2288–2294 *Nucleic Acids Research, 1996, Vol. 24, No. 12*

Pbx1–Hox heterodimers bind DNA on inseparable half-sites that permit intrinsic DNA binding specificity of the Hox partner at nucleotides 3′ to a TAAT motif

Paul S. Knoepfler*, Qiang Lu and Mark P. Kamps

Department of Pathology, University of California, San Diego, School of Medicine, 9500 Gilman Drive, La Jolla, CA 92093, USA

Received March 11, 1996; Revised and Accepted April 22, 1996

ABSTRACT

Heterodimers between the Pbx/Exd and Hox/HOM-C classes of homeodomain proteins bind regulatory elements in tissue-specific and developmentally regulated genes. In this work, we characterize the half-site bound by both Pbx1 and Hox proteins on a prototypic element (TGATTAAT) and determine how the orientation of the Hox protein contributes to the DNA binding specificity of Pbx–Hox heterodimers. We demonstrate that the Hox protein binds the 3′ TAAT sequence as its recognition core and exhibits sequence-specific binding at positions 3′ to the TAAT core. Unfavored sequences at this position, such as two cytosines, abrogate binding to the element. The upstream Pbx1 core sequence, TGAT, must immediately juxtapose the Hox core. This geometry maintains the preference of Hox/HOM-C proteins for a T base at position –1, as T represents the fourth position of the Pbx1 core, and suggests that this T base is bound by both Pbx1 and Hox proteins, Pbx1 binding in the major grove and the Hox protein binding in the minor grove. Pbx1 also exhibits base selectivity 5′ to its TGAT recognition sequence.

INTRODUCTION

Homeobox genes regulate cell fate and segmental patterning (1–4) and encode a conserved DNA binding motif known as the homeodomain (HD; 4). HD proteins act with great biological specificity in development and their functional differences are largely mediated by their HDs (5–8). The vertebrate Hox proteins and *Drosophila* HOM-C proteins are a subset of HD proteins that specify structural development along the anterior–posterior (A-P) axis (1,3,9) through specific recognition of DNA motifs in target genes (10–13). Hox and HOM-C proteins share structural and functional features, including the primary sequence of their HDs, the chromosomal organization of their respective genes and their spatial and temporal expression patterns along the A-P axis (1,3,9). Mutations in these genes induce specific segments to develop morphological structures of another.

Despite the large number of different biological functions of homeoproteins, most HDs bind DNA motifs containing a TAAT core (12,14–16). Within the three α-helices and flexible N-terminal arm that comprise the HD (17), basic residues at positions 3 and 5 bind the 5′ TA, and Asp51 and a hydrophobic residue at position 47, both of which lie in helix 3, bind the 3′ A and T bases of the core, respectively (18,19). Residues in helix 3 make sequence-specific contacts 3′ of the TAAT core that permit different HD proteins to discriminate among related binding sites. When residue 50 is Lys, two consecutive cytosines are strongly favored immediately 3′ to the TAAT core (20,21), but when residue 50 is Gln two guanines are favored. Residue 50 independently designates this specificity, because interconversion of Gln and Lys at position 50 in Bicoid or in Hox/HOM-C proteins interconverts DNA binding specificity between TAATGG and TAATCC (20,21).

While all Hox/HOM-C proteins encode Gln at residue 50, they maintain different intrinsic specificities for sequences positioned up to four bases 3′ to the TAAT core. For example, Ubx binds TAATGGCT with a 6-fold slower off-rate than TAATCGAC and activates transcription 50-fold better on the TAATGGCT element (14). By contrast, Dfd binds TAATGGCT with only a 50% slower off-rate than for TAATCGAC and activates transcription only 2.6-fold better via the TAATGGCT element. This difference is due to the fact that Ubx prefers a C three bases 3′ to the TAAT core, while Dfd prefers an A, and this specificity is intrinsic to helix 3 (14). Differential binding mediated by the unstructured N-terminal arm (14,22) creates binding specificity at positions 5′ to the TAAT core, however, the magnitude of these effects are smaller than those that occur for bases 3′ to the TAAT and the major effect is limited to the identity of bases at the –1 position (14,23).

Members of the Pbx/Exd family of HD proteins cooperate with Hox proteins to regulate target gene expression and bind DNA cooperatively with many Hox and HOM-C proteins (24–28). The Pbx/Exd family includes the human proto-oncoprotein Pbx1 (29–30), its mammalian homologs, Pbx2 and Pbx3 (31), *Drosophila* Extradenticle (Exd; 32) and *Caenorhabditis elegans* ceh-20 (33). The importance of Pbx1 in differentiation is also implicated by the fact that E2A-Pbx1, an oncogenic and transcriptionally activated derivative of Pbx1 found in human

* To whom correspondence should be addressed

pre-B cell leukemia (29,30), can block normal hematopoietic differentiation (34–36). Cooperative binding of Pbx and Hox proteins was first described on TTGATTGAT (26,37), which was derived from selection experiments using degenerate oligonucleotides and recombinant Pbx proteins. Heterodimer formation requires only the HD and adjacent C-terminal 17 residues of Pbx1 (38) and only the HD and N-terminal sequences sufficient to include an essential pentapeptide sequence (YPWMR) of Hox proteins (25,39,40). Because transfer of the Hox pentapeptide sequence onto the HD of Pbx1 causes two molecules of Pbx1 to exhibit cooperative binding to TTGATTGAT, this element was likely initially isolated due to binding of Pbx1 as a dimer (38). Thirty two genes of the Hox loci that are either known or proposed to exist (A1–A8, B1–B8, C1–C8 and D1–D8) encode HD proteins containing variants of this pentapeptide sequence and thus are predicted to be able to heterodimerize with Pbx1 on appropriate DNA elements. In addition, the HD protein STF-1, which is not encoded by the Hox loci and is expressed in the pancreas and small intestine, also contains a pentapeptide sequence (41) and binds TTGATTGAT cooperatively with Pbx1 (42).

Elements similar to TTGATTGAT are bound by Hox/Pbx heterodimers *in vivo* and are required for expression of tissue-specific genes. The Hox B1 promoter contains TGATGGAT and AGATTGAT elements that are essential for stage-specific and tissue-specific expression and bind Hox B1 and Exd (Table 1; 27). In the somatostatin promoter a Pbx–STF-1 heterodimer binds TGATTAAT (42), a site important for expression of somatostatin in pancreatic islet-cells (41). This element can also be activated synergistically by STF-1 and E2A-Pbx1 (42). Cooperative DNA binding by Pbx and Hox proteins raises the question of how each heterodimer partner is oriented on its half-site, whether these half-sites are separable and whether heterodimer formation alters the DNA binding specificity of Hox proteins. In this work, we address these issues.

We demonstrate that the Hox protein binds to these elements by utilizing the 3′ TGAT or TAAT sequence as its recognition core, that its HD assumes the same orientation on DNA as that of Hom-C proteins and that it exhibits sequence-specific contacts at positions 3′ to sequences currently considered to comprise such elements. The upstream Pbx1 core sequence, TGAT, must be immediately juxtaposed to the Hox core. This geometry maintains the preference of Hox/HOM-C proteins for a T residue at position –1 (14) because T represents the fourth position of the Pbx1 core. Pbx1 also exhibits additional 5′ base selectivity. We suggest that an internal portion of the Pbx–Hox element is bound by both partners, Pbx1 approaching from the major grove and the Hox protein binding in the minor grove.

MATERIALS AND METHODS

Construction of recombinant plasmids

All cDNAs were cloned in vectors pGEM 3Z F$^-$, pGEM 4Z or pGEM 3Z (Promega).

Mutagenesis

Residue 50 of the HD of Hox A5 and Hox B8 cDNAs was changed to Lys using the Muta-gene Phagemid *In Vitro* Mutagenesis kit (BioRad) according to the manufacturer's protocol. All mutations were verified by DNA sequence analysis.

Table 1. DNA probes used in these studies, including numerical designations (a comparison of these sequences to known promoter elements is listed at the bottom)

DNA probes for experiments

Probe 1	TCAC- G T - T G A T T A A T - **G A** -G CGACTGCTCGG
Probe 2	TCAC- G T - T G A T T A A T - **G G** -G CGACTGCTCGG
Probe 3	TCAC- G T - T G A T T A A T - **G T** -G CGACTGCTCGG
Probe 4	TCAC- G T - T G A T T A A T - **G C** -G CGACTGCTCGG
Probe 5	TCAC- G T - T G A T T A A T - **A A** -G CGACTGCTCGG
Probe 6	TCAC- G T - T G A T T A A T - **A G** -G CGACTGCTCGG
Probe 7	TCAC- G T - T G A T T A A T - **A T** -G CGACTGCTCGG
Probe 8	TCAC- G T - T G A T T A A T - **A C** -G CGACTGCTCGG
Probe 9	TCAC- G T - T G A T T A A T - **T A** -G CGACTGCTCGG
Probe 10	TCAC- G T - T G A T T A A T - **T G** -G CGACTGCTCGG
Probe 11	TCAC- G T - T G A T T A A T - **T T** -G CGACTGCTCGG
Probe 12	TCAC- G T - T G A T T A A T - **T C** -G CGACTGCTCGG
Probe 13	TCAC- G T - T G A T T A A T - **C A** -G CGACTGCTCGG
Probe 14	TCAC- G T - T G A T T A A T - **C G** -G CGACTGCTCGG
Probe 15	TCAC- G T - T G A T T A A T - **C T** -G CGACTGCTCGG
Probe 16	TCAC- G T - T G A T T A A T - **C C** -G CGACTGCTCGG
5'	
probe 17	TCAC- G **A** - T G A T T A A T - G G -G CGACTGCTCGG
probe 18	TCAC- G **C** - T G A T T A A T - G G -G CGACTGCTCGG
probe 19	TCAC- G **G** - T G A T T A A T - G G -G CGACTGCTCGG
probe 20	TCAC- **A** A - T G A T T A A T - G G -G CGACTGCTCGG
probe 21	TCAC- **T** A - T G A T T A A T - G G -G CGACTGCTCGG
probe 22	TCAC- **C** A - T G A T T A A T - G G -G CGACTGCTCGG
Probe 23	TCAC- G G - T G A T T **G** A T - **C C** -G CGACTGCTCGG
Probe 24	TCAC- G G - T G A T T **G** A T - **G G** -G CGACTGCTCGG

Hox positions:	-1 1 2 3 4 5 6 7 8 9
Pbx positions:	-1 1 2 3 4 5 6 7
Hox-Pbx positions:	-1 1 2 3 4 5 6 7 8 9 10 11 12 13

In vivo Pbx binding sites

Hox B1 promoter Repeat 1	C - A G A T G G A T - G G
Hox B1 promoter Repeat 3	T - T G A T G G A T - G G
Somatostatin promoter Element TSEII	A - T G A T T A A T - T A
CYP17 $P450_{17\alpha}$ promoter Element CRE	T - T G A T G G A C - A G
Position designations:	1 2 3 4 5 6 7 8 9 10 11

In vitro transcription/translation

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

Electrophoretic mobility shift assays (EMSA)

Double-stranded oligonucleotides were labeled with [^{32}P]ATP to the same specific activities by phosphorylating a common reverse oligonucleotide that was annealed to the oligonucleotides shown in Table 1 and made double-stranded by addition of dNTPs and the Klenow fragment of DNA polymerase I. Bound and free probe were separated by electrophoresis in native 5% acrylamide gels formed in 0.5× TBE (27 mM Tris, 27 mM boric acid, 0.6 mM EDTA) and run in the same buffer. Binding reactions contained 20 000 c.p.m. probe, 4 µl *in vitro* translated proteins, 1 µg poly(dI·dC), 10 mM Tris, pH 7.5, 1 mM EDTA, 1 mM dithiothreitol, 0.1% Nonidet P-40 and 5% glycerol. Reactions were incubated for 30 min at 23°C. For Hox A5 and Hox B8 proteins containing point mutations at residue 50 of the HD, equivalent amounts of wild-type and mutant proteins were used

in binding reactions, as determined by labeling during transcription/translation reactions using [^{35}S]methionine.

RESULTS

Sequence-specific binding of Pbx1–Hox heterodimers suggest that the Hox protein binds the 3′ TAAT motif

The half-sites bound by Pbx1 and Hox proteins on a single element have only been partially characterized. DNase I protection analysis suggests that Pbx1 is bound 5′ to Hox proteins (43). Mutation of the 5′ G in TGATTGATT to A abrogates cooperative DNA binding, while mutation of the 3′ G to A maintains cooperativity, consistent with the proposal that Pbx1 binds a 5′ half-site containing TGAT and that the Hox protein binds a 3′ half site consisting of either TGAT or TAAT. On each proposed half-site these core nucleotides are numerically designated 2–5, and for the combined Pbx/Hox site that we propose they are designated 2–5 and 6–9 for Pbx1 and Hox proteins, respectively (Table 1). Hox proteins exhibit negative selection for CC at the two positions 3′ to their TAAT core and positive selection for G or T at the first 3′ base (6 in the monomeric site) and either G or A at the second 3′ base (7 in the monomeric site; 14,15,23). Therefore, if the Hox protein binds the 3′ TAAT or TGAT core in the Pbx1–Hox heterodimer site, positions 10 and 11 should prove crucial for strong binding of this element by either Hox monomers or Hox/Pbx heterodimers.

To develop a model system in which we could observe the effect of sequence alterations on the binding of both monomeric Hox proteins as well as Pbx1–Hox heterodimers, we selected the Pbx/STF-1 binding site contained in the somatostatin promoter (TGATTAAT), which contains an optimal TAAT core. Another TAAT core sequence is encoded by the opposite strand, however, this sequence is followed by a CA dinucleotide which creates an unfavorable Hox monomer binding site. If changes at positions 10 and 11 of the combined Pbx1/Hox binding site altered both Hox and Hox–Pbx1 binding, it would be very unlikely that the Hox protein was bound to the reverse TAAT core because nucleotides located 3–4 bases 5′ of a TAAT site have little to no effect on Hox binding. Due to the co-migration of the Pbx–Hox–DNA complex with a non-specific complex produced by the binding of a factor in reticulocyte lysate, we used the larger E2A-Pbx1 fusion form of Pbx1 for these experiments. E2A-Pbx1 contains all the determinants required for cooperative DNA binding with Hox proteins and, in a previous study, E2A-Pbx1 and Pbx1 were found to exhibit indistinguishable behavior in their cooperative binding to DNA with Hox B7, Hox B8, Hox A5 and STF-1 (43). DNA probes containing GG, TA or CC at positions 10 and 11 exhibited dramatic differences in their abilities to bind monomers of Hox B7 (lanes 5–7), Hox B8 (lanes 8–10), Hox A5 (lanes 11–13) and STF-1 (lanes 14–16) or heterodimers of E2A-Pbx1 and each Hox protein (Fig. 1). Like Pbx1, E2A-Pbx1 failed to bind probe in the absence of a Hox protein (lane 1) and each of the Hox proteins alone failed to form the upper complex in the absence of E2A-Pbx1 (lanes 2–4). This suggested that the Hox protein bound the 3′ TAAT sequence (positions 6–9) in a manner consistent with the sequence-specific DNA binding characteristics of Hox proteins.

Next, the preference of Hox B7, Hox B8 and Hox A5 for all 16 combinations of dinucleotides at positions 10 and 11 was examined. The known preferences for the HOM-C HD proteins, Ubx and Dfd, are G > T > A > C at position 6 and A,G > C,T at position 7 (14). Likewise, naturally occurring Hox D4 binding sites, TAATGG and TAATGA (44), and Hox A5 binding sites, TAATGG, TAATGA (45), exhibit a similar preference. The abundance of monomeric complexes containing Hox B7 and Hox B8, as well as all dimeric complexes containing E2A-Pbx1 plus Hox proteins, paralleled these preferences remarkably. For residue 10 the preferred order of nucleotides was G > T > A > C (lanes 1–4 > lanes 9–12 > lanes 5–8 > lanes 13–16) and for position 11 it was G > A,T > C (Fig. 2). The only consistent exception to this rule occurred when A at position 10 followed A at position 11, in which case binding was poorer than for the AC dinucleotide (lane 5 < lane 8). This analysis strongly supports the hypothesis that each Hox protein bound the 3′ TAAT sequence in the same orientation as Engrailed (Eng) and Antennapedia (Ant) bind DNA, based on their crystal structures (17,18).

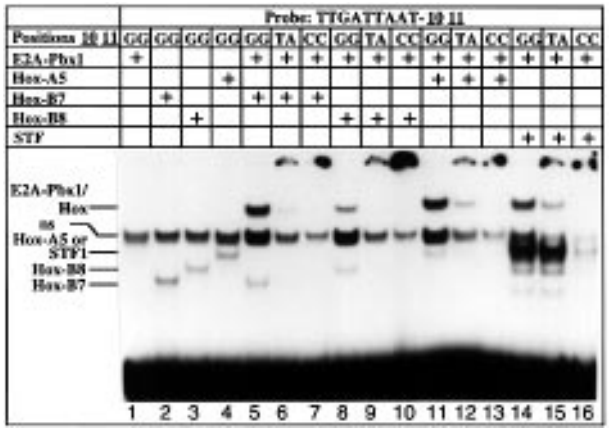

Figure 1. Bases 3′ to the TAAT core alter the binding affinity of E2A-Pbx1–Hox heterodimers. The binding affinity of Hox proteins and STF-1 were examined as monomers and as heterodimers with E2A-Pbx1, using EMSA. The identity of the probe, which varied only at positions 10 and 11, is indicated at the top. E2A-Pbx1 and Hox proteins were generated by coupled transcription/translation and STF was produced as a recombinant protein in bacteria. Proteins added to each binding reaction are indicated by a + sign above their respective lanes and the sequence of the complete probes are indicated in Table 1. NS refers to a non-specific band, which arises from the binding of some factor in the reticulocyte lysate to a sequence in the probe and which does not supershift with antibodies to Pbx1.

Residue 50 of the Hox HD specifies the identity of bases bound at positions 5 and 6, affirming the proposed position of the Hox protein

In the crystal structures of Eng and Ant, N51 contacts the A at position 4 and residue 50 makes base-specific contacts at positions 6 and 7 (17,18). In Bicoid, residue 50 is Lys and specifies CC at these positions, while HD proteins encoded by the HOM-C/Hox gene contain Gln at this residue (20). In order to prove that the Hox protein in the E2A-Pbx1–Hox complex was bound in the hypothesized orientation, Gln50 of Hox A5 was converted to Lys in order to provide biochemical evidence that this would alter binding and favor a TAATCC motif (Fig. 3). Indeed, the K50 mutation in Hox A5 eliminated monomeric Hox A5 and E2A-Pbx1–Hox A5 binding to TGATTAATGG (lane 3 versus 1) and revealed a new ability to bind TGATTAATCC (lane 4 versus 2). The same shift in binding specificity was observed

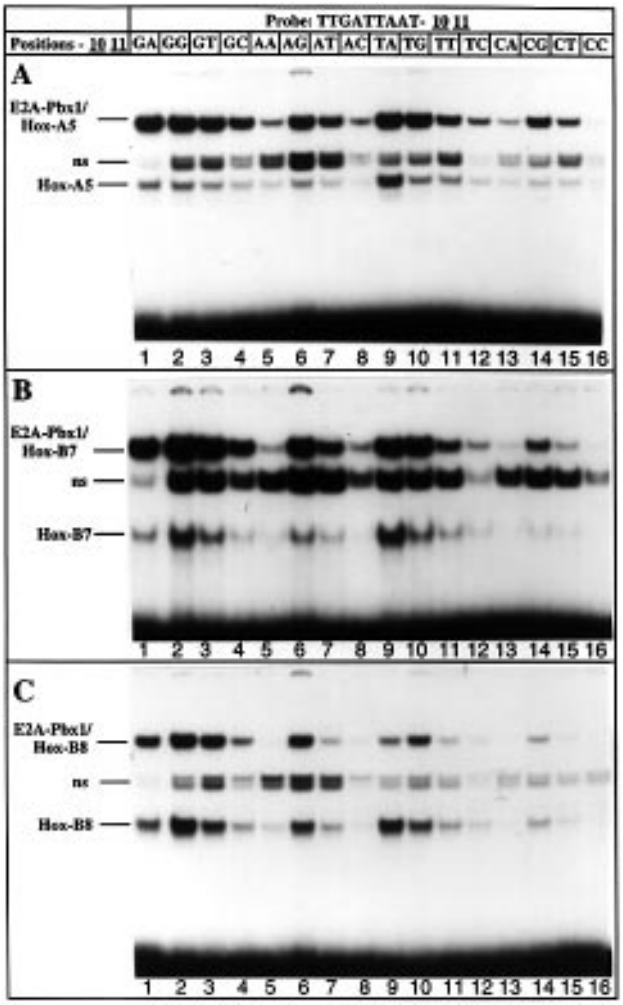

Figure 2. Analysis of the binding affinities of heterodimers of Hox and E2A-Pbx1 for DNA probes containing all dinucleotide combinations 3′ to TGATTAAT. All proteins were synthesized by coupled transcription/translation and 2 µl E2A-Pbx1 plus 2 µl specified Hox protein was added to all binding reactions and the complexes analyzed by EMSA. The identity of the probe, which varied only at positions 10 and 11, is indicated at the top. This analysis was performed using Hox A5 (**A**), Hox B7 (**B**) and Hox B8 (**C**).

when an identical analysis was performed with an oligonucleotide probe containing a TGAT Hox core sequence instead of a TAAT core (lanes 5–8). TGAT is a common half-site variant of Pbx–Hox heterodimer binding sites (Table 1). Mutation of Gln50 to Lys in Hox B8 produced the same effect on DNA binding preference as observed with Hox A5 (data not shown). These experiments demonstrate that when Hox proteins bind DNA elements with Pbx proteins they are positioned on a 3′ half-site in an orientation that permits Q50-specific contacts 3′ of the core sequence.

The binding of Pbx1 is influenced by bases 5′ to the TGAT core

Pbx1 is strongly implicated to bind DNA in the same orientation as the Hox protein binds its 3′ half-site, because fusion of the Hox pentapeptide N-terminal to Pbx1 causes it to bind the internally repeated sequence TGATTGATT as a strong and cooperative

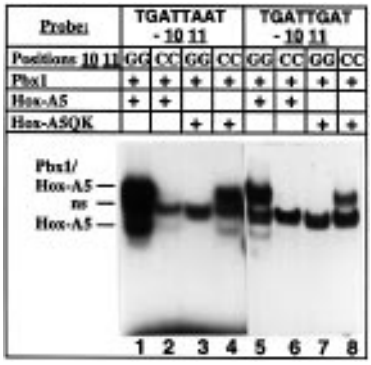

Figure 3. Residue 50 of the Hox HD specifies the identity of the dinucleotide pair 3′ to the TAAT core. Analysis of the binding specificity of wild-type Hox A5 and a mutant form containing a substitution of Lys for Gln at residue 50 of the HD (Hox A5 Q50K) was analyzed by EMSA, using probes containing either a GG or CC dinucleotide 3′ to the TAAT core. All proteins were synthesized by coupled transcription/translation and added to binding reactions as indicated by + signs. The identities of bases 10 and 11 of the heterodimer binding site are indicated above each lane.

homodimer (38). Thus, Pbx1 is likely to bind the upstream half-site core TGAT (bases 2–5), employing N51 to bind the invariant A at position 4 (designations in Table 1). Because HD proteins such as Ftz, Ubx and Dfd prefer C or T located just 5′ of their TAAT core site (15), we examined whether Pbx1 exhibited sequence specificity at this location (position 1 in the Pbx1/Hox site) and whether it was similar or different to that predicted for Hox proteins (Fig. 4). Changes at position 1 had a dramatic effect on the degree of cooperative binding with Hox proteins, A (lanes 2 and 6) being favored over T or G (lanes 1, 4, 5 and 8) and C permitting very little binding (lanes 3 and 7). By contrast, changes at position –1 had little effect on formation of the complex containing E2A-Pbx1 and Hox B7 or Hox A5, consistent with the observation that the –1 position has little effect on sequence-specific binding of HD proteins (lanes 9–16).

Spacing of the Pbx1 and Hox half-sites abrogates cooperative DNA binding by Pbx and Hox proteins

The cooperative binding of Pbx and Hox proteins requires an essential pentapeptide sequence that is positioned variably from the Hox HD (25,39,40). In Hox B8 it resides five amino acids N-terminal to the HD, while in Hox A5 it is positioned 15 residues N-terminal to the HD. Thus, interaction of Pbx1 with this pentapeptide sequence could be permitted for DNA elements containing spacing between the Pbx1 and Hox half-sites. Alternatively, precise spacing geometry between Pbx and Hox half-sites may be required in order to achieve formation of the heterodimer. To test directly whether these defined half-sites can be separated, a T, G or GG sequence was inserted between residues 5 and 6. The T preserves the Hox preference for T at its position 1 and maintains a possible preference of Pbx1 for T at position 6. All insertions completely abrogated formation of the E2A-Pbx1–Hox complexes, indicating that the Pbx1 and Hox half-sites cannot be separated (Fig. 5). Failure to form heterodimers was not due to diminished binding of the Hox protein, because insertion of a single T actually enhanced the binding of Hox A5.

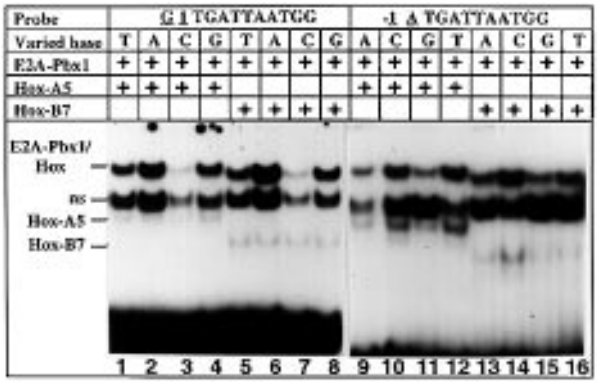

Figure 4. Residues 5′ to the TGAT core influence the binding affinity of E2A-Pbx1–Hox heterodimers. The importance of the dinucleotide 5′ to the Pbx1 TGAT core was examined by EMSA, using probes that vary at these positions and combinations of E2A-Pbx1 with either Hox A5 or Hox B7. All proteins were synthesized by coupled transcription/translation and were added as indicated above each lane. The probes used in lanes 1–8 contain G-1-TGATTAATGG, where only position 1 is varied as indicated. The probes used in lanes 9–16 contain –1-A-TGATTAATGG, where only position –1 is varied.

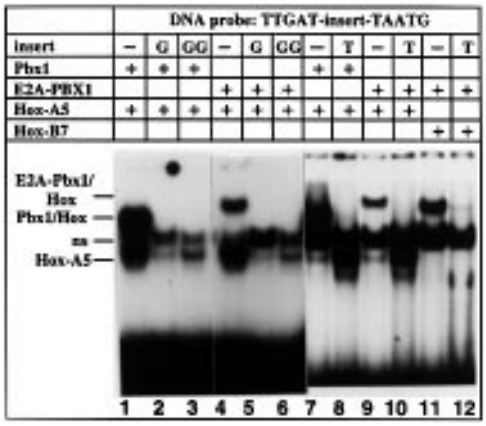

Figure 5. Spacing of the Pbx1 and Hox half-site core sequences abrogates heterodimer formation by Pbx and Hox proteins. The effect of separating the core sequences bound by Pbx1 and Hox proteins was examined by EMSA, using probes containing no insert (lanes 1, 4, 7, 9 and 11) or insertions of G (lanes 2 and 5), T (lanes 8, 10 and 12) or GG (lanes 3 and 6). Combinations of Pbx1 or E2A-Pbx1 and Hox A5 or Hox B7, produced by coupled transcription/translation, were added as indicated by the + signs above the lanes. All probes contained the TTGAT 5′ half-site and the TAATG 3′ half-site. Autoradiography for lanes 1–6 was 16 h and for lanes 7–12 60 h.

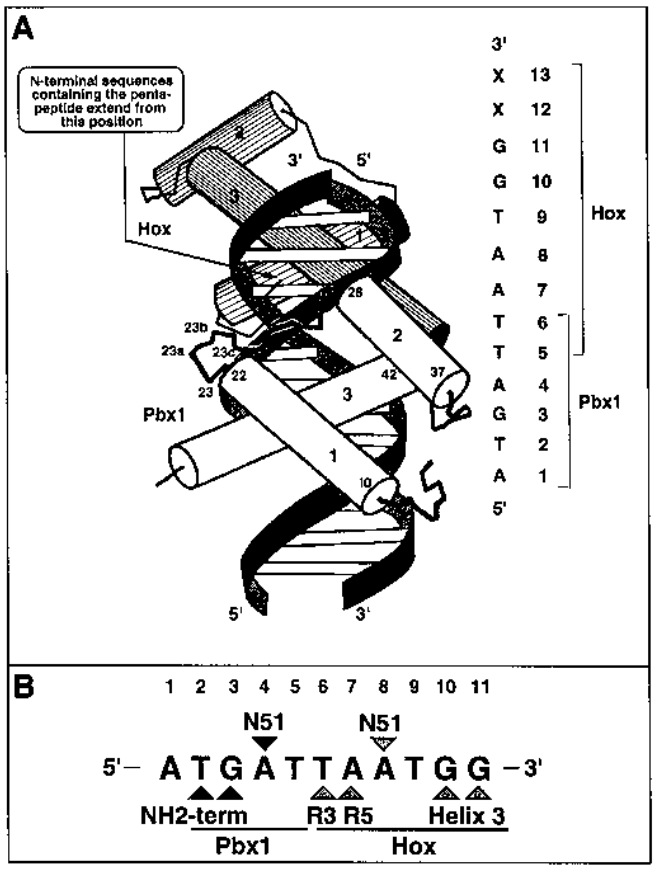

Figure 6. Proposed model of spacing and orientation of Pbx1 with Hox proteins in a heterodimer bound to the DNA sequence ATGATTAATGGXX. Numerical designations of nucleotide positions are identical to those in Table 1. XX designates bases known to influence DNA binding affinity of Hox monomers when the Hox HD is bound as indicated.

DISCUSSION

In this paper we define the binding orientation of Hox proteins in Pbx–Hox complexes and the sequence specificity of each partner at positions 5′ and 3′ to its half-site. The binding of E2A-Pbx1 and Hox proteins on DNA provides sequence-specific contacts in at least 11 positions (Table 1 and Fig. 6A). Mutation of Q50 to K50 in Hox proteins altered specificity of positions 10 and 11 to CC and binding analysis on all 16 dinucleotide possibilities mirrored the specificity of monomeric binding by Hox B7 and Hox B8 and the general specificity of Hox and HOM-C proteins characterized by others (14,44–46). Thus, the Hox protein binds the TAAT core spanned by positions 6–9 using the same conformation as that used by Ant or Eng, based on crystallographic studies (17,18). Dimer formation of Pbx with Hox A5, Hox B7 and Hox B8 on DNA thus permits the Hox protein to exhibit its major sequence-specific contacts. In this conformation N51 would bind A at position 8 and the N-terminal arm of the HD would contact bases 5–7 (Fig. 6B). Extension of Pbx1–Hox–DNA interactions to positions 10 and 11 is also consistent with earlier methylation interference studies, in which a heterodimer of E2A-Pbx1 and Hox A5 exhibited partial protection of guanine 10 and 12 in the DNA sequence TTGATTGATGCG (positions 10 and 12 underlined; 43). We would predict that all pentapeptide-containing Hox and HOM/C HD proteins that form heterodimers with Pbx proteins will exhibit this same geometry on DNA.

Pbx1 is proposed to bind TGAT (positions 2–5) as its core, with N51 recognizing A at position 4 (Fig. 6B). This orientation is based on the fact that fusion of a Hox pentapeptide sequence to the Pbx1 N-terminus produces a protein that exhibits strong cooperative binding to the repetitive element TGATTGAT (38), suggesting that adjacent Pbx1 proteins bind in the same orientation and that the Pbx1 pentapeptide fusion protein binds the 3′ site in the same orientation as would a Hox protein. Having

defined the orientation of Hox protein binding, we propose that Pbx1 binds in this same orientation. In this orientation, the specificity of Pbx1 would be very similar to that of yeast a1. When Pbx1 was first identified, the sequence of its HD exhibited the highest degree of amino acid identity with that of yeast a1 (19/30 residues in helices 2 and 3). Yeast a1 binds a very similar sequence (TGATGT-A/G) and forms cooperative heterodimers with a second HD protein, α2 (46). The crystal structure of an al–α2 heterodimer bound to DNA has recently been solved (47) and there are several similarities in binding properties to those proposed in our model for heterodimers of Pbx and Hox proteins. In the a1–α2 heterodimer, a1, like Pbx1, binds the 5′ half-site containing a TGAT. During heterodimerization with α2, residue N51 of a1 contacts the third base, A, of the 5′ half-site, which is the same HD–base contact occurring in our model of Pbx binding during heterodimerization with Hox proteins. Further, an unstructured C-terminal helix of α2 is responsible for making contact with the HD of a1, which parallels the interaction of the unstructured pentapeptide motif N-terminal to the HD of Hox proteins with the HD of Pbx1 during cooperative DNA binding. Finally, the crystal structure of al–α2 indicates that al and α2 bind in a head-tail/head-tail fashion, also consistent with the model of Pbx–Hox heterodimerization that we have proposed. However, while the Pbx and Hox half-sites are unspaced, those of a1 and α2 are separated by 6 bp (46) and a1 and α2 bind DNA on the same side of the double helix.

In contrast to the binding of a1 and α2, unspaced binding of Pbx1 and Hox to adjacent 4 bp cores would position each protein approximately on opposite sides of the double helix, in a conformation similar to that demonstrated by crystallographic studies of the binary $POU_{(homeo)}$ and $POU_{(specific)}$ DNA binding domains that typify the POU family of HD proteins (prototyped by the Brn-2, Oct-1 and Pit-1 transcription factors). The $POU_{(homeo)}$ domain has a structure almost identical to that of a Hox HD and the adjacent $POU_{(specific)}$ domain contains a structure similar to the HD but which is stabilized by a fourth helix. When Brn-2 binds the element GCATTAAT, the $POU_{(specific)}$ domain binds GCAT and the $POU_{(homeo)}$ domain binds TAAT (48), and both domains bind in the same orientation. Proteins in the POU family of transcription factors retain high affinity binding even when the individual half-sites are separated by up to 3 bp (48).

Unlike the opposing DNA binding domains of POU proteins, the Pbx1 and Hox half-sites could not be separated by 1 (T or G) or 2 (GG) bp without a dramatic loss in cooperative binding. This was not due to a reduction in the binding of the Hox protein, as evidenced by the fact that both Hox A5 and Hox B7 bound the site containing the T insertion as effectively as they did the parental site. The inseparability of the half-sites suggests that precise protein–protein interactions occur between each HD. Specific protein–protein interactions were also suggested by the behavior of mutations at Glu28 of the Pbx1 HD (38). An E28R mutation abolished cooperative DNA binding, maintained monomeric binding and maintained the ability of the HD to exhibit increased DNA binding upon addition of a synthetic Hox A5 pentapeptide sequence, suggesting that this mutation did not abrogate interaction with the pentapeptide, but eliminated a second direct interaction between the Pbx1 and Hox A5 HDs. In this regard, dimer formation between Pbx1 and Hox proteins is more similar to dimers of the *Drosophila* Paired HD, which binds cooperatively to a palindrome composed of two inverted TAAT sites separated by 2 bp and in which case cooperative DNA binding is drastically impaired by further separation of the TAAT sites (49).

The monomeric DNA binding specificity of Hox proteins has been examined previously (50,51) and those studies suggest that an optimal site that binds a Hox monomer is somewhat different to the optimal half-site that binds a Hox protein in the Pbx1–Hox heterodimer. Class one Hox proteins vary minimally in their monomeric DNA binding specificity, preferring C/G-TAATTG, with TG being a more optimal 3′ dinucleotide than GG (51). Although we find that a 3′ TG dinucleotide creates a good Pbx1–Hox heterodimer site, a 3′ GG nucleotide is clearly better for all Hox proteins tested. This may suggest that heterodimer formation with Pbx1 alters, to some degree, the specificity of DNA binding by helix 3 of the Hox HD at its contacts just 3′ to the core sequence. Similarly, the optimal nucleotides 5′ to the TAAT Hox monomeric core also differ from what we observe in an optimal Pbx1–Hox heterodimer. In the heterodimer the adjacent Pbx1 binding site presumably dictates the AT dinucleotide 5′ to the TAAT half-site. While this dinucleotide combination is invariantly retained in optimal DNA motifs selected by combinations of Pbx1 plus Hox proteins (Lu and Kamps, unpublished), the optimal dinuclotide at this position for Hox monomers is AC, and AT is disfavored (51). Furthermore, as observed in Figure 5, when the two half-sites of the Pbx1/Hox element were spaced by addition of a T, which destroys formation of the Pbx1–Hox–DNA complex, the resulting TT dinucleotide 5′ to the TAAT half-site strongly enhanced monomeric binding by Hox A5, suggesting that Hox A5 prefers a TT dinucleotide 5′ to its TAAT core when binding as a monomer. Thus, the site bound by Hox proteins in optimal Pbx1–Hox heterodimers is different to the optimal site bound by monomeric Hox proteins, and this may serve as one mechanism *in vivo* to prevent Hox proteins from binding as monomers to optimal Pbx1/Hox motifs.

Although Hox A5, Hox B7 and Hox B8 all display approximately the same preference for binding dinucleotides 3′ to the TAAT half-site during heterodimeric binding with Pbx1, there is now strong evidence that different Pbx1–Hox heterodimers do indeed bind different DNA motifs, however, the variation is not within the 3′ dinucleotide, but within the second position of the proposed Hox core (position 7 of Table 1 and Fig. 6). We have found that heterodimers of Pbx1 plus Hox A1 favor G at this position and bind much less tightly to probes containing an A or T at this position. By comparison, Hox D4 and Hox D5 form the most abundant complex with Pbx1 on DNA containing A in this position and somewhat less with those containing G or T, while Hox B7, Hox B8 and Hox C8 strongly prefer a T (52). This same binding specificity was also recently reported by Chang *et al* (53). We demonstrated that as a monomer, Hox B8 binds most stably to the Pbx1/Hox DNA motif containing a TAAT Hox core, while as a heterodimer with Pbx1 it binds more stably to the DNA motif containing a TTAT-containing probe. Thus, we postulate that heterodimer formation with Pbx1 alters interaction between the N-terminal arm of the Hox HD and nucleotide 7 (Fig. 6) in the Pbx1/Hox motif, shifting the stability from A to T or G, depending on the unique Hox sequences within the N-terminal arm.

The Pbx/Exd family of HD proteins is unique in its content of Gly at position 50. The close proximity of Pbx1 and Hox proteins in the Pbx1–Hox–DNA complex may indicate that Gly50, which lacks a side chain, allows Pbx to exhibit a lower degree of sequence specificity for nucleotides 3′ to its TGAT core, permitting the adjacent Hox protein to dictate most of the sequence specificity associated with the heterodimer element.

The HDs of Pbx1, Pbx2 or Pbx3 select sequences containing TGATG and TGATT as monomer binding motifs (26), indicating that Pbx proteins exhibit some sequence-specific selection for the T at position 6, but little specificity for bases 3′ to this position. Thus, in the Pbx1–Hox–DNA complex both HD proteins are proposed to contact positions 5 and 6, the N-terminal arm of the Hox protein making contacts in the minor groove and helix 3 of Pbx1 binding in the major groove. In this manner, Pbx proteins would differ from yeast a1, which encodes Ile at position 50 and displays DNA-specific contacts three bases 3′ of its core (TGATGT-A/G).

The half-sites we propose for Pbx and Hox proteins, as well as their inseparable nature, is consistent with the sequences of known Pbx/Hox sites in cellular promoters, as well as with sites known to bind Pbx1 and unidentified partners. In addition to the Pbx1/Hox sites found in the Hox B1 and somatostatin promoters, Kagawa *et al.* (54) have described a site in the bovine CYP17 ($P450_{17\alpha}$) gene that binds Pbx1 and a second factor and functions as a cAMP-responsive element (Table 1). This element contains TTGAT at its 5′-half, consistent with the general positioning of Pbx1 on the 5′ half-site and GGACAG on its 3′ half-site. Likewise, we have used antibodies to Pbx1 and nuclear extracts from pre-B cells to select the best heterodimer binding motif for Pbx1 and derived the consensus TTGATTGAC-A/G-G, again finding that the Pbx1 half-site consensus is positioned at the 5′ half-site (Knoepfler, unpublished). In addition, all elements (Table 1 and the aforementioned elements) contain invariant A residues in the fourth and eighth positions, consistent with the proposal that N51 of helix 3 contacts this base and that spacing between the half-sites is not allowed and suggesting that the undefined partners may be HD proteins as well. Positions 10 and 11 of all elements also contain dinucleotide combinations favorable for the binding of Hox proteins. Interestingly, there is no consistent requirement for the T and A residues at positions 6 and 7, which is normally specified by the N-terminal arm of the Hox protein and which is present in almost all naturally occurring monomeric Hox binding sites (11,12,14–16). This may indicate that altered DNA binding by the Hox protein occurs upon heterodimer formation with Pbx1. Confirmation of this general model, as well as elucidation of specific variations in binding induced at the heterodimer interface, must await further experimentation and resolution of the crystal structure of a Pbx–Hox–DNA complex.

ACKNOWLEDGEMENTS

This work was supported by Public Health Services grants 2 PO1 CA50528 and CA56876. MPK is a PEW scholar in the Biomedical sciences. PSK is a fellow of the Lucille P.Markey Charitable Trust.

REFERENCES

1 Krumlauf,R. (1994) *Cell*, **78**, 191–201.
2 Lawrence,P.A. and Morata,G. (1994) *Cell*, **78**, 181–189.
3 McGinnis,W. and Krumlauf,R. (1992) *Cell*, **68**, 283–302.
4 Scott,M.P., Tamkun,J.W. and Hartzel,G.W. (1989) *Biochim. Biophys. Acta*, **989**, 25–48.
5 Gibson,G., Schier,A., LeMotte,P. and Gehring,W.J. (1990) *Cell*, **62**, 1087–1103.
6 Hayashi,S. and Scott,M. (1990) *Cell*, **63**, 883–894.
7 Kuziora,M.A. and McGinnis,W. (1989) *Cell*, **59**, 563–571.
8 Mann,R.S. and Hogness,D.S. (1990) *Cell*, **60**, 597–610.
9 Akam,M. (1989) *Cell*, **57**, 347–349.
10 Desplan,C., Theis,J. and O'Farrell,P.H. (1988) *Cell*, **54**, 1081–1090.
11 Gehring,W.J., Affolter,M. and Bürglin,T. (1994) *Annu. Rev. Biochem.*, **63**, 487–526.
12 Gehring,W.J., Qian,Y.Q., Billeter,M., Furukubo-Tokunaga,K., Schier,A.F., Resendez-Perez,D., Affolter,M., Otting,G. and Wüthrich,K. (1994) *Cell*, **78**, 211–223.
13 Hoey,T. and Levine,M. (1988) *Nature*, **332**, 858–861.
14 Ekker,S., von Kessler,D. and Beachy,P. (1992) *EMBO J.*, **11**, 4059–4072.
15 Ekker,S., Jackson,D., von Kessler,D., Sun,B., Young,K. and Beachy,P. (1994) *EMBO J.*, **13**, 3551–3560.
16 Laughon,A. (1991) *Biochemistry*, **30**, 11357–11367.
17 Qian,Y.Q., Billeter,M., Otting,G., Müller,M., Gehring,W.J. and Wüthrich,K. (1989) *Cell*, **59**, 573–580.
18 Kissinger,C.R.B., Liu,S., Martin-Blabco,E., Kornberg,T.B. and Pabo,C.O. (1990) *Cell*, **63**, 579–590.
19 Wolberger,C., Vershon,A.K., Liu,B.S., Johnson,A.D. and Pabo,C.O. (1991) *Cell*, **67**, 517–528.
20 Ades,S. and Sauer,R. (1994) *Biochemistry*, **33**, 9187–9194.
21 Treisman,J., Gonczy,P., Vashishtha,M., Harris,E. and Desplan,C. (1989) *Cell*, **59**, 553–562.
22 Percival-Smith,A., Muller,M., Affolter,M. and Gehring,W. (1990) *EMBO J.*, **9**, 3967–3974.
23 Ekker,S., Young,K., von Kessler,D. and Beachy,P. (1991) *EMBO J.*, **10**, 1179–1186.
24 Chan,S.-K., Jaffe,L., Capovilla,M., Botas,J. and Mann,R. (1994) *Cell*, **78**, 603–615.
25 Chang,C.P., Shen,W.F., Rozenfeld,S., Lawrence,H.J., Largman,C. and Cleary,M.L. (1995) *Genes Dev.*, **9**, 663–674.
26 Lu,Q., Wright,D.D. and Kamps,M.P. (1994) *Mol. Cell. Biol.*, **6**, 3938–3948.
27 Pöpperl,H., Bienz,M., Studer,M., Chan,S.K., Aparicio,S., Brenner,S., Mann,R.S. and Krumlauf,R. (1995) *Cell*, **81**, 1031–1042.
28 Van Dijk,M.A. and Murre,C. (1994) *Cell*, **78**, 617–624.
29 Kamps,M.P., Murre,C., Sun,X. and Baltimore,D. (1990) *Cell*, **60**, 547–555.
30 Nourse,J., Mellentin,J.D., Galili,N., Wilkinson,J., Stanbridge,E., Smith,S.D. and Cleary,M.L. (1990) *Cell*, **60**, 535–545.
31 Monica,K., Galili,N., Nourse,J., Saltman,D. and Cleary,M.L. (1991) *Mol. Cell. Biol.*, **11**, 6149–6157.
32 Rauskolb,C., Peifer,M. and Wieschaus,E. (1993) *Cell*, **74**, 1101–1112.
33 Bürglin,T.R. and Ruvkun,G. (1992) *Nature Genet.*, **1**, 319–320.
34 Dedera,D.A., Waller,E.K., Lebrun,D.P., Sen-Majumdar,A., Stevens,M.E., Barsh,G.S. and Cleary,M.L. (1993) *Cell*, **74**, 833–843.
35 Kamps,M.P. and Baltimore,D. (1993) *Mol. Cell. Biol.*, **13**, 351–357.
36 Kamps,M.P. and Wright,D. (1994) *Oncogene*, **9**, 3159–3166.
37 Van Dijk,M., Voorhoeve,P. and Murre,C. (1993) *Proc. Natl. Acad. Sci. USA*, **90**, 6061–6065.
38 Lu,Q. and Kamps,M.P. (1996) *Mol. Cell. Biol.*, **16**, 1632–1640.
39 Johnson,F.B., Parker,E. and Krasnow,M.A. (1995) *Proc. Natl. Acad. Sci. USA*, **92**, 739–743.
40 Knoepfler,P. and Kamps,M.P. (1995) *Mol. Cell. Biol.*, **15**, 5811–5819.
41 Leonard,J., Peers,B. Johnson,T. Ferreri,K., Lee,S. and Montminy,M. (1993) *Mol. Endocrinol.*, **7**, 1275–1283.
42 Peers,B., Sharma,S., Johnson,T., Kamps,M. and Montminy,M. (1995) *Mol. Cell. Biol.*, **15**, 7091–7097.
43 Lu,Q., Knoepfler,P.S., Scheele,J., Wright,D.D. and Kamps,M.P. (1995) *Mol. Cell. Biol.*, **15**, 3786–3795.
44 Popperl,H. and Featherstone,M. (1992) *EMBO J.*, **11**, 3673–3680.
45 Odenwald,W., Garbern,J., Arnheiter,H., Tournier-Lasserve,E. and Lazzarini,R. (1989) *Genes Dev.*, **3**, 158–172.
46 Goutte,C. and Johnson.A. (1994) *EMBO J.*, **13**, 1434–1442.
47 Li,T., Stark,M.R., Johnson,A.D. and Wolberger,C. (1995) *Science*, **270**, 262–269.
48 Lebrun,D. and Cleary,M. (1994) *Oncogene*, **9**, 1641–1647.
49 Wilson,D., Sheng,G.J., Lecuit,T., Dostatni,N. and Desplan,C. (1993) *Genes Dev.*, **7**, 2120–2134.
50 Catron,K.M., Iler,N. and Abate,C. (1993) *Mol. Cell. Biol.*, **13**, 2354–2365.
51 Pellerin,I., Schnabel,C., Catron,K.M. and Abate,C. (1994) *Mol. Cell. Biol.*, **14**, 4532–4545.
52 Lu,Q. and Kamps,M.P. (1996) *Oncogene*, submitted for publication.
53 Chang,C.P., Brochierri,L., Shen,W.F., Largman,C. and Cleary,M. (1996) *Mol. Cell. Biol.*, **16**, 1734–1745.
54 Kagawa,N., Ogo,A., Takahashi,Y., Iwamatsu,A. and Waterman,M.R. (1994) *J. Biol. Chem.*, **269**, 18716–18719.