# Conformational Changes in the PBX Homeodomain and C-Terminal Extension upon Binding DNA and HOX-Derived YPWM Peptides[†,‡]

Tara Sprules,[§] Nancy Green,[||] Mark Featherstone,[§,||] and Kalle Gehring*[,§,⊥]

*Department of Biochemistry, McGill Cancer Centre and Montreal Joint Centre for Structural Biology, McIntyre Medical Sciences Building, McGill University, 3655 Drummond, Montreal, PQ, Canada, H3G 1Y6*

*Received January 19, 2000; Revised Manuscript Received May 2, 2000*

ABSTRACT: PBX is a member of the three amino acid loop extension (TALE) class of homeodomains. PBX binds DNA cooperatively with HOX homeodomain proteins that contain a conserved YPWM motif. The amino acids immediately C-terminal to the PBX homeodomain increase the affinity of the homeodomain for its DNA site and HOX proteins. We have determined the structure of the free PBX homeodomain using NMR spectroscopy. Both the PBX homeodomain and the extended PBX homeodomain make identical contacts with a 5′-TGAT-3′ DNA site and a YPWM peptide. A fourth α-helix, which forms upon binding to DNA, stabilizes the extended PBX structure. Variations in DNA sequence selectivity of heterodimeric PBX-HOX complexes depend on the HOX partner; however, a comparison of five different HOX-derived YPWM peptides showed that each bound to PBX in the same way, differing only in the strength of the association.

*In vitro* HOX proteins recognize DNA binding sites containing a common 5′-TAAT-3′ core, although their biological activities are very different (*1*). HOX-PBX heterodimers bind cooperatively to the consensus DNA site 5′-ATGATTNAT-3′ (*2*, *3*). Both the HOX homeodomain and a highly conserved YPWMR/K pentapeptide motif (YPWM motif) N-terminal to the homeodomain are required for cooperative interactions with PBX (*4*−*8*). HOX-derived peptides containing the YPWM motif and its flanking amino acids increase the affinity of PBX for DNA (*9*−*11*). These peptides also prevent the formation of HOX-PBX heterodimers. The identity of the variable nucleotide in the DNA binding site depends on residues within the HOX homeodomain (*2*, *12*). The sequence of the pentapeptide motif and linker region can also influence the choice of binding site (*2*).

PBX and its homologues, *Drosophilia* extradenticle (Exd) and *Caenorhabditis elegans ceh*-20, the PBC family of proteins, belong to the three amino acid loop extension (TALE)[1] superclass of homeodomains (*13*). The extra amino acids of PBC homeodomains are essential for cooperative interactions with HOX proteins (*10*, *14*). The 15 amino acids immediately C-terminal to the homeodomain are highly conserved among PBC proteins (*15*−*18*). Full-length monomeric PBX is unable to interact with DNA, and its isolated homeodomain binds weakly. Maximal monomer and cooperative binding of the PBX homeodomain to DNA are seen in the presence of the conserved C-terminal amino acids (C-terminal extension) (*5*, *10*, *19*).

The crystal structures of a PBX-HOXB1-DNA complex and an extradenticle-ultrabithorax (Exd-Ubx) DNA complex have recently been solved (*20*, *21*). The PBX-HOXB1 complex includes the conserved C-terminal extension of PBX and shows that it forms a fourth α-helix which packs against the homeodomain. This potentially stabilizes its structure and accounts for its increased affinity for DNA. The Exd-Ubx structure contains only the homeodomain of Exd. The N-terminal YPWM motif, linker regions and homeodomains of HOXB1 and Ubx are present. The two homeodomains bind to DNA in a head-to-tail orientation. The only interprotein interaction is the insertion of the YPWM motif of the HOX partner into a hydrophobic binding pocket in PBX formed from the extended loop between helices one and two and the C-terminus of helix three. In the PBX-HOXB1 crystal structure the turn between helices three and four of PBX also contributes to the binding pocket.

Recently Jabet and co-workers (*22*) reported the results of their NMR studies showing that the fourth α-helix forms only when the extended PBX homeodomain binds to DNA. They compared the secondary structures of the homeodomain and the extended homeodomain in the absence of DNA and found that the C-terminal extension destabilizes the PBX homeodomain in solution.

We report here results from our NMR studies of PBX. We have determined the solution structure of the truncated PBX homeodomain. This represents the first structure of a TALE class homeodomain in the absence of DNA and cooperative binding partners. A comparison of the structures

[†] This work was supported by Medical Research Council of Canada operating grants to M.F. and K.G.

[‡] The atomic coordinates for the truncated PBX homeodomain have been deposited in the Protein Data Bank (PDB ID code 1DU6). Chemical shift assignments have been deposited with the BioMagRes Bank (BRMB ID code 4572).

* To whom correspondence should be addressed. Phone: (514) 496-2558. Fax: (514) 496-5143. E-mail: kalle@bri.nrc.ca.

[§] Department of Biochemistry.

[||] McGill Cancer Centre.

[⊥] Montreal Joint Centre for Structural Biology.

[1] Abbreviations: TALE, three amino acid loop extension; NMR, nuclear magnetic resonance; NOE, nuclear Overhauser effect; HCM, HOX cooperativity motif.

10.1021/bi0001067 CCC: $19.00 
Published on Web 07/21/2000

of the PBX homeodomain and the PBX homeodomain plus the C-terminal extension was made to examine more completely the role of the C-terminal extension in stimulating binding to DNA and HOX proteins. Both proteins were studied in the free state, bound to a DNA site, and bound to DNA in the presence of a HOXA1-derived YPWM peptide. The interaction between five HOX-derived YPWM peptides and the extended PBX homeodomain was observed by NMR spectroscopy to investigate the influence of the pentapeptide motif on heterodimer formation and DNA sequence specificity.

## MATERIALS AND METHODS

*Sample Preparation.* His-tagged PBX fusion proteins were expressed in *E. coli* BL21(DE3)pLysS cells (*19*). Cultures were grown in LB media for unlabeled samples or M9 media supplemented with $^{15}N$ $NH_4Cl$ and $U^{13}C_6$ D-glucose as the sole nitrogen and carbon sources for labeled samples. Cells were grown at 37 °C to an $A_{600}$ of 0.4 OD then induced with 1 mM isopropyl-1-thio-$\beta$-D-galactopyranoside and grown for a further 3−4 h. For the truncated homeodomain, PBX 9−59, the resulting cell pellet was resuspended in binding buffer, lyzed by sonication, and purified in one step by affinity chromatography with $Ni^{2+}$-charged chelating sepharose. The protein was dialyzed extensively against $ddH_2O$, the pH was adjusted to 4.7, and the sample was concentrated with Ultrafree-4 centrifugal concentrators (Millipore, Bedford, MA). NMR samples were 1−4 mM protein in 90% $H_2O$/ 10% $^2H_2O$ or 100% $^2H_2O$ supplemented with 1 mM ETDA to prevent aggregation of the polyhistidine tag and 1 mM DTT to avoid dimerization through the lone cysteine residue. Approximately 40 mg of protein were obtained from 1 L of cell culture. A more lengthy purification protocol was required for the remaining constructs, in which the polyhistidine tag was enzymatically removed. Cells were pelleted by centrifugation, resuspended in binding buffer containing 4 M urea and lyzed. PBX was purified by affinity chromatography on a $Ni^{2+}$-charged chelating sepharose column and refolded on the column by stepwise decreases in urea concentration in the wash buffer. The histidine tag was removed by treatment with Factor Xa (1 unit/mg PBX) (Haematologic Technologies Inc., Essex Jct., VT) for 3 h at room temperature and the Factor Xa was removed with *p*-aminobenzamidine agarose resin (Sigma, St. Louis, MO). The protein was further purified by HPLC on a Mono S 5/5 column (Amersham Pharmacia Biotech, Piscataway, NJ) followed by a HiLoad 16/60 Superdex 75 prep grade column (Amersham Pharmacia Biotech). The samples were finally dialyzed against 20 mM sodium phosphate, pH 7.0, and concentrated to 1−2 mM. Approximately 5 mg of pure protein were obtained from 1 L of cell culture.

The 14-mer DNA oligonucleotides, d(GCGCATGATTGCCC) and d(GGGCAATCATGCGC) were purchased from AlphaDNA (Montreal, Canada) and purified on a Mono Q 5/5 column (Amersham Pharmacia Biotech). Each strand was dialyzed extensively into water and desalted on a NAP-25 column (Amersham Pharmacia Biotech). Equimolar amounts of each strand were combined, annealed and concentrated.

HOX-derived peptides HOXA1, AQTFDWMKVKRNPPKTGKV; HOXD4, AVVYPWMKKVHVNSVNPNY; HOXA1t, AQTFDWMKVKRN; HOXA1tDP, AQTFPWMKVKRN; and HOXA1tDA, AQTFAWMKVKRN; were purchased (Peptide Chemistry, BRI−NRC, Montreal, Canada) and purified by reverse phase chromatography on a C-18 column.

*NMR Spectroscopy.* NMR experiments were recorded on Bruker DRX 500 or Varian Unity 750 spectrometers equipped with pulsed field gradient probes at 30 °C. Proton chemical shifts were referenced to an internal sample of DSS at 0 ppm. $^{15}N$ and $^{13}C$ chemical shifts were referenced to the proton spectrum using the ratio of gyromagnetic moments ($\gamma^N/\gamma^H$= 0.101 329 05, $\gamma^C/\gamma^H$= 0.251 449 52). Data were processed using the Gifa program, Version 4.2 (*23*). Resonance assignments and coupling constants were obtained based on standard $^{15}N$-edited HMQC-TOCSY and HMQC-NOESY, HNHA and HNHB experiments on $^{15}N$ labeled samples and CBCACONH, HNCACB, TOCSY-CONH, and $^{13}C$ HSQC experiments on doubly labeled $^{15}N$, $^{13}C$ samples. Assignments for aromatic protons were determined from 2D HBCBCGCDHD and HBCBCGCDCEHE experiments. Distance restraints for the PBX 9−59 structure calculation were derived from a 200 ms mixing time $^{15}N$-HMQC-NOESY and a series of homonuclear NOESY experiments with mixing times of 50, 100, and 200 ms in $D_2O$ and a 200 ms NOESY in $H_2O$. $^{15}N$ $T_1$ and $T_2$ values were measured for a 2 mM sample of $^{15}N$-labeled PBX 9−59 and a 1.4 mM sample of $^{15}N$ labeled PBX 9−84 (*24*). Transferred NOEs for the HOXA1t peptide were recorded in a 200 ms mixing time NOESY of 5.3 mM HOXA1t with 5% (mol/mol) $^{15}N$ PBX 1−59-DNA. The program XEASY (*25*) was used for analysis of NOE data.

*Structure Calculation.* An ensemble of structures for the 75 residue His-tagged truncated PBX homeodomain was calculated based on the constraints listed in Table 1. No NOEs were observed for the first 11 residues ($GH_{10}$) of the polypeptide; these amino acids were omitted from the structure determination. The final structures include the 10 C-terminal residues of the polyhistidine tag followed by residues 9−59 of the PBX homeodomain (residues 241−294 of the full-length protein). The starting conformation for dynamical annealing structure calculations was a homology model based on the des (1−6) Antennapedia structure (*26*). The last 10 amino acids of the histidine tag were added in an extended conformation, and the extended loop between helices one and two was modeled using the generate loops feature in InsightII (Molecular Simulations, Inc.). Initially 75% of the NOEs were assigned based on chemical shifts, independently of the model structure. The remaining ambiguous NOEs were later unambiguously assigned based on preliminary structure calculations. The statistics for the 30 lowest energy structures of 100 are presented in Table 1. An ensemble of structures for the HOXA1t peptide was calculated based on 47 sequential and 31 long-range NOEs. The program CNS, versions 0.3 and 0.5 (Yale University), was used for all structure calculations.

## RESULTS

A series of recombinant polypeptides were constructed in order to compare the structure and DNA binding properties of the PBX homeodomain with and without the conserved C-terminal extension (Figure 1). The lone cysteine at position 38 of the homeodomain was found to dimerize readily in

Table 1: Structure Statistics of the Truncated PBX Homeodomain

| restraints for structure calculations | |
|---|---|
| total restraints used | 1338 |
| total NOE restraints | 1268 |
| intraresidue | 517 |
| sequential ($\|i-j\|=1$) | 309 |
| medium range ($1<\|i-j\|\le 4$) | 277 |
| long range ($\|i-j\|>4$) | 165 |
| hydrogen bond restraints | 22 |
| dihedral angle restraints | 37 |
| *J*-coupling restraints | 11 |

| statistics for structure calculations | ⟨SA⟩[a] |
|---|---|
| rmsd from idealized covalent geometry | |
| bonds (Å) | 0.0041 ± 0.0002 |
| bond angles (deg) | 0.56 ± 0.02 |
| improper torsions (deg) | 0.43 ± 0.03 |
| rmsd from experimental restraints[b] | |
| distances (Å) | 0.027 ± 0.002 |
| final energies (kcal mol$^{-1}$) | |
| $E_{total}$ | 284 ± 19 |
| $E_{bonds}$ | 17.8 ± 1.7 |
| $E_{impropers}$ | 16.3 ± 2.5 |
| $E_{vdW}$[c] | 77.4 ± 8.5 |
| $E_{NOE}$ | 75.1 ± 9.3 |

| coordinate precision[d] (Å) | ⟨SA⟩ versus $\overline{\langle SA\rangle}$ |
|---|---|
| rmsd of backbone atoms (N, Cα, C′) for residues 11−52 | 0.50 ± 0.13 |
| rmsd of all heavy atoms for residues 11−52 | 1.08 ± 0.12 |

[a] ⟨SA⟩ refers to the ensemble of the 30 structures with lowest energy from 100 calculated structures. [b] No distance restraint in any of the structures included in the ensemble was violated by more than 0.43 Å. [c] Repel = 4.0 for the final step of calculations. [d] Rmsd between the ensemble of structures ⟨SA⟩ and the average structure of the ensemble $\overline{\langle SA\rangle}$.

the absence of reducing agents and was mutated to a serine in all constructs except for PBX 9−59. This mutation did not significantly affect the structure or DNA-binding affinity of the PBX homeodomain (data not shown). The complete PBX homeodomain consists of residues 233−294 of the full length protein, which will be referred to as residues 1−59 in the following discussion. The truncated PBX homeodomain, PBX 9−59, lacks the arginine-rich N-terminal arm of the full homeodomain and is incapable of DNA binding. This construct was used to determine the solution structure of the PBX homeodomain. The spectra of the first homeodomain studied, PBX 1−84, corresponding to the polypeptide used in the published crystal structure and NMR studies (*20*, *22*), had broad line widths in comparison to PBX 1−59 and PBX 9−59. Measurement of amide $^{15}$N $T_1$ and $T_2$ values and comparison of correlation times (*27*, *28*) for PBX 9−59 and PBX 9−84 suggested that the extended homeodomain aggregated. Higher $^{15}$N relaxation times indicated that the C-terminus was unstructured. The spectra of a shorter construct, PBX 1−78, were much better than those of the longer polypeptide.

*Solution Structure of the Truncated PBX Homeodomain.* An ensemble of 100 structures of PBX 9−59 was calculated from an average of 33 constraints/residue. The 30 lowest energy structures were chosen to represent the conformation of the truncated PBX homeodomain (Figure 2). The rmsd for the backbone atoms of the well-defined portion of the polypeptide, from Q11 to K52, is 0.5 Å. Statistical analysis

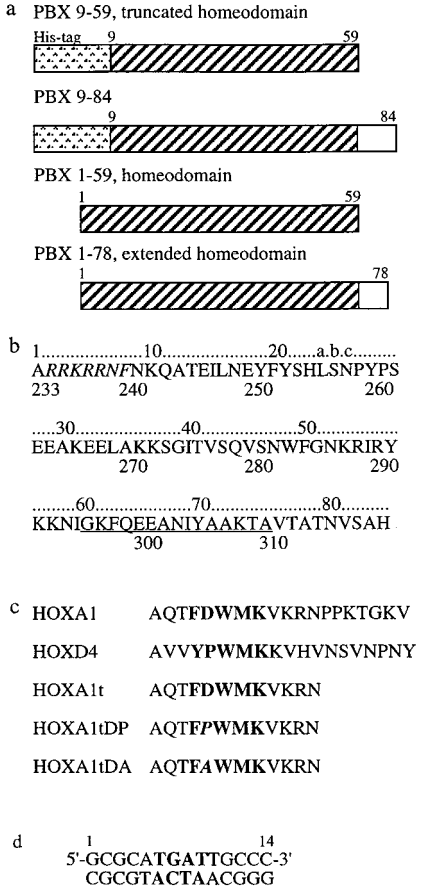


FIGURE 1: (a) Schematic representation of the PBX homeodomain constructs employed in this study. (b) Amino acid sequence of PBX. The numeric position in the full-length PBX protein is shown below the amino acid sequence, above is the standard homeodomain numbering. The extra 3 amino acids are numbered 23a−c to preserve consensus. The N-terminal arm is italicized and the highly conserved C-terminal amino acids are underlined. The polyhistidine tag, $GH_{10}$SSGHIEGRHM, was enzymatically cleaved from the PBX 1−59 and PBX 1−78 constructs. (c) HOX-derived YPWM peptides. The YPWM motif is highlighted in bold type. (d) Nucleic acid sequence of the 14-mer DNA duplex. The PBX recognition site is highlighted in bold type.

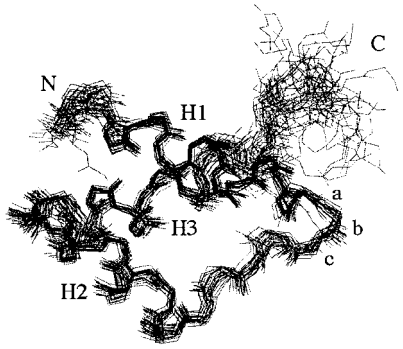


FIGURE 2: Structure of the truncated PBX homeodomain, residues PBX 9−59. The superposition of the polypeptide backbone of the 30 lowest energy conformations is shown. The 21 amino acids of the His-tag were unstructured and have been omitted for clarity. N and C represent the positions of N9 (N241) and N59 (N294). The three α-helices are labeled H1, H2, and H3 and the three amino acid loop insertion is labeled a−c.

of the representative structures is presented in Table 1.

The truncated PBX homeodomain consists of three α-helices, which fold around a hydrophobic core. The first α-helix extends from residues Q11 to Y21, the second from

E28 to C38, and the third from V42 to K52. Residues R53 to Y56 have $^3J_{HNH\alpha}$ values below 5 Hz, but exhibit few NOEs characteristic of α-helical conformation. Amide $^{15}N$ $T_1$ and $T_2$ values increase from Y56 onward. This indicates that residues R53−Y56 retain some helical character, while the remainder of the polypeptide is unstructured. The 21 amino acid N-terminal polyhistidine tag is unstructured and does not interact with the homeodomain, with the exception of the final methionine, which contacts the hydrophobic core. The three amino acid insertion is located in the loop between helices one and two. Two different orientations that satisfy the experimental constraints are observed for the side chain of H23 and the P24 carbonyl, which flank this insertion. Despite these differences in orientation the extended loop between helices one and two has a well-defined conformation in the free PBX homeodomain.

The solution structure of PBX 9−59 was compared with the structure of PBX 1−84 in the PBX-HOXB1 heterodimer (*20*). The rmsd between the ensemble of NMR structures and the crystal structure is 0.98 Å for the backbone atoms of residues Q11−K52. The two structures differ in the length of the third helix, which ends at K52 in PBX 9−59, but extends to K58 in the crystal structure. The beginning of the first helix and the loop between helices one and two are in slightly different positions. These differences reflect the conformational changes that occur when the homeodomain binds to DNA. Neither conformation of the imidazole ring of H23 seen in the NMR structures matches the conformation observed in the crystal structure, whereas one orientation of the P24 carbonyl is the same.

*Free PBX 1−59 and PBX 1−78 Differ in the Length of Helix 3.* Complete backbone amide and side chain assignments have been made for the PBX homeodomain, PBX 1−59. The chemical shifts for residues common to PBX 9−59 and PBX 1−59 are the same, indicating that the tertiary structure of the two polypeptides is the same. In a fashion analogous to the Antennapedia homeodomain (*26*), the N-terminal arm is flexible in the absence of DNA and does not interact with the globular homeodomain core. Determination of the structure of the extended PBX homeodomain in the free state was unfeasible due to poor quality of the NMR spectra. Although broad line widths caused by aggregation of PBX 1−84 were reduced in PBX 1−78, the amide protons of residues N7, N9, K10, E14, S43, K52, and Y56-G61 were not observed at pH 7.0. As well, the cross-peaks for the amide protons in the N-terminal arm were weak in comparison to PBX 1−59. The secondary structure of PBX 1−78 was obtained by analysis of chemical shift index (*29*), $^3J_{HNH\alpha}$ coupling constants, and NOE patterns and is the same as that of PBX 9−59. The N-terminal arm and C-terminal extension are unstructured. Chemical shifts for PBX 1−59 and PBX 1−78 are very similar, with the exception of the end of the third helix and the C-terminal half of the first helix. As was previously reported (*22*), addition of the C-terminal extension to the PBX homeodomain results in some differences in the third helix of the homeodomain, and generally destabilizes its structure. We found that the third helix of PBX 1−59 is well-defined to K52. Progressively sharper line widths and fewer NOEs are observed from R53 until the end of the polypeptide at N59. The third helix of PBX 1−78 ends at N51. Amide cross-peaks are very weak or absent for residues K52 to G61. Some

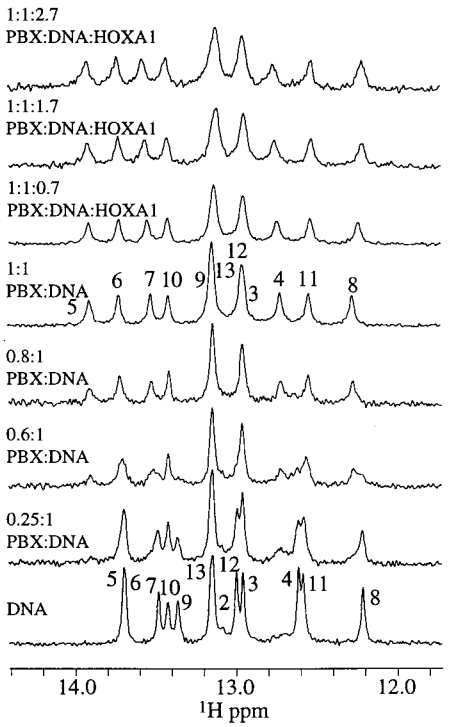


FIGURE 3: Changes in DNA imino resonances upon complex formation. 1d proton spectra were recorded at 500 MHz and 30 °C. Imino peaks are numbered according to the position of the corresponding base pair.

sequential NOEs can be seen for K62 to A71, but $^3J_{HNH\alpha}$ values are above 5 Hz, and NOEs characteristic of α-helical form are not observed. Residues A72 to A78 of the C-terminal extension are unstructured and highly mobile in the free polypeptide. The differences at the end of the third helix, which crosses the first helix, account for differences between the spectra of the homeodomain and the extended homeodomain.

*Both PBX 1−59 and PBX 1−78 Form Stable Complexes with a TGAT DNA Site.* DNA-bound PBX 1−59 and PBX 1−78 were compared to determine the role of the C-terminal extension in increasing DNA-binding affinity. The DNA duplex used in these studies was designed to contain a single PBX 5′-TGAT-3′ recognition site to avoid the formation of homodimers (Figure 1). The titration of the DNA duplex with PBX was followed in 1D proton spectra (Figure 3). The formation of the DNA−PBX complex is in slow exchange on the NMR time scale. The imino resonances of the free DNA disappear and new peaks appear as PBX is titrated into the DNA. Upon binding to the TGAT DNA site there was an increase in spectral dispersion in the HSQC spectra. Both the PBX homeodomain and the extended PBX homeodomain formed very stable complexes with the DNA duplex. In contrast to published results (*22*), cross-peaks for all of the backbone amides in the extended homeodomain could be observed in our binary complex. Complete backbone amide assignments were made for both the PBX 1−59 and PBX 1−78 binary complexes. In addition, nearly all the protein side chain and DNA aromatic, imino, H1′, H2′, H2″, and H3′ assignments have been made for the PBX 1−78-DNA complex. As was observed for the free proteins, the chemical shifts are very similar for PBX 1−59 and PBX 1−78 except at the ends of the first and third α-helices. The largest changes in amide chemical shift between the free and

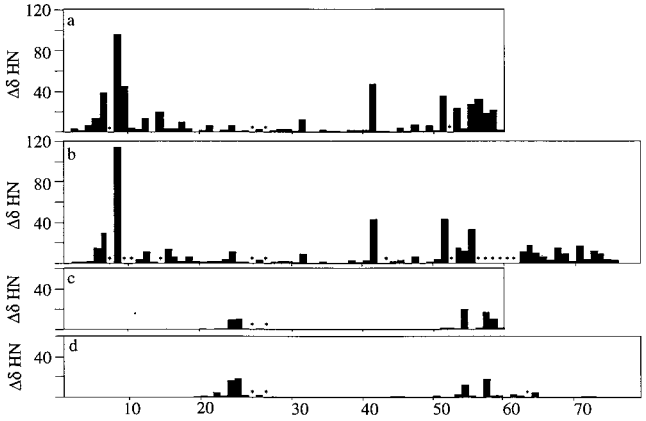


FIGURE 4: Graphical representation of chemical shift changes in PBX $^{15}$N HSQC spectra upon binding to DNA and HOXA1. ΔδHN is the difference in the position of the amide cross-peak for a given residue in two different forms. $\Delta\delta HN = ((\delta Hb - \delta Ha) \times 10)^2 + ((\delta Nb - \delta Na))^2$. (a) ΔδHN between free PBX 1−59 and DNA bound PBX 1−59. (b) ΔδHN between free PBX 1−78 and DNA bound PBX 1−78. (c) ΔδHN on addition of HOXA1 to the PBX 1−59 binary complex. d) ΔδHN on addition of HOXA1 to the PBX 1−78 binary complex.

DNA-bound forms of the proteins are seen for residues that interact with the DNA site (Figure 4). The N-terminal arm contacts the minor groove of the DNA site, and additional contacts to the hydrophobic core are formed by the aromatic ring of F8. Secondary structure analysis shows that contact with the DNA site results in the lengthening of the third, or recognition helix, to K58 upon formation of the binary complex for both PBX 1−59 and PBX 1−78. The most striking observation upon binding of PBX 1−78 to DNA is the formation of the fourth α-helix from K62 to A72 in the C-terminal extension. This helix does not contact the DNA site. NOEs between the aromatic ring Y70 and the amide protons of N17 and E18 in helix one fold the fourth helix toward the rest of the homeodomain. As Jabet et al. postulated (*22*), these additional contacts could stabilize the homeodomain structure, resulting in reduced mobility in the third helix and higher DNA-binding affinity. In the conditions of our NMR study, however, both the PBX homeodomain and the extended homeodomain form equally stable binary complexes with the DNA duplex. Steady-state binding experiments also showed that the two constructs bind DNA with similar efficiency, while in competition experiments the extended homeodomain has a 5-fold higher affinity for the DNA probe (*19*).

*Binding Site for a HOXA1-Derived YPWM Peptide Can be Identified Based on Chemical Shift Changes.* Upon addition of the HOXA1 peptide (Figure 1) to the binary complexes, chemical shift changes are seen in two localized regions of the protein. Intermediate exchange kinetics are observed in $^{15}$N HSQC spectra as the backbone amide cross-peaks of residues S22−N23c (loop 1) and I54−N59 (helix 3) first broaden and then sharpen during titration with increasing amounts of peptide. The side chain amides of Q64 also shift substantially. Some small shifts are seen in the C-terminal extension (Figure 4). These shifts identify the three amino acid loop extension and the end of helix 3 as the YPWM binding site, as is seen in the crystal structures of PBC−HOX heterodimers (*20*, *21*). The imino protons of the DNA also broaden and shift as the peptide is added (Figure 3), indicating that the interface between the DNA and PBX changes. $^{15}$N HMQC-NOESY spectra show changes in side chain chemical shifts and the number of NOEs observed for residues in the binding site (Figure 5). The quality of the NMR spectra, especially for PBX 1−78, was improved upon binding to DNA. However, increased line widths were observed in the spectra of the ternary complexes. This may result from aggregation or conformational exchange.

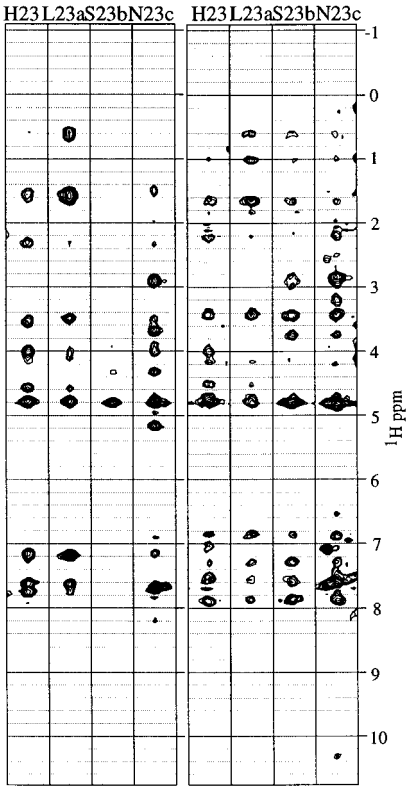


FIGURE 5: Comparison of strips from $^{15}$N-edited HMQC-NOESY spectra of the binary PBX-1−78-DNA complex (left) and the ternary PBX 1−78-DNA-HOXA1 complex (right). The selected strips are from the loop between helices one and two. Experiments were recorded at 500 MHz and 30 °C. Sample concentration was approximately 1 mM protein in 20 mM sodium phosphate, pH 7.0.

*Protein−Peptide Interactions Are Observed Using Transferred NOEs.* The low affinity of the HOXA1 peptide for the PBX 1−59-DNA complex allows the use of transferred NOEs to study the bound conformation of the peptide. When there is a large excess of peptide in comparison to protein, each peptide molecule spends a short time bound to the complex. Information is transferred from the bound state to the free state of the peptide, where it is observed. Intermolecular NOEs were identified between the phenylalanine and tryptophan rings of HOXA1t and L23a of PBX in the NOESY spectrum of 20:1 peptide:binary complex (Figure 6). A structure for the HOXA1t peptide was calculated (data not shown). The conformation of the backbone from residues T3 to M7 is similar to the corresponding region of HOXB1 in the PBX-HOXB1 crystal structure (*20*), although not as tightly folded.

*Binding Modes for Different HOX-Derived YPWM Peptides Are the Same.* To examine the role of the HOX YPWM motif in determining DNA binding site preferences a series of HOX-derived peptides were constructed (Figure 1) and titrated with the PBX 1−78-DNA complex. The binding of HOXD4 and mutant HOXA1 peptides to the PBX−DNA

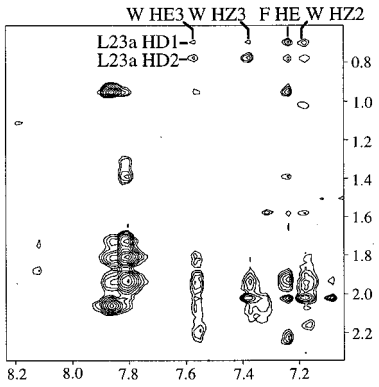


FIGURE 6: Detail from a 200 ms $^1$H-$^1$H NOESY of 5.3 mM HOXA1t with 5% (mol/mol) PBX 1−59, 20 mM sodium phosphate, pH 6.4. Intermolecular PBX-HOXA1t NOEs are labeled.

complex also affected only residues S22−N23c and I54−N59. No differences in NOE patterns or side chain chemical shifts were seen in a comparison of $^{15}$N HMQC-NOESY spectra for the HOXA1 and HOXD4 ternary complexes. The only variation among the peptides was in their affinity for the PBX−DNA complex. HOXA1 had the highest affinity for the PBX 1−78-DNA complex, with a $K_d$ of about 10 $\mu$M. HOXD4 bound to PBX 1−78-DNA with a $K_d$ of 30 $\mu$M. The three shorter peptides had $K_d$s of around 100 $\mu$M for the binary complex. HOXA1 bound to the PBX 1−59-DNA complex with a $K_d$ of 100 $\mu$M, a 10-fold reduced affinity compared to the extended homeodomain. This agrees well the results of dissociation rate experiments with the HOXA1 homeodomain, in which the half-life of the PBX 1−84-HOXA1-DNA complex was six times longer than that of the PBX 1−59 ternary complex (*19*). HOX-derived peptides were also titrated with PBX in the absence of DNA. Addition of one equivalent of HOXD4 to a sample of $^{15}$N-labeled PBX 9−84 did not result in any change in the HSQC spectrum; neither did addition of three equivalents of HOXA1t to PBX 9−59.

## DISCUSSION

We have reported the first structure for a free TALE class homeodomain. The solution structure of the truncated PBX homeodomain is very similar to other homeodomain structures. Three α-helices, separated by two loops, fold around a hydrophobic core. The amino acids that make up the hydrophobic core are highly conserved across all classes of homeodomains and determine the overall fold of the protein (*30*). The three extra amino acids typical of TALE class homeodomains extend the loop between the first and second α-helices, forming a bulge that allows the rest of the loop to retain a conformation like that of non-TALE homeodomains. The rmsd for the backbone atoms of the three α-helices of *Antennapedia* (*26*) and PBX is 1.0 Å. The extended loop and globular core of free PBX 9−59 are quite similar to the corresponding regions in the crystal structure of the PBX-HOXB1 heterodimer (*20*).

Although the overall fold of all homeodomains is the same, there is some variation in the length of the third (recognition) helix in the absence of DNA. It has been shown that the identity of the amino acids at positions 52 and 56 can determine the length of this helix and influence protein stability in the absence of DNA (*31*, *32*). The length of the third helix does not, however, influence DNA-binding affinity. There are two α-helices in the C-terminus of Antennapedia homeodomain: one from residues 42 to 52 and a less defined helix from residues 53 to 59 (*33*). Residues 52 and 56 of Antp are arginine and tryptophan. Tryptophan 56 of Antp interacts with its hydrophobic core, potentially stabilizing the fourth helix. Helix three ends at residue 52 in the solution structures of the vnd/NK-2 and fushi tarazu (ftz) homeodomains, which have threonine and serine at position 56, respectively (*31*, *34*). A double H52R, T56W mutation in the vnd/NK-2 homeodomain extends this helix to residue 60 (*32*). The presence of lysine and tyrosine at positions 52 and 56 of the PBX homeodomain could account for its increased thermal stability and the extension of helical character past residue 52. Antp and ftz homeodomains denature at 48 °C and 27 °C (*31*). The truncated PBX homeodomain is stable in solution to at least 50 °C, and its third helix runs from residues 42 to 52, with helical character continuing to residue 56. The structure of PBX 9−59 was determined at 30 °C. In comparison, the highest temperature used for structural studies of homeodomains displaying an elongated recognition helix was 25 °C (*26*, *32*, *35*−*37*). Secondary structural analysis of the PBX homeodomain (1−61) at 25 °C by Jabet et al. (*22*) located the end of helix three at Y56.

The C-terminal extension of PBX increases the homeodomain's affinity for DNA (*5*, *10*, *19*). The crystal structure of a PBX-HOXB1 heterodimer showed that a fourth α-helix, which packs against the rest of the homeodomain, is found in the C-terminal extension of PBX (*20*). This helix was proposed to stabilize the structure of the PBX homeodomain and hold the third helix in an optimal position for binding to DNA. Subsequent NMR studies (*22*) demonstrated that this helix forms only upon binding of the extended homeodomain to DNA. In comparison to the PBX homeodomain, the third helix of the extended PBX homeodomain is less structured in the free state. Upon binding to the TGAT DNA-site this helix lengthens to K58 in both PBX 1−59 and PBX 1−78. Our results, like those of Jabet et al. (*22*), show that small differences in the orientation of the helices relative to one another, and the extension of the third α-helix, induce the formation of the fourth helix in the DNA-bound extended homeodomain. The packing of this helix against the homeodomain results in the higher affinity for DNA of PBX 1−78 compared to PBX 1−59, due to decreased mobility in the recognition helix.

Removal of the C-terminal extension weakens, but does not abolish, both monomeric and cooperative binding of the PBX homeodomain to DNA (*19*). Addition of a HOXA1-derived peptide to DNA-bound PBX 1−59 revealed that loop one and helix three are sufficient to form the hydrophobic pocket where the YPWM motif binds, as is seen in the Ubx-Exd crystal structure (*21*). HOX-derived peptides only interact with the DNA-bound form of PBX. The conformation of the loop between helices one and two, which contains the three amino acid insertion necessary for cooperative interactions with HOX (*14*), is very similar in the free and DNA-bound forms of the homeodomain. The end of helix three is somewhat disordered in the free protein. The extension of the third helix from K52 to K58 must be necessary for recognition of the binding pocket by the HOX YPWM motif. Residues Y25, R53, I54, R55, and K57 are

involved in interactions with both the YPWM motif and DNA. The increased affinity of PBX for DNA observed in the presence of HOX-derived YPWM peptides (*9*−*11*) may result from a slight repositioning of these residues that improves contacts with the DNA. The affinity of the HOXA1 peptide for PBX 1−78 is approximately 10 times higher than for PBX 1−59. The only C-terminal residue involved in interprotein contacts identified in the PBX-HOXB1 crystal structure is I60. The side chain amide protons of Q64, which are not observed in the crystal structure due to poor electron density, are the only resonances in the C-terminal extension that exhibit a significant change in chemical shift when the HOXA1 peptide binds; they likely form a hydrogen bond with the peptide. These additional contacts could contribute to the increased affinity of HOXA1 for PBX 1−78. Decreased mobility of the third helix of the extended homeodomain could also improve interaction with the HOX YPWM motif.

The HOX half-site in DNA sequences recognized by PBX-HOX heterodimers contains a variable position: 5′-ATGATTNAT-3′. The two homeodomains contact opposite faces of the DNA site at this point in the sequence, the HOX N-terminal arm in the minor groove, and the third helix of PBX in the major groove. The identities of the nucleotides 3′ to the core homeodomain binding site are influenced by the amino acid at position 50 of the homeodomain (*38*−*40*). In PBX this is a glycine, therefore the preferred base at the variable nucleotide is determined by the HOX N-terminal arm. The presence of a glycine, instead of the more common glutamine, also explains the relatively low DNA-binding affinity of PBX compared to other homeodomains. Construction of chimeric proteins has demonstrated that for some HOX homeodomains the sequence of the linker region and YPWM motif can also affect the binding site selectivity (*2*). However, spectra of PBX 1−78-DNA bound to either the HOXD4 or HOXA1 peptide were identical. Only the YPWM (FDWM) motif of these peptides contacts the PBX homeodomain. Of the 39 known HOX proteins that contain a YPWM motif, 36 have a proline at position 2. Mutation of this residue to an alanine has little effect on the formation of cooperative complexes with PBX (*9*, *41*, *42*). HOXA1 peptides with the aspartic acid at position 2 replaced by either a proline or alanine were synthesized and titrated with the DNA-bound extended homeodomain. The mutant peptides bound to the PBX homeodomain in the same way as HOXA1. For HOXA1 and HOXD4, at least, differences in heterodimer binding site selectivity are not a result of different modes of pentapeptide interaction with the PBX homeodomain.

Comparison of the solution structure of the truncated PBX homeodomain and the ternary PBX-HOXB1 crystal structure shows that relatively small changes in protein conformation induced by binding interactions can have a large effect on the properties of a protein. The formation of trimeric PBX-HOX-MEIS or PBX-HOX-PREP complexes that enhance transcriptional activation or DNA binding have recently been reported (*43*−*48*). In addition, the N-terminus of PBX was found to increase the efficiency of heterodimer formation with HOXD4 and HOXD9. The presence of other components in the transcriptional apparatus, and regions of the proteins outside of the homeodomains, may be necessary to realize the full specificity of the PBX-HOX-DNA interaction.

## ACKNOWLEDGMENT

We thank Andrew Nong for purification of PBX constructs and HOX-derived peptides. Spectra at 750 MHz were acquired at the Environmental Molecular Sciences Laboratory (a national scientific user facility sponsored by the DOE Office of Biological and Environmental Research) located at Pacific Northwest National Laboratory, operated by Battelle for the DOE. T.S. is supported by a MRC Doctoral Award. K.G. and M.F. are Chercheur-Boursiers of the Fonds de la Recherche en Santé du Québec.

## REFERENCES

1. Gehring, W. J., Qian, Y. Q., Billeter, M., Furukubo, T. K., Schier, A. F., Resendez, P. D., Affolter, M., Otting, G., and Wüthrich, K. (1994) *Cell 78*, 211−23.
2. Chang, C. P., Brocchieri, L., Shen, W. F., Largman, C., and Cleary, M. L. (1996) *Mol. Cell. Biol. 16*, 1734−45.
3. Chan, S. K., and Mann, R. S. (1996) *Proc. Natl. Acad. Sci. U.S.A. 93*, 5223−8.
4. van Dijk, M. A., and Murre, C. (1994) *Cell 78*, 617−24.
5. Chang, C. P., Shen, W. F., Rozenfeld, S., Lawrence, H. J., Largman, C., and Cleary, M. L. (1995) *Genes Dev. 9*, 663−74.
6. Lu, Q., Knoepfler, P. S., Scheele, J., Wright, D. D., and Kamps, M. P. (1995) *Mol. Cell. Biol. 15*, 3786−95.
7. Johnson, F. B., Parker, E., and Krasnow, M. A. (1995) *Proc. Natl. Acad. Sci. U.S.A. 92*, 739−43.
8. Phelan, M. L., Rambaldi, I., and Featherstone, M. S. (1995) *Mol. Cell. Biol. 15*, 3989−97.
9. Shanmugam, K., Featherstone, M. S., and Saragovi, H. U. (1997) *J. Biol. Chem. 272*, 19081−7.
10. Lu, Q., and Kamps, M. P. (1996) *Mol. Cell. Biol. 16*, 1632−40.
11. Knoepfler, P. S., and Kamps, M. P. (1995) *Mol. Cell. Biol. 15*, 5811−19.
12. Phelan, M. L., and Featherstone, M. S. (1997) *J. Biol. Chem. 272*, 8635−43.
13. Bürglin, T. R. (1997) *Nucleic Acids Res. 25*, 4173−80.
14. Peltenburg, L. T., and Murre, C. (1997) *Development 124*, 1089−98.
15. Monica, K., Galili, N., Nourse, J., Saltman, D., and Cleary, M. L. (1991) *Mol. Cell. Biol. 11*, 6149−57.
16. Rauskolb, C., Peifer, M., and Wieschaus, E. (1993) *Cell 74*, 1101−1112.
17. Bürglin, T. R., and Ruvkun, G. (1992) *Nat. Genet. 1*, 319−20.
18. Flegel, W. A., Singson, A. W., Margolis, J. S., Bang, A. G., Posakony, J. W., and Murre, C. (1993) *Mech. Dev. 41*, 155−61.
19. Green, N. C., Rambaldi, I., Teakles, J., and Featherstone, M. S. (1998) *J. Biol. Chem. 273*, 13273−9.
20. Piper, D. E., Batchelor, A. H., Chang, C.-P., Cleary, M. L., and Wolberger, C. (1999) *Cell 96*, 587−597.
21. Passner, J. M., Ryoo, H. D., Shen, L., Mann, R. S., and Aggarwal, A. K. (1999) *Nature 397*, 714−9.
22. Jabet, C., Gitti, R., Summers, M. F., and Wolberger, C. (1999) *J. Mol. Biol. 291*, 521−30.
23. Pons, J. L., Malliavin, T. E., and Delsuc, M. A. (1996) *J. Biomol. NMR 8*, 445−52.
24. Farrow, N. A., Muhandiram, R., Singer, A. U., Pascal, S. M., Kay, C. M., Gish, G., Shoelson, S. E., Pawson, T., Forman-Kay, J. D., and Kay, L. E. (1994) *Biochemistry 33*, 5984−6003.
25. Bartels, C., Xia, T.-H., Billeter, M., Guntert, P., and Wüthrich, K. (1995) *J. Biomol. NMR 5*, 1−10.
26. Qian, Y. Q., Resendez, P. D., Gehring, W. J., and Wüthrich, K. (1994) *Proc. Natl. Acad. Sci. U.S.A. 91*, 4091−5.
27. Palmer, A. G., Rance, M., and Wright, P. E. (1991) *J. Am. Chem. Soc. 113*, 4371−80.

28. Stone, M. J., Fairbrother, W. J., Palmer, A. G. I., Reizer, J., Saier, M. H. J., and Wright, P. E. (1992) *Biochemistry 31*, 4394−406.
29. Wishart, D. S., Sykes, B. D., and Richards, F. M. (1992) *Biochemistry 31*, 1647−51.
30. Gehring, W. J., Affolter, M., and Bürglin, T. (1994) *Annu. Rev. Biochem. 63*, 487−526.
31. Qian, Y. Q., Furukubo, T. K., Resendez, P. D., Müller, M., Gehring, W. J., and Wüthrich, K. (1994) *J. Mol. Biol. 238*, 333−45.
32. Weiler, S., Gruschus, J. M., Tsao, D. H., Yu, L., Wang, L. H., Nirenberg, M., and Ferretti, J. A. (1998) *J. Biol. Chem. 273*, 10994-1000.
33. Qian, Y. Q., Billeter, M., Otting, G., Müller, M., Gehring, W. J., and Wüthrich, K. (1989) *Cell 59*, 573−80.
34. Tsao, D. H., Gruschus, J. M., Wang, L. H., Nirenberg, M., and Ferretti, J. A. (1995) *J. Mol. Biol. 251*, 297−307.
35. Schott, O., Billeter, M., Leiting, B., Wider, G., and Wüthrich, K. (1997) *J. Mol. Biol. 267*, 673−83.
36. Ippel, H., Larsson, G., Behravan, G., Zdunek, J., Lundqvist, M., Schleucher, J., Lycksell, P. O., and Wijmenga, S. (1999) *J. Mol. Biol. 288*, 689−703.
37. Esposito, G., Fogolari, F., Damante, G., Formisano, S., Tell, G., Leonardi, A., Di, L. R., and Viglino, P. (1996) *Eur. J. Biochem. 241*, 101−13.
38. Knoepfler, P. S., Lu, Q., and Kamps, M. P. (1996) *Nucleic Acids Res. 24*, 2288−94.
39. Wilson, D. S., Sheng, G., Jun, S., and Desplan, C. (1996) *Proc. Natl. Acad. Sci. U.S.A. 93*, 6886−91.
40. Tucker-Kellogg, L., Rould, M. A., Chambers, K. A., Ades, S. E., Sauer, R. T., and Pabo, C. O. (1997) *Structure 5*, 1047−54.
41. Shen, W. F., Chang, C. P., Rozenfeld, S., Sauvageau, G., Humphries, R. K., Lu, M., Lawrence, H. J., Cleary, M. L., and Largman, C. (1996) *Nucleic Acids Res. 24*, 898−906.
42. Neuteboom, S. T., Peltenburg, L. T., van, D. M., and Murre, C. (1995) *Proc. Natl. Acad. Sci. U.S.A. 92*, 9166−70.
43. Swift, G. H., Liu, Y., Rose, S. D., Bischof, L. J., Steelman, S., Buchberg, A. M., Wright, C. V., and MacDonald, R. J. (1998) *Mol. Cell. Biol. 18*, 5109−20.
44. Shen, W. F., Rozenfeld, S., Kwong, A., Köm ves, L. G., Lawrence, H. J., and Largman, C. (1999) *Mol. Cell. Biol. 19*, 3051−61.
45. Jacobs, Y., Schnabel, C. A., and Cleary, M. L. (1999) *Mol. Cell. Biol. 19*, 5134−42.
46. Shanmugam, K., Green, N. C., Rambaldi, I., Saragovi, H. U., and Featherstone, M. S. (1999) *Mol. Cell. Biol. 19*, 7577−88.
47. Berthelsen, J., Zappavigna, V., Mavilio, F., and Blasi, F. (1998) *EMBO J. 17*, 1423−33.
48. Ferretti, E., Schulz, H., Talarico, D., Blasi, F., and Berthelsen, J. (1999) *Mech. Dev. 83*, 53−64.

BI0001067