hox HD are non-canonical. Therefore, we tested the hypothesis that heterodimer formation with Pbx1 or E2a-Pbx1 might alter sequence-specific DNA-binding by Hox proteins.

Here we demonstrate that Hox-Pbx1 complexes exhibit unique DNA-binding specificities at positions predicted to be bound by Arg3 and Arg5 of the NTA of the Hox HD. The same specificities are retained by complexes containing E2a-Pbx1. Substituting the NTA sequences of a recipient Hox protein with those of a doner Hox protein shifts the DNA-binding specificity of the Pbx complex toward that of the donor. Because the proposed location of the Hox NTA is at the heterodimer interface and because at least a portion of the pentapeptide sequence upstream of the NTA is thought to bind the Pbx1 HD, we suggest that the conformation of the NTA may be altered by interaction with Pbx1, therein changing Hox DNA-binding specificity. Because different paralogs of Hox genes (1–8) encode different sequences adjacent to Arg3 and Arg5, heterodimer formation with Pbx proteins could reconfigure the tertiary structure of the Hox NTA such that it binds optimally to a unique half-site. This represents a general mechanism that could account for differential gene targeting by different Hox proteins, and thus explain some of the unique developmental functions of Hox proteins. It also implies that oncoprotein E2a-Pbx1 may be able to interfere with multiple programs of development that are induced by the sequential or simultaneous expression of Hox proteins during B cell or myeloid differentiation.

Results

Dimers of Hox proteins and Pbx1 bind different optimal DNA motifs

To determine whether varying the Hox protein in Pbx1-Hox or E2a-Pbx1-Hox heterodimers produces different optimal DNA-binding sequences, an oligonucleotide degenerate at 20 consecutive internal positions was mixed with recombinant histidine-tagged Pbx1 and GST-tagged HoxA5, HoxB8, or HoxC8, and DNA motifs bound by Pbx1-Hox heterodimers were selected by sequential selection for the histidine and GST tags, followed by PCR amplification. This selection was repeated five times (see Materials and methods) and yielded a single class of binding site in which the 5′ half site contained the invariant TGAT Pbx1-recognition sequence and the juxtaposed 3′ half site contained variations of the TAAT Hox recognition core, followed by a preference for a GG, GA, TA or TG dinucleotide 3′ to the Hox core (Figure 1d). The unspaced positioning of the Pbx1 and Hox cores, as well as the consistent orientation of the Hox protein (based on 3′ dinucleotide preference) are identical to those predicted from our original characterization of Pbx1-Hox complexes binding a TGATTAATGG sequence. All motifs contained an A at the third position of both Pbx1 and Hox core sequences, consistent with the proposed interaction of this base with the invariant Asn51 of both the Pbx1 and Hox HD's. This analysis demonstrated that the orientation and spacing of Pbx1 and Hox proteins does not vary when the identity of the Hox partner is altered, and positions the Pbx1 and Hox HDs on opposing sides of the double helix. A striking variation in sequence specificity occurred at position two of the Hox half-site, which is predicted to contact the conserved Arg3 in a Hox-DNA complex. While monomeric Hox proteins bind A at this position, heterodimers of Pbx1 and HoxB8 or HoxC8 selected T and those containing HoxA5 bound motifs containing A or G.

Differential recognition of DNA motifs by different Pbx1-Hox heterodimers was next characterized by electrophoretic mobility shift analysis (EMSA), using HoxA1, HoxA5, HoxB8, and HoxC8, and probes containing a 5′ TGAT Pbx1 core flanked by a 3′ TAAT, TGAT, or TTAT Hox core and a GA 3′ dinucleotide. While heterodimers containing HoxA1 strongly preferred a TGAT Hox core (Figure 2, lane 3 *vs* 12, 21), both HoxB8 and HoxC8 strongly preferred TTAT (lane 25 and 27 *vs* 7, 9, 16, 18), and HoxA5 formed heterodimers on all three Hox core sequences (lanes 5, 14, 23). This analysis demonstrated that different Hox-Pbx1 heterodimers bind optimally to unique DNA sequences that vary at position 2 of the Hox recognition core.

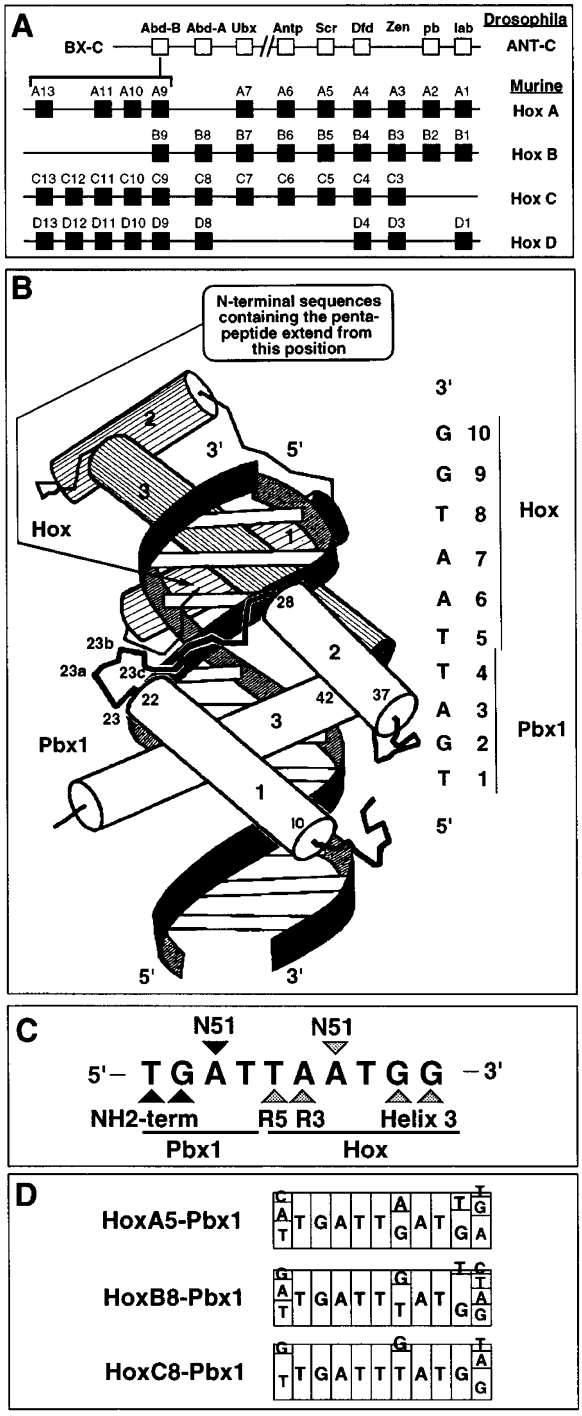

Figure 1 (**a**) Schematic representation of the ANT-C and BX-C of Drosophila and of the Hox-A through Hox-D loci in higher eucarocytes. (**b**) Model of the Pbx1-Hox-DNA complex. (**c**) Proposed sites of specific protein-DNA contacts in the Pbx1-Hox-DNA complex. (**d**) DNA motifs selected from populations of degenerate oligonucleotides by combinations of recombinant Pbx1 plus HoxA5, HoxB8, or HoxC8

Dimerization alters the DNA-binding specificity at residues predicted to bind the Hox N-terminal arm

To investigate the biochemical basis responsible for shifting the DNA-binding specificity of Hox proteins from TAAT in Hox-DNA monomers to TGAT or TTAT in Pbx1-Hox heterodimers, the stability of Hox, Hox-Pbx1, and Hox-E2a-Pbx1 complexes on probes containing the TAAT, TTAT or TGAT Hox cores were measured (Table 1). The dissociation half-life for bacterially-expressed HoxA5 on the TAAT probe was 40 s and dropped to 12 s for binding to the TTAT or TGAT probes (Table 1). Likewise, the dissociation half-life for recombinant HoxB8 was 18 s for the TAAT probe, and dropped below detection (<8 s) for the TGAT and TTAT probes. This affirmed that both HoxA5 and HoxB8 bound more tightly to a TAAT core than to either a TGAT or a TTAT Hox core. By contrast, dissociation of HoxB8 plus E2a-Pbx1 from the TTAT core was much slower ($t_{1/2}$ of 48 min) than dissociation from either the TGAT or TAAT cores ($t_{1/2}$ of 11 and 6.5 min). Dissociation of HoxA5 plus E2a-Pbx1 from the TTAT or TGAT Hox cores (67 and 42 min, respectively) was also slower than that measured from the TAAT core (20 min). Dissociation half-lives using bacterially-expressed Pbx1 plus HoxB8 or HoxA5 yielded similar results (Table 1), suggesting that heterodimerization with either Pbx1 or E2a-Pbx1 shifts the stability of Hox DNA-binding away from a TAAT core and toward a TTAT or TGAT core.

An expanded analysis of the DNA-binding specificity of Hox-Pbx and Hox-E2a-Pbx1 complexes was next conducted on a panel of binding sites containing variations in positions one and two of the TAAT Hox core, using HoxA1, HoxB7, HoxB8, HoxC8, and HoxD4 (Figure 3). As was observed for HoxA5 and HoxB8, the HoxA1, HoxB7 and HoxD4 proteins also bound TAAT cores preferentially as monomers, as judged from the intensity of the monomeric band in EMSA (quantitated in Table 1). Complexes containing HoxB7 and Pbx1 or E2a-Pbx1 exhibited a shift in affinity from a TAAT Hox core to a TTAT Hox core, similar to that exhibited by HoxB8 or HoxC8. The dissociation half-lives of HoxB7-E2a-Pbx1 heterodimers bound to the TTAT, TGAT and TAAT probes were 75, 8 and 8 min, respectively, demonstrating that the heterodimer with HoxB7 has an even greater selectivity for the TTAT Hox half site than do those containing HoxB8 or HoxC8 (Table 1). The heterodimer of HoxD4 and E2a-Pbx1 exhibited a relative binding specificity similar to that of HoxA5 plus E2a-Pbx1. Interestingly, the Pbx1-HoxA1 heterodimer bound uniquely to probes containing a TCAT or CGAT Hox half site. Likewise, the Pbx1-HoxB7, -HoxB8 and -HoxC8 heterodimers bound ATAT Hox half-sites with an affinity much greater than did those containing HoxA1, HoxA5 or HoxD4. Both oncogenic forms of E2a-Pbx1 (E2a-Pbx1a or E2a-Pbx1b), which arise from differential mRNA splicing, exhibited the same heterodimer DNA-binding specificity with Hox proteins as did Pbx1. As we observed earlier (Lu *et al*., 1995), complexes containing E2a-Pbx1 were somewhat more abundant than those containing Pbx1. This likely reflects their greater stability (slower off-rate), which we originally reported for E2a-Pbx1 complexes formed with either HoxB7 or HoxA5 (Lu *et al*., 1995). Using deletion mutants of Pbx1, only the Pbx1 HD was required to exhibit Hox-dependent heterodimer specificity, and inclusion of 15 amino acids C-terminal to the Pbx1 HD strongly increased the abundance of specific complex formation (data not shown). These minimal Pbx1 sequences are contained in all forms of Pbx1 and E2a-Pbx1.

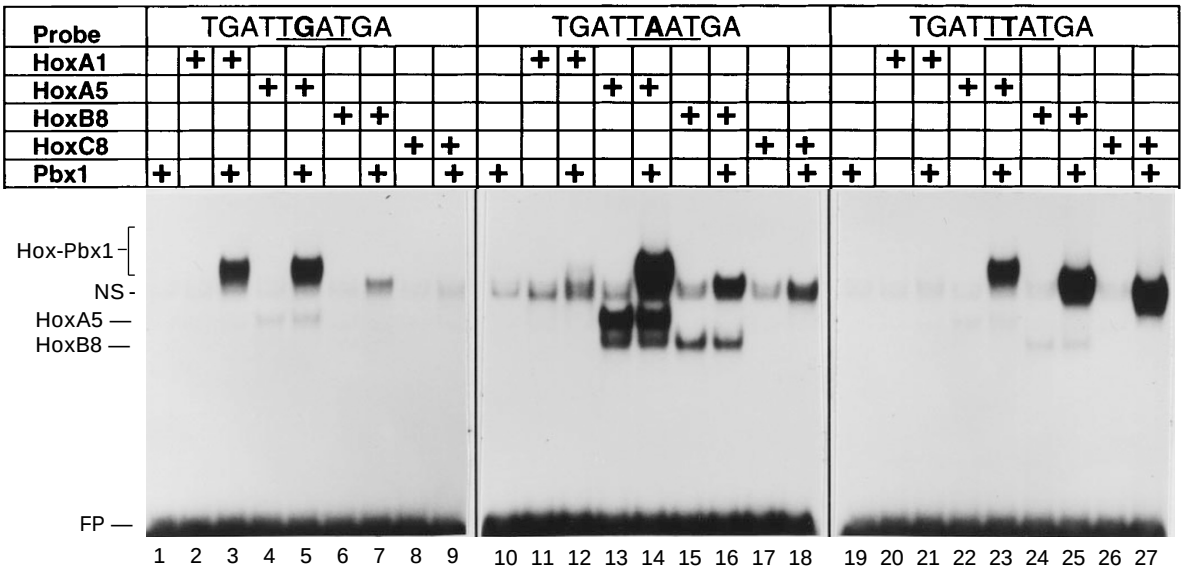

Figure 2 Analysis of DNA-binding specificities of Pbx1-Hox complexes, using EMSA. Sequences of the DNA probe are indicated at top, and content of Pbx1 or Hox proteins are indicated by plus signs above each lane. NS represents a non-specific complex that forms with somewhat different abundances on each probe. All probes were synthesized at the same specific activity, and the same exposure times are illustrated. 2 μl of Pbx1 and Hox proteins were added to each gel-shift reaction

Table 1

	Monomeric proteins						*Heterodimers of Hox proteins plus Pbx1 or E2A-Pbx1*											
	Abundance ratio[2]			*Off-rate (s^3)*[4]			*Abundance ratio*[2]						*Off-rate (min)*					
							Hox/Pbx1a			*Hox/E2a-Pbx1a*			*Hox/E2a-Pbx1a*[2]			*Hox-Pbx1b*[4]		
Probe[1]	*TAAT*	*TGAT*	*TTAT*	*TAAT*	*TGAT*	*TTAT*	*TAAT*	*TGAT*	*TTAT*	*TAAT*	*TGAT*	*TTAT*	*TAAT*	*TGAT*	*TTAT*	*TAAT*	*TGAT*	*TTAT*
HoxA1	1.0	0.26	0.16				1.0	8.60	0.32	1.0	7.60	0.43	2.5	11.5				
HoxA5	1.0	0.17	0.10	40	12	12	1.0	0.49	0.40	1.0	0.39	0.32	20.0	42.0	67.0	33.0	37.0	60.0
HoxB7	1.0	0.06	0.21				1.0	0.97	4.32	1.0	0.77	1.80	8.0	8.0	75.0			
HoxB8	1.0	0.06	0.39	18	<8	<8	1.0	0.91	8.01	1.0	0.77	3.30	6.5	11.0	48.0	10.0	18.0	60.0
HoxC8	–	–	–				1.0	0.62	15.8	1.0	0.42	2.50	6.0		42.0			
HoxD4	1.0	0.12	0.09				1.0	0.35	0.36	1.0	0.44	0.28				7.5	10.0	14.0

[1]Probe contains TGAT_ _ _ _GA, where designated sequence occupies the underlined region. [2]Using proteins produced by coupled transcription/translation *in vitro*. [3]Data fit an exponential decay function using a 1.0 min correction factor to account for dissociation of the complex before entrance into the gel. [4]Using purified proteins produced in bacteria

The N-terminal arm of the Hox protein is partially responsible for dictating binding-specificity

Generation of unique DNA-binding properties for individual Pbx1-Hox complexes might arise if heterodimer formation with Pbx1 altered the conformation of the Hox protein at the Pbx1-Hox interface such that residues in the NTA no longer bind optimally to A at the second position of the Hox TAAT core, but rather to T or G. To test whether NTA residues confer specificity of DNA-binding by Pbx1-Hox complexes, hybrid Hox proteins were synthesized containing the NTA of HoxB7 in HoxA5, the NTA of HoxA1 in HoxA5 and the NTA of HoxA1 in HoxB8, and tested for their specificity of DNA-binding as complexes with E2a-Pbx1 on probes containing a 5′ TGAT Pbx core and a 3′ TAAT, TGAT, or TTAT Hox core. The DNA-binding specificities of each hybrid protein was compared with that of the parental Hox proteins (Figure 4). NTA mutations encompassed the first eight residues of each HD, and are indicated in the upper left corner of each panel in Figure 4. Insertion of the NTA of HoxB7 into HoxA5 shifted predominant binding from the HoxA5 specificity (TAAT Hox core) to the HoxB7 specificity (TTAT core; Figure 4a). Similarly, insertion of the NTA of HoxA1 into HoxA5 shifted predominant binding from the HoxA5 specificity (TAAT Hox core) to the HoxA1 specificity (TGAT core; Figure 4b). Likewise, insertion of the NTA of HoxA1 into HoxB8 shifted predominant binding from the HoxB8 specificity (TTAT Hox core) to the HoxA1 specificity (TGAT core; Figure 4c). In no case was the DNA-binding specificity of one Hox protein converted completely into that of another simply by acquiring its NTA sequence, indicating that unique sequences in the NTA of Hox proteins are partially but not completely responsible for determining sequence-specific DNA-binding by different Pbx1-Hox heterodimers.

The Hox-Pbx binding code dictates targetting of oncoprotein E2a-Pbx1

Finally, the specificity of Hox-E2a-Pbx1 heterodimers was examined *in vivo* by testing their ability to activate transcription from CAT constructs driven by the motifs TGATTAAT or TGATTTAT. While many cell lines exhibited activation of these constructs by E2a-Pbx1 alone, suggesting the presence of substantial quantities of heterodimerizing partners, activation of the con-

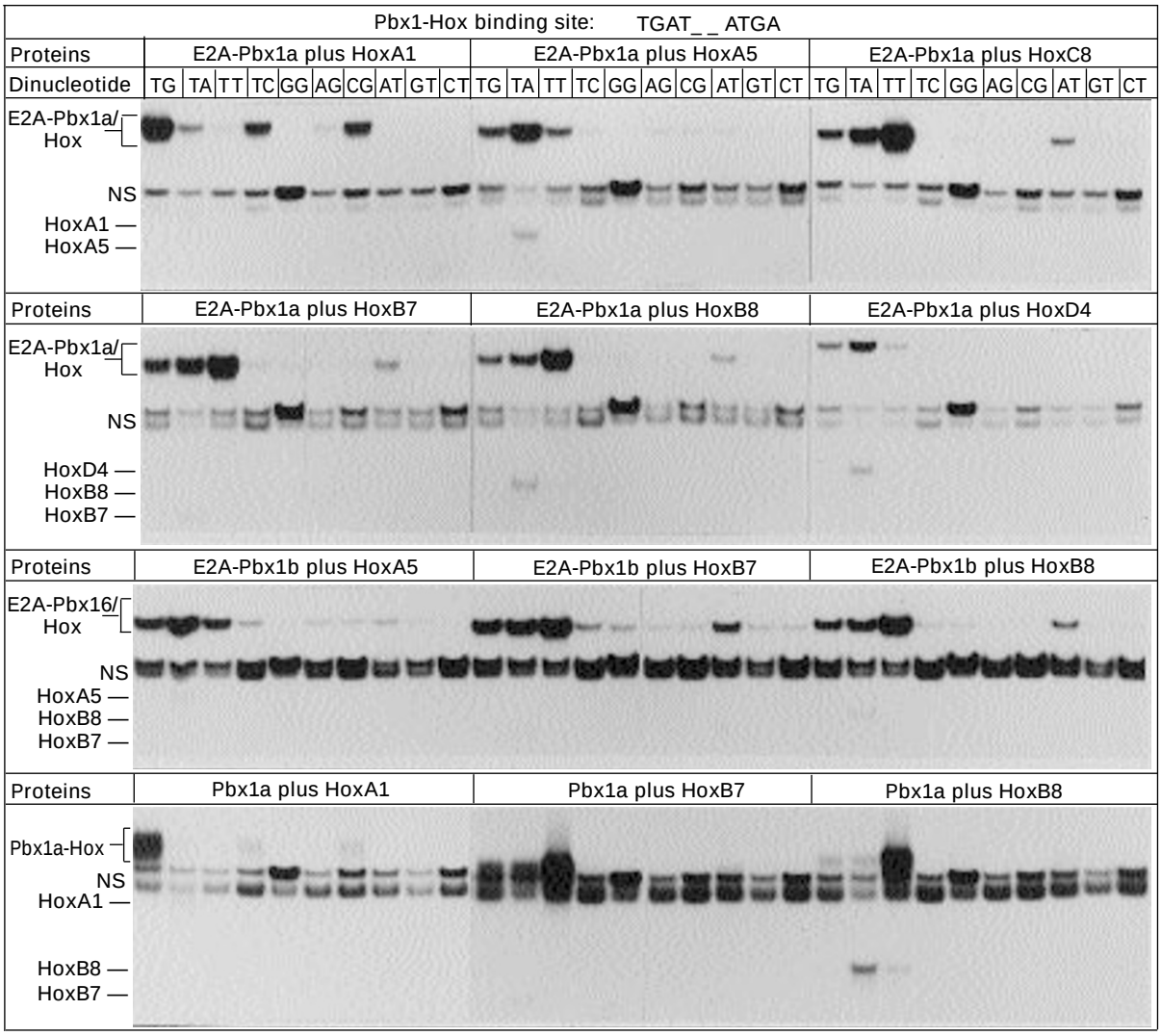

Figure 3 Analysis of binding-site specificity of Pbx1 heterodimers with HoxA1, HoxA5, HoxB7, HoxB8, HoxC8, and HoxD4. Probes containing variable dinucleotides within the first two bases of the Hox core (as designated) were combined with Pbx1 proteins (Pbx1, E2a-Pbx1a, E2a-Pbx1b) and the indicated Hox proteins. Complexes were resolved by EMSA, and are designated at left. NS designates nonspecific complex. Only the relevant portions of the gels containing Hox and Pbx1-Hox complexes are shown

structs by E2a-Pbx1 in the human pre-B cell line, Nalm-6, was very low. Likewise, in COS cells, the CAT construct driven by 6xTGATTAAT was not activated by E2a-Pbx1, even though that driven by 6xTGATTTAT was strongly activated. In no cases did expression of HoxA5, HoxB8, or HoxC8 alone activate transcription of any reporter constructs. In Nalm-6 cells, coexpression of HoxA5 with E2a-Pbx1 activated transcription through by the TGATTAAT motif (lane 3) somewhat better than through the TGATTTAT motif (lane 7), consistent with its marginal preference for binding TGATTAAT in conjunction with E2a-Pbx1. Likewise, specificity of HoxC8-E2a-Pbx1 complexes were recapitulated *in vivo*, activating transcription better through the TGATTTAT site than through the TGATTAAT site. Because the ability of E2a-Pbx1 to block myeloid differentiation is dependent on its DNA-binding ability while induction of foci in fibroblasts is not, we tested whether E2a-Pbx1 N682S, the DNA-binding mutant E2a-Pbx1, was able to activate transcription in conjunction with Hox proteins. In this case, COS cells were used because they yield the greatest degree of cooperative transactivation by E2a-Pbx1 and the representative Hox protein, HoxC8. E2a-Pbx1-N682S was absolutely incapable of activating transcription in conjunction with HoxC8, indicating that the ability of E2a-Pbx1 block differentiation correlates with its ability to activate transcription in conjunction with Hox proteins.

Discussion

The mechanism by which individual Hox genes induce specific developmental outcomes is difficult to explain based on sequence-specific DNA-binding alone because all pentapeptide-containing Hox proteins bind similar DNA motifs containing a TAAT core, followed by a preference for a GG, GA, TA or TG dinucleotide. Earlier studies have attributed the functional specificity of the Drosophila Ant and Scr HD proteins to residues in the NTA of the HD (Furukubo-Tokunaga *et al*., 1993; Zeng *et al*., 1993), Our discovery that binding as a dimer with Pbx1 produces Hox-dependent sequence-specificity mediated, in part, by the HD NTA suggests that

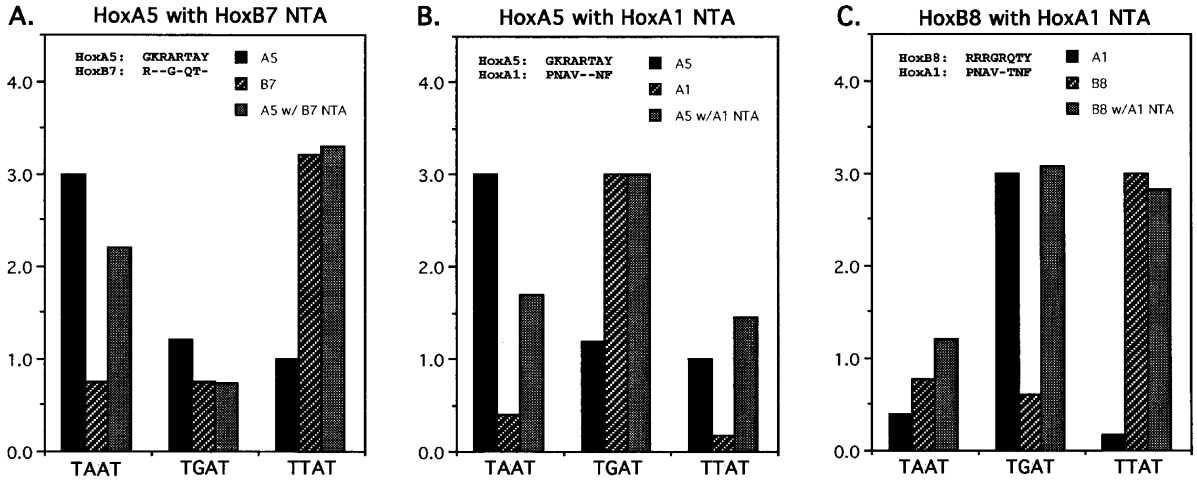

Figure 4 Binding site specificity of NTA exchange Hox proteins. Heterodimer formation with E2a-Pbx1b was quantitated for native Hox proteins and NTA chimera's on DNA probes containing a 5′ TGAT Pbx core and a 3′ TAAT, TGAT, or TTAT core, followed by a 3′ GA dinucleotide. (**a**) Complex abundance obtained with HoxA5, HoxB7 and an NTA mutant of HoxA5 replacing the first eight residues of its HD with those of the HoxB7 HD. (**b**) Complex abundance obtained with HoxA5, HoxA1 and an NTA mutant of HoxA5 replacing the first eight residues of its HD with those of the HoxA1 HD. (**c**) Complex abundance obtained with HoxA1, HoxB8 and an NTA mutant of HoxB8 replacing the first 8 residues of its HD with those of the HoxA1 HD

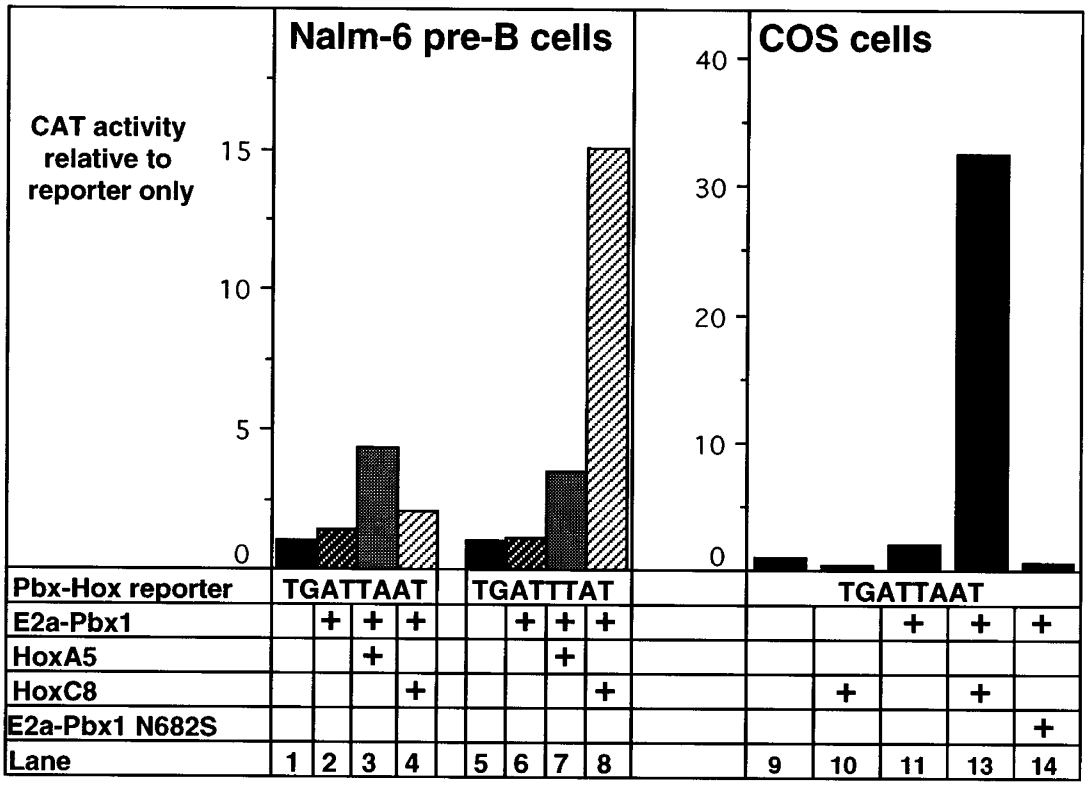

Figure 5 Differential targeting of E2a-Pbx1 transactivation by Hox proteins. CAT reporter genes driven by six tandom binding sites containing either a TGATTAAT or TGATTTAT motif were introduced into human Nalm-6 pre-B cells (lanes 1–8) or primate COS cells (lanes 9–14) and examined for transcriptional activation by E2a-Pbx1 alone (lanes 2, 6 and 11) or in combination with HoxA5 (lanes 3 and 7) or HoxC8 (lanes 4, 8, 13). Stimulation of CAT activity is measured relative to background activity observed in transfections containing the reporter only

Hox proteins can achieve a high degree of sequence-specificity when they interact with other transcription factors, and that the developmental specificity of Hox proteins may be accounted for, at least in part, by differential targeting to cellular promoters in conjunction with other transcription factors, such as Pbx1. While this is an attractive hypothesis, there is little direct proof as yet that specific Hox-Pbx1 sites mediate unique effects of Hox proteins on gene expression during development – only that the TGATGGAT motif, which strongly contributes to tissue-specific and developmental regulation of the HoxB1 promoter, forms an efficient heterodimer of Exd/Pbx plus HoxB1, suggesting that HoxB1 exerts certain developmental effects by targeting a GGAT core in conjunction with Pbx proteins.

Hox-dependent, sequence-specific, DNA-binding of E2a-Pbx1-Hox heterodimers, suggests that E2a-Pbx1 will activate transcription of different target genes in different cell types, depending on their content of heterodimerizing partners, which could vary at different stages of differentiation. Hox genes are expressed during hematopoiesis. Studies in leukemic cell lines demonstrate that HoxA genes (A1–A7) exhibit strong preferential expression in the myeloid lineage (Vieille-Grosjean *et al*., 1992), that HoxB genes (B2–B8) exhibit preferential expression in the erythroid lineage (Mathews *et al*., 1991), with some expression in myelomonocytic (B2 and B4), T cell (B4 and B7), and B cell (B4 and B7) lineages (Petrini *et al*., 1992) and that HoxC genes exhibit moderate specificity for B and T lymphoid lineages (C4 and C8; Lawrence *et al*., 1993). While analysis of Hox gene expression during normal hepatopoiesis is still in its early stages, the lineage-specific expression and coexpression of these Hox genes suggets that Hox proteins may contribute to lineage definition and differentiation state. Therefore, the fact that E2a-Pbx1 binds different optimal recognition elements in conjunction with different Hox proteins increases the number of possible target genes impacted by E2a-Pbx1 during B cell or myeloid differentiation. Under the former assumption that E2a-Pbx1 heterodimerizes on similar elements with Hox proteins, there would be one group of target genes that could be differentially impacted by different combinations of E2a-Pbx1-Hox heterodimers. In the presence of multiple Hox proteins, the heterodimer that bound with the highest affinity would exhibit the greatest impact on transcription of all target genes. Now, however, we would suggest that multiple groups of genes are each regulated by distinct optimal Pbx1-Hox DNA motifs, and that E2a-Pbx1, in the presence of multiple Hox proteins, would target all genes regulated by each class of Hox gene expressed. A potential reason why E2a-Pbx1 may be associated with pre-B cell leukemia rather than pro-B cell leukemia could be that its oncogenic potential is realized only when an appropriate partner or combination of partners is expressed, which might not occur until the pre-B cell stage of differentiation.

Because some *in vitro* regulatory motifs apparently bind Hox monomers, conforming to the sequence TAAT-GG GA, TA or TG, while others bind Hox-Pbx1 heterodimers, an unanswered question is what are the different functions of Hox monomer and Hox-Pbx heterodimers. Most of the biologically relevant sites described to date in both Drosophila and higher eucaryotes are monomer sites, such as the autoregulatory TAATGA motifs in the Drosophila Deformed gene (Regulski *et al*., 1991), and the autoregulatory TAATGG sites in the HoxD4 promoter (Popperl and Featherstone, 1992). One possibility is that heterodimer formation with Pbx1 simply represents a method to recruit two independent functions of the Hox and Pbx proteins to the same promoter. Alternatively, heterodimerization with Pbx1 could alter the transcriptional activity, as well as the DNA-binding activity, of the Hox protein, targeting this new activity to a subset of cellular promoters.

We suggest two possible mechanisms to explain the fact that unique sequences in the NTA of a Hox protein alter the specificity of DNA-binding from that observed for a Hox monomer (TAAT core) to that observed for a Pbx1-Hox or E2a-Pbx1-Hox heterodimer (TTAT or TGAT). Because the Hox pentapeptide is located just upstream of the Hox NTA, dimer formation may alter the DNA-binding surface of Pbx1, the Hox protein, or both, resulting in altered interaction with the first two bases of the Hox core. We have argued that residue 50 of Pbx1, which would normally bind a dinucleotide 3′ to the Pbx1 core, may not interact with DNA because it is glycine, and therefore does not extend an R group that could bind these positions. However, structural changes at the dimer interface induced by unique NTA sequences of individual Hox proteins could reposition helix 3 of Pbx1 within the major groove, permitting base contacts by residues adjacent to Gly50. Alternatively, dimer formation with Pbx1 could reconfigure the NTA of Hox proteins such that Arg3 and Arg5, which lie in the minor groove opposing helix 3 of Pbx1, are now repositioned to bind a different nucleotide at position 6 of the Pbx-Hox motif (Figure 1b). Finally, specific DNA-binding by heterodimers could arise from new contacts by both Pbx1 and the Hox protein. The crystal structure of a Pbx1-Hox-DNA complex should reveal the unique mechanism of how protein interactions at the dimer interface establish DNA-binding specificity.

Materials and methods

Construction of GST-HoxA5, GST-HoxB8, GST-HoxC8 and Histidine-tagged Pbx1

His6-Pbx1 was constructed by mutating the ATG start codon of Pbx1 into a BamH1 site, and subcloning in frame with the His6 tag of pRsetB (Invitrogen). GST-Hox plasmids were made by mutating the first codon of each cDNA into BamHI or BglII site, and subcloning in frame with GST into pGEX2TK (Promega).

Purification of recombinant proteins

For purification of recombinant Pbx and Hox proteins, 20 ml of B121 cells containing the GST-Pbx1 or polyHis-Hox expression vectors was grown at 37°C to an A_{600} of 0.5. Expression of recombinant proteins was induced by addition of 0.2 mM IPTG for an additional 3 h. Cells were collected by centrifugation, lysed by three brief sonications (15 s each) in 1 ml of 25 mM Tris (pH 7.5), 1 mM EDTA, 150 mM NaCl (Buffer B). Insoluble material was removed by centrifugation in a 1.5 ml Eppendorf tube and the soluble fraction was added to 100 μl glutathione-agarose beads or Nickel resin pre-equilibrated with Buffer B at 4°C and mixed for 2 h at 4°C. The resins were then washed seven times with 1.0 ml Buffer B. GST fusion proteins were eluted with 200 μl of 10 mM glutathione and dialysed against PBS. His-tagged Hox proteins were used as immobilized conjugates.

Selection of Pbx1-Hox binding sites

Optimal DNA motifs were selected using a two-step affinity purification. 50 picomoles of a population of double-stranded 70mers containing 20 random internal nucleotides flanked by *m*ultiple *c*loning *s*ites (MCS) were synthesized from a single-stranded template, a 3′ primer complimentary to the MCS, and PCR extension. A second cycle of PCR, using both 5′ and 3′ oligos complementary to the MCS was

used to amplify the library twofold. The amplified, double-stranded oligonucleotides were incubated for 1 h at 4°C with 25 μl recombinant His6-Pbx1b, immobilized on Ni-affinity resin, 2–10 μg of GST-Hox protein, 25 μg poly(dI-dC), in a total volume of 300 μl of buffer A (1 μg/ml BSA, 0.1% NP-40, 20 mM imidazole in PBS). Unbound oligonucleotides were removed by washing the resin six times in 500 μl of buffer A, and bound oligonucleotides were removed by incubation for 15 min at 4°C with 300 μl of buffer B (1 μg/ml BSA, 0.1% NP-40, 100 mM imidazole in PBS). The eluate was then incubated with 25 μl glutathione affinity resin at 4°C for 30 min. The resin was washed six times with PBS and bound oligonucleotides were released by boiling in 50 μl water. 20% of the eluted oligonucleotides were subjected to 15 cycles of amplification by PCR. This DNA was then used for the next round of double-selection/amplification. After six rounds of double selection, oligonucleotides were cloned and sequenced.

Site-directed mutagenesis

Point mutations and deletions were constructed using the Muta-Gene system (Bio-Rad) of oligonucleotide-directed mutagenesis of single-stranded uridine-containing template DNA according to the manufacturers instructions, as originally described (Ausubel *et al*., 1989; Kunkel *et al*., 1987).

In vitro *transcription/translation*

In vitro transcription/translation was performed using the Promega TNT Coupled Reticulocyte Lysate System, in accordance with manufacturer's protocol and employing SP6 polymerase.

Electrophoretic mobility shift assays (EMSA)

Double-stranded oligonucleotides were labeled with [^{32}P]ATP by phosphorylating a short oligonucleotide that annealed to the 3′ portion of the binding-site oligonucleotide, and then synthesizing the complementary strand using dNTP's and Klenow. All probes had the sequence TCACGGTGATTAATGAGCGACTGCTCGG and varied only at the two nucleotides underlined, which comprises the first two positions of the Hox core. For EMSA, 15 000 c.p.m. of probe was incubated with 2–4 μl of *in vitro* translated proteins of appropriate amount of recombinant proteins in the presence (for *in vitro* translated proteins) or absence (for purified recombinant protein) of 1 μg of poly (dI-dC) : (dI-dC) in a buffer containing 10 mM Tris (pH 7.5), 1 mM EDTA, 1 mM DTT, 2.5 mM $MgCl_2$, 0.1% NP-40, and 5% glycerol for 20 min at room temperature. Bound and free probe were separated by electrophoresis in 5–8% acrylamide gels formed in 0.5 × TBE and run in the same buffer. After drying the gel, the protein-DNA complexes were visualized by autoradiography. Complex abundance was measured on a Bio-Rad GS-250 Molecular Imager. Off-rates were calculated by fitting an exponential decay to the abundance of complex remaining after addition of a 1000-fold molar excess of competitor. In experiments determining the DNA-binding specificities of Hox-Pbx complexes, the molarity of the Pbx and Hox proteins were not measured and no attempt was made to add equal concentrations of Hox proteins. For this reason no claims are made about the varying affinities of different Pbx-Hox complexes for the same DNA probe; rather, the only valid determination is that of relative affinities of the same Pbx-Hox complex for different DNA probes. This is appropriate because differences in the specificities of Pbx-Hox complexes can be measured accurately at a range of protein concentrations below the Kd for complex formation.

Acknowledgements
This work was supported by Public Health Services Grants CA56876-05 and 2 PO1 CA50528. MPK is a Pew Scholar in the Biomedical Sciences.

References

Ausubel FM, Brent R, Kingston RE, Moore DD, Seidman JG, Smith JA and Struhl K. (1989). *Current Protocols in Molecular Biology*, **1,** 8.1.1–8.1.6.
Carroll A, Crist W, Parmley R, Roper M, Cooper M and Finley W. (1984). *Blood*, **63,** 721–724.
Chan SK, Jaffe L, Capovilla M, Botas J and Mann R. (1994). *Cell*, **78,** 603–615.
Chan SK and Mann RS. (1996). *Proc. Natl. Acad. Sci. USA*, **93,** 5223–5228.
Chang CP, Shen WF, Rozenfeld S, Lawrence HJ, Largman C and Cleary ML. (1995). *Genes & Dev.*, **9,** 663–674.
Dedera DA, Waller EK, Lebrun DP, Sen-Majumbar A, Stevens ME, Barsh GS and Cleary ML. (1993). *Cell*, **74,** 833–843.
Furukbo-Tokunaga K, Flister S and Gehring WJ. (1993). *Proc. Natl. Acad. Sci. USA*, **90,** 6360–6364.
Gehring WJ, Affolter M and Bürglin T. (1994a). *Annu. Rev. Biochem.*, **63,** 487–526.
Gehring WJ, Qian YQ, Billeter M, Furukubo-Tokunaga K, Schier AF, Resendez-perez D, Affolter M, Otting G and Wüthrich K. (1994b). *Cell*, **78,** 211–223.
Johnson FB, Parker E and Krasnow MA. (1995). *Proc. Natl. Acad. Sci.*, **92,** 739–743.
Kamps MP, Murre C, Sun X and Baltimore D. (1991). *Cell*, **60,** 547–555.
Kamps MP and Baltimore D. (1993). *Mol. Cell. Biol.*, **13,** 351–357.
Kamps MP and Wright D. (1994). *Oncogene*, **9,** 3159–3166.
Kamps MP, Wright DD and Lu Q. (1996). *Oncogene*, **12,** 19–30.
Kissinger CR, Liu BS, Martin-Blabco E, Kornberg TB and Pabo CO. (1990). *Cell*, **63,** 579–590.
Knoepfler P and Kamps MP. (1995). *Mol. Cell. Biol.*, **15,** 5811–5819.
Knoepfler P, Lu Q and Kamps M. (1996). *Nuc. Acid. Res.*, **24,** 2288–2294.
Krumlauf R. (1994). *Cell*, **78,** 191–201.
Kunkel TA, Roberts JD and Zakour RA. (1987). *Meth. Enzymol.*, **154,** 367–378.
Laughon A. (1991). *Biochemistry*, **30,** 11357–11367.
Lawrence H, Stage K, Mathews C, Detmer K, Scibienski R, MacKenzi M, Migliaccio E, Boncinelli E and Largmen C. (1993). *Cell Growth and Differentiation*, **4,** 665–669.
Lebrun DP and Cleary ML. (1994). *Oncogene*, **9,** 1641–1647.
Lu Q, Wright DD and Kamps MP. (1994). *Mol. Cell. Biol.*, **6,** 3938–3948.
Lu Q, Knoepfler P, Scheele J, Wright DD and Kamps MP. (1995). *Mol. Cell. Biol.*, **15,** 3786–3795.
Lu Q and Kamps MP. (1996). *Mol. Cell. Biol.*, **16,** 1632–1640.
Mathews C, Detmer K, Boncinelli E, Lawrence H and Largmen C. (1991). *Blood*, **78,** 2248–2252.
McGinnis N, Kuziora M and McGinnis W. (1990). *Cell*, **63,** 969–976.
Monica K, Galili N, Nourse J, Saltman D and Cleary ML. (1991). *Mol. Cell. Biol.*, **11,** 6149–6157.

Neuteboom STC, Peltenburg LTC, Van Dijk MA and Murre C. (1995). *Proc. Natl. Acad. Sci. USA*, **92,** 9166 – 9170.
Nourse J, Mellentin JD, Galili N, Wilkinson J, Stanbridge E, Smith SD and Cleary ML. (1990). *Cell*, **60,** 535 – 545.
Peers B, Sharma S, Johnson T, Kamps MP and Montminy M. (1995). *Mol. Cell. Biol.*, **15,** 7091 – 7097.
Petrini M, Quaranta M, Testa U, Samoggia P, Tritarelli E, Care A, Cianetti L, Valtieri M, Barletta C and Peschle C. (1992). *Blood*, **80,** 185 – 193.
Pöpperl H, Bienz M, Studer M, Chan SK, Aparicio S, Brenner S, Mann RS and Krumlauf R. (1995). *Cell*, **81,** 1031 – 1042.
Pöpperl H and Featherstone M. (1992). *EMBO J.*, **11,** 3673 – 3680.
Regulski M, Dessain S, McGinnis N and McGinnis W. (1991). *Genes & Dev.*, **5,** 278 – 286.
Van Dijk MA, Voorhoeve P and Murre C. (1993). *Proc. Natl. Acad. Sci. USA*, **90,** 6061 – 6065.
Van Dijk MA and Murre C. (1994). *Cell*, **78,** 617 – 624.
Vieille-Grosjean I, Roullot V and Courtois G. (1992). *BBRC*, **183,** 1124 – 1130.
Zeng W, Andrew DJ, Mathies LD, Horner MA and Scott MP. (1993). *Development*, **118,** 339 – 352.