MOLECULAR AND CELLULAR BIOLOGY, June 1994, p. 3938–3948
0270-7306/94/$04.00+0

Vol. 14, No. 6

Fusion with E2A Converts the Pbx1 Homeodomain Protein into a Constitutive Transcriptional Activator in Human Leukemias Carrying the t(1;19) Translocation

QIANG LU,[1] DWAINE D. WRIGHT,[2] AND MARK P. KAMPS[2]*

Department of Chemistry[1] and Department of Pathology, School of Medicine,[2] University of California, San Diego, La Jolla, California 92093

Received 7 January 1994/Returned for modification 4 March 1994/Accepted 16 March 1994

***E2A-PBX1* is a chimeric gene formed by the t(1;19)(q23;p13.3) chromosomal translocation of pediatric pre-B-cell leukemia. The E2A-Pbx1 fusion protein contains sequences encoding the transactivation domain of E2A joined to a majority of the Pbx1 protein, which contains a novel homeodomain. Earlier, we found that expression of E2A-Pbx1 causes malignant transformation of NIH 3T3 fibroblasts and induces myeloid leukemia in mice. Here we demonstrate that the homeodomains encoded by *PBX1*, as well as by the highly related *PBX2* and *PBX3* genes, bind the DNA sequence ATCAATCAA. E2A-Pbx1 strongly activates transcription in vivo through this motif, while Pbx1 does not. This finding suggests that E2A-Pbx1 transforms cells by constitutively activating transcription of genes regulated by Pbx1 or by other members of the Pbx protein family.**

A large number of genes contain a conserved 180-bp sequence termed the homeobox, which encodes a sequence-specific, DNA-binding domain called the homeodomain (23, 35, 42). Homeoboxes were first discovered in a set of *Drosophila* genes whose mutation caused entire body segments to differentiate along the pathways programmed for other body segments; therefore, homeodomain proteins were proposed to be master regulators of developmental programs (25, 26, 36, 42). Homeodomain proteins have proven functions as both activators (6) and repressors (10, 11) of transcription and are proposed to regulate development by regulating transcription of target genes containing their specific DNA recognition sequences. It is now clear that all higher organisms also contain a large family of homeobox genes (35) and that these genes are essential for both normal development and normal gene expression in differentiated cells (23, 35).

While the appropriate expression of homeobox genes can regulate normal differentiation, the inappropriate expression of homeobox genes, as well as the expression of mutant homeodomain proteins, is suggested to cause abnormal differentiation in human T-cell (4, 12, 20) and B-cell (19, 32) leukemias. In this report, we have focused on characterizing the biochemical properties of a mutant form of the Pbx1 homeodomain protein, which is created by the t(1;19) chromosomal translocation in human pre-B-cell acute lymphocytic leukemia. *PBX1* was first identified as a gene on chromosome 1 that was located at the breakpoint of the t(1;19) chromosomal translocation, which is found in 20% of pediatric pre-B-cell acute lymphocytic leukemias (19, 32). This translocation joins the 5′ half of the *E2A* transcription factor gene, which encodes the amino-terminal half of E2A, with the majority of the *PBX1* homeobox gene beginning with sequences that encode residue 90 of Pbx1 (19, 28, 32) (Fig. 1A). The normal E2A protein is a transcriptional activator and is one of many factors that activate expression of the immunoglobulin kappa light-chain gene (13, 29). It belongs to the helix-loop-helix family of transcription factors and must dimerize with itself or with other members of this family in order to bind DNA (30). The dimerization and DNA-binding functions of E2A are closely linked in a domain located in the carboxy-terminal half of the protein (30). The amino-terminal half of E2A contains a transactivation domain that can confer a strong transactivation function upon other heterologous DNA-binding domains (13). In the *E2A-PBX1* fusion gene, sequences that encode the dimerization and DNA-binding functions of E2A have been replaced by Pbx1 sequences, including those that encode the Pbx1 homeodomain. While *PBX1* is normally not expressed in pre-B cells (18, 28), *E2A* is strongly expressed in the B-cell lineage, and therefore the *E2A* promoter drives efficient expression of the *E2A-PBX1* fusion gene in pre-B leukemic cells containing the t(1;19) translocation. Alternative splicing of *E2A-PBX1* transcripts leads to the production of two different E2A-Pbx1 proteins, designated E2A-Pbx1a and E2A-Pbx1b, which vary in their carboxyl-terminal sequences (Fig. 1A).

Transformation assays have clearly established that both E2A-Pbx1a and E2A-Pbx1b are oncoproteins. Expression of either form of E2A-Pbx1 at levels approximately 20-fold greater than those of endogenous Pbx1 induces malignant transformation of NIH 3T3 fibroblasts (18). In a mouse hematopoietic model, reconstitution with bone marrow infected by a retrovirus expressing E2A-Pbx1a produces acute myeloid leukemia (17), and mice transgenic for constructs expressing either E2A-Pbx1a or E2A-Pbx1b driven by the immunoglobulin heavy-chain enhancer develop T-cell leukemia (1).

On the basis of the hypothesis that Pbx1 is also a sequence-specific transcription factor, a simple model explaining the molecular basis of transformation by E2A-Pbx1 suggests that fusion with E2A alters the biochemical function of Pbx1 and that E2A-Pbx1 constitutively activates transcription of genes normally regulated by Pbx1 through a Pbx1-binding motif. This model, however, must also consider the fact that *PBX1* represents only one member of a larger *PBX* gene family, which includes at least two other members, *PBX2* and *PBX3*, which

* Corresponding author. Mailing address: Department of Pathology, University of California, San Diego, School of Medicine, 9500 Gilman Dr., La Jolla, CA 92093. Phone: (619) 534-5822. Fax: (619) 534-7415.

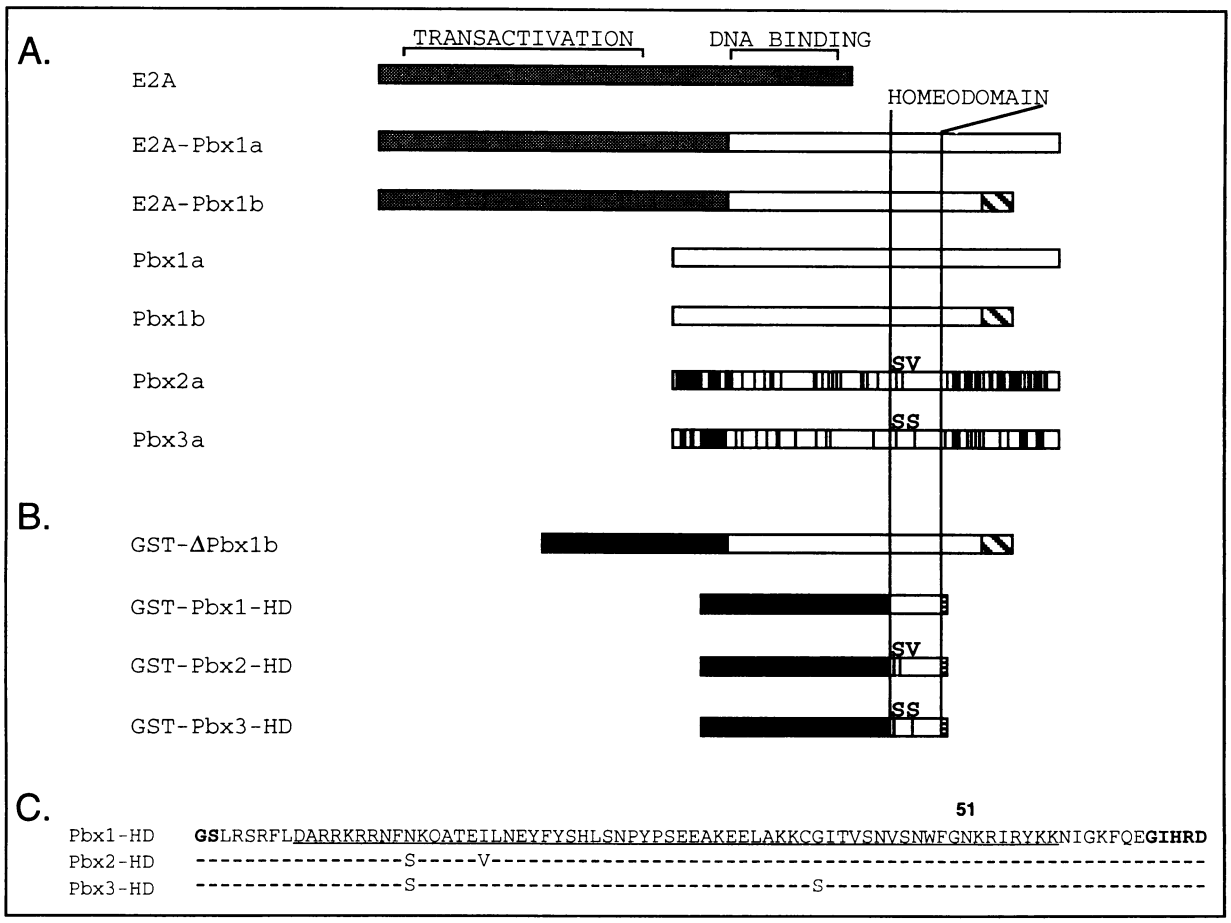

FIG. 1. (A) Structures of E2A, the Pbx protein family, and the fusion proteins formed by rearrangement of the E2A and PBX1 genes. In the E2A-Pbx1 fusion proteins, stippled zones represent E2A sequences and unfilled areas represent Pbx sequences. The cross-hatched region at the carboxyl terminus of E2A-Pbx1b represents unique Pbx1b sequences resulting from differential mRNA splicing. Differential splicing in the normal Pbx1 transcript also produces the corresponding forms of Pbx1, designated Pbx1a and Pbx1b. The vertical bars in Pbx2a and Pbx3a represent amino acid differences from the sequence of Pbx1a. (B) Structures of the recombinant fusion proteins used for selection of binding sites. GST-ΔPbx1b contains residues 90 to 346 of Pbx1b. GST-Pbx-HD proteins contain residues corresponding to 226 to 300 of Pbx1. Filled regions designate GST sequences provided by the expression vector, and unfilled regions correspond to Pbx homeodomain sequences. The horizontal line at the carboxyl terminus of each GST-Pbx-HD fusion protein represents the sequence Gly-Ile-His-Arg-Asp, which is appended to each homeodomain by translation of vector sequences preceding the stop codon. The thrombin cleavage site (Arg ↓ Gly) lies between GST and Pbx sequences, and cleavage results in the addition of a Gly-Ser dipeptide to the amino terminus of each homeodomain. The two single-amino-acid changes in the homeodomains of Pbx2 and Pbx3 are indicated by SV and SS, respectively. (C) Sequences of the Pbx1, Pbx2, and Pbx3 homeodomains used for measurement of K_ds. The homeodomain is underlined in Pbx1-HD and corresponds to residues 226 to 300 of Pbx1. The boldface sequences at the amino and carboxyl termini are encoded by the expression vector, not by the PBX gene. Dashes represent amino acid sequence identities of Pbx2 and Pbx3 with Pbx1. Asparagine 51 (N-51), whose mutation to serine abolishes specific DNA binding, is designated immediately above the sequence of Pbx1-HD.

are also expressed in pre-B cells (28). Because their homeodomains are highly related to that of Pbx1, Pbx2 and Pbx3 may bind the same DNA motif as Pbx1 and E2A-Pbx1. If this were the case, E2A-Pbx1 might need to compete with other Pbx proteins to activate transcription of target genes. To investigate this model for transformation by E2A-Pbx1, we have identified the best DNA-binding sequences for the Pbx1, Pbx2, and Pbx3 homeodomains and have examined differences in the transcriptional functions of E2A-Pbx1 and Pbx1, using reporter constructs driven by the Pbx1-binding motif.

In this report, we demonstrate that the homeodomains of Pbx1, Pbx2, and Pbx3 specifically bind the sequence ATCAATCAA, which we designate the Pbx1-responsive sequence (PRS), and that E2A-Pbx1 strongly activates transcription in vivo of vectors containing the PRS, while Pbx1 does not. These data suggest that the transforming ability of E2A-Pbx1 arises from its distinctive function as a constitutive transactivator and raise the possibility that E2A-Pbx1 induces transformation by stimulating transcription of genes that are normally regulated by Pbx1 or other Pbx proteins through sequences similar to ATCAATCAA.

MATERIALS AND METHODS

Expression and purification of recombinant Pbx proteins. ΔPbx1b was constructed as a glutathione *S*-transferase (GST) fusion protein in pGEX-2T (Pharmacia) (37) as described previously (18) and contains residues 90 to 346 of Pbx1b, which represents all of the Pbx1 sequences contained in E2A-Pbx1b (Fig. 1A). The construction of GST-Pbx-HD (HD stands for

homeodomain) proteins is described below. For purification of recombinant Pbx proteins, 200 ml of NP42 cells containing the GST-Pbx expression vectors were grown at 37°C to an A_{600} of 0.5. Expression of GST-Pbx proteins was induced by addition of 0.2 mM isopropylthiogalactopyranoside (IPTG) for an additional 3 h. Cells were collected by centrifugation and lysed by three cycles of freezing and thawing in 5 ml of 20 mM Tris (pH 7.5)–1 mM EDTA–100 mM NaCl (buffer B). Insoluble material was removed by centrifugation, and the soluble fraction was applied to a 1-ml glutathione-agarose column preequilibrated with buffer B at 4°C. The column was washed three times with 3 ml of buffer B. ΔPbx1b sequences were cleaved from GST sequences by adding 1 ml of buffer B containing thrombin (100 U/ml) to the column, sealing both ends, and rocking the column for 30 min at room temperature. ΔPbx1b was recovered by collecting the flowthrough from the column after the 30-min cleavage reaction. The protein was stored at −70°C in buffer B containing 50% glycerol (0.5× buffer B, 50% glycerol).

Construction of recombinant fusion proteins between GST and the homeodomains of Pbx1, Pbx2, and Pbx3. PCR was used to amplify HeLa cell cDNA sequences encoding the homeodomains of Pbx1, Pbx2, and Pbx3, each of which is represented in Fig. 1C. Each of the 5′ oligonucleotides used for PCR contained a *Bam*HI site positioned such that subsequent cloning of the *Bam*HI-cut PCR product into the expression vector pGEX-2T resulted in an in-frame fusion between the GST vector sequences and the homeodomain. An *Eco*RI site was included in the 3′ amplification oligonucleotide such that translation of the amplified homeobox fragment in pGEX-2T would append the same five amino acids, Gly-Ile-His-Arg-Asp, to the carboxyl terminus of each homeodomain prior to the vector-encoded termination codon. Clones were sequenced to verify their identity with the published sequences of Pbx1, Pbx2, and Pbx3 (28). Oligonucleotides used to amplify Pbx1-HB were GACAG GATCCCTGCGTTCCCGATTTCTGGAT (5′) and ATAGGTA AATTTCAAGAGGAATTCCA (3′). Oligonucleotides used to amplify the Pbx2-HB were GACAGGATCCCTGCGCTCCCGTT TCCTGGAT (5′) and ATCGGAAAGTTCCAAGAGGAAT TCCA (3′). Oligonucleotides used to amplify the Pbx3-HB were CAGGATCCTTAAGATCAAGGTTCCTTGAT (5′) and ATTG GCAAGTTTCAGGAAGAATTCCA (3′). Production of the GST-Pbx-HD proteins and cleavage of Pbx-HD sequences from GST sequences are described above. The N51S mutant of Pbx1-HD was constructed by oligonucleotide-directed mutagenesis.

Oligonucleotides and DNA probes. Oligonucleotides containing 30 internal random bases that were used for the selection of the Pbx1-binding site had the sequence CGCGGATCCTGCAGCTC GAG...30 random bases...GTCGACAAGCTTCTAGAGCA. Competitor oligonucleotides were the PRS oligonucleotide (CAG TACATCAATCAAATGGTC), the CATCAA oligonucleotide (CCAGTACATCAAATGGTCC), the CGTGA oligonucleotide (CCAGTACGTGAAATGGTCC), and a nonspecific oligonucleotide designated Nde I (TATGACGCGTACGTAGATCTGGTA ACCAA). The chloramphenicol acetyltransferase (CAT) reporter construct driven by seven tandem Pbx1-binding-site repeats was made by using the oligonucleotide TCGACGCCTCATCAAT CAAATCGGC. DNA probes for measurement of the apparent equilibrium dissociation constant (K_d) contained the specified oligonucleotides cloned into the *Sma*I site of pBSK$^-$ and were excised from these plasmids by digestion with *Bam*HI and *Eco*RI. All specific probes therefore had the following general sequence: GATCCCCC—specific oligonucleotide—GGGCTGCAGG.

Selection of DNA recognition sequences of Pbx1. A method employing PCR amplification in conjunction with affinity purification was used to isolate the Pbx1 DNA recognition sequence. A schematic representation is shown in Fig. 2A. In brief, 20 pmol of a population of double-stranded 70-mers containing 30 random internal nucleotides flanked by 20-nucleotide multiple cloning sites (MCSs) was synthesized from a single-stranded template, a 3′ primer complementary to the MCS, and PCR extension. A second cycle of PCR using both 5′ and 3′ oligonucleotides complementary to the MCSs was used to amplify the library twofold. This library contains 10^{13} different internal 30-mer sequences. One-tenth (10 μl) of this library of double-stranded oligonucleotides was then incubated for 30 min at 4°C with 25 μl of recombinant GST-ΔPbx1b immobilized on glutathione-Sepharose, 25 μl of poly(dI-dC) (1.0 mg/ml), 25 μl of bovine serum albumin (BSA; 1.0 mg/ml), 15 μl of water, and 25 μl of 5× binding buffer (125 mM *N*-2-hydroxyethylpiperazine-*N*′-2-ethanesulfonic acid [HEPES; pH 7.5], 500 mM KCl, 5 mM EDTA, 50 mM $MgCl_2$, 0.5% Nonidet P-40, 25% glycerol, 5 mM dithiothreitol). Unbound oligonucleotides were removed by washing the resin four times in 500 μl of 1× binding buffer, and bound oligonucleotides were removed by boiling the resin in 50 μl of water. Ten microliters of oligonucleotides released from the resin by boiling in water was subject to 10 cycles of PCR amplification. Amplified DNA was then used for consecutive rounds of selection and amplification. Oligonucleotides obtained after six rounds of selection were cloned and sequenced, and the sequences of the 30 internal bases are designated the F (first) series in Fig. 2B. The same population of oligonucleotides obtained after six rounds of PCR selection was used for complex formation with recombinant Pbx1 in an electrophoretic mobility shift assay (EMSA). Two closely spaced gel shift complexes were observed. The DNA bound to Pbx1 in each of these complexes was isolated, amplified, cloned, and sequenced and is represented by the S (second) series of oligonucleotides in Fig. 2B. No differences were observed in the content of specific sequences in oligonucleotides isolated from each complex, and we would suggest that the lower complex contained a slightly smaller form of Pbx1b generated by proteolysis during isolation of GST-ΔPbx1b from bacteria.

Calculation of K_d. K_ds were defined by the formula K_d = [DNA][HD]/[HD-DNA], where [DNA] is the concentration of free DNA, [HD] is the concentration of free protein, and [HD-DNA] is the concentration of the C1 homeodomain-DNA complex (see Fig. 6B). The concentration of all ^{32}P-labeled DNA probes was 0.32 nM, which was greater than 10-fold below the lowest K_d. The concentration of Pbx-HD was kept at least 10-fold in excess of that of DNA, allowing the approximation [HD] ≅ [HD_{total}] to be used in calculating K_ds. The DNA probes contained the target sequence CATCAAT CAA, CATCAA, or CGTGAA and are described above. K_ds were calculated from binding reactions containing Pbx1-HD, Pbx2-HD, or Pbx3-HD at concentrations of 3.4 and 10 nM for the probe containing CATCAATCAA, at concentrations of 10 and 30 nM for the protein containing CATCAA, and at concentrations of 30 and 91 nM for the probe containing CGTGAA. The value of [DNA]/[HD-DNA] was determined by taking the ratio of free counts per minute to bound counts per minute.

Preparation of nuclear extracts. Cultured cells were harvested by centrifugation. Cell pellets were resuspended in six volumes of buffer A (10 mM HEPES [pH 7.9], 1.5 mM $MgCl_2$, 10 mM KCl, 0.5 mM dithiothreitol, 1 mM phenylmethylsulfonyl fluoride), swelled for 10 min on ice, and disrupted by Dounce homogenization. Nuclei were collected by centrifugation at 3,000 rpm for 15 min at 4°C, resuspended in 16 volumes of buffer A, and lysed by addition of ammonium sulfate to 0.3 M followed by incubation at 4°C for 45 min. The viscous extract was centrifuged 1 h in an SW41 rotor at 40,000 rpm. The supernatant was transferred to a new tube, and transcription factors were precipitated by dissolving 0.2 g of solid ammonium sulfate per ml and incubating the mixture for 30 min on ice. The precipitated proteins were collected by cen-

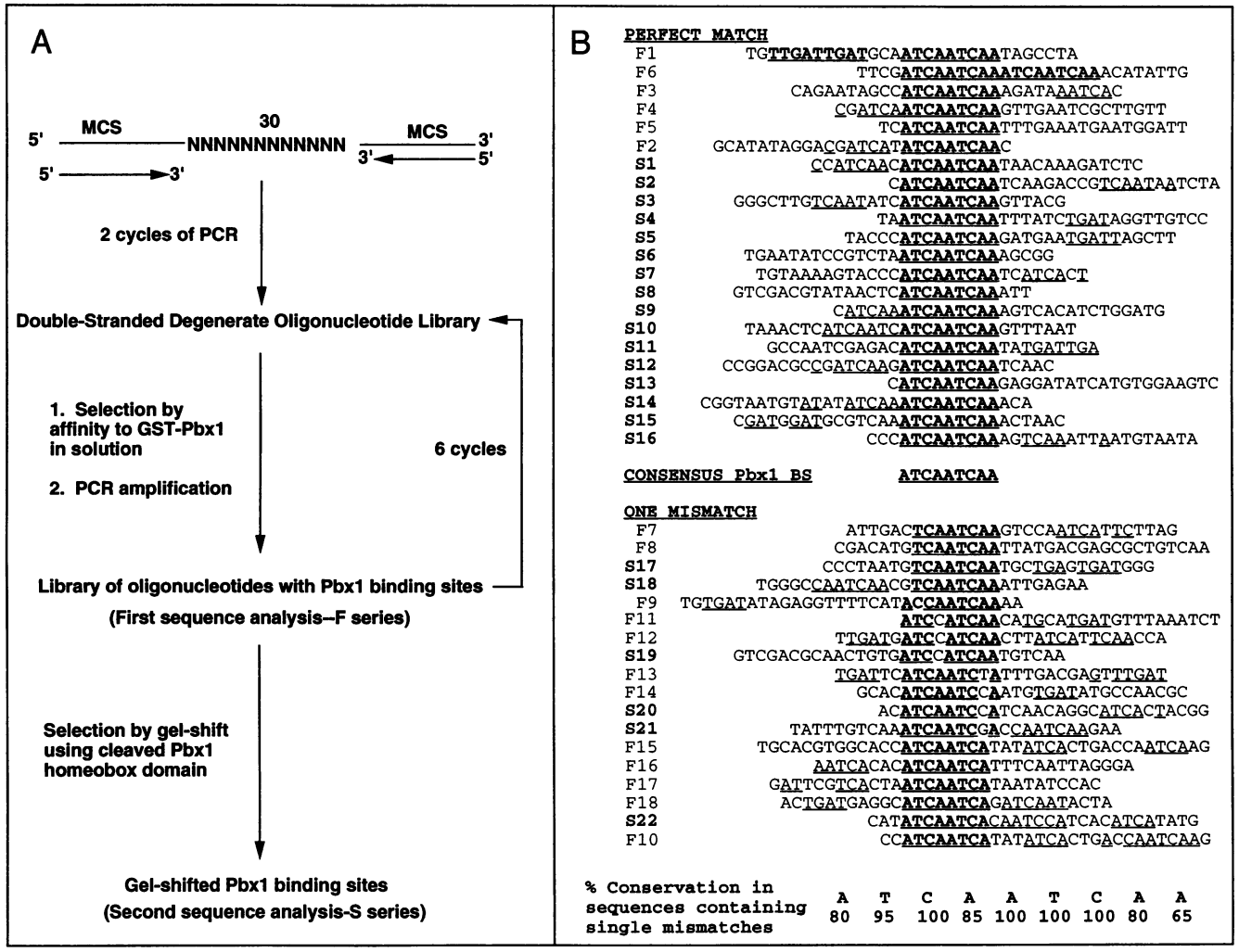

FIG. 2. Selection of DNA-binding sequences by GST-ΔPbx1b. (A) Schematic representation of the procedure used to select PRSs listed in panel B. A degenerate population of double-stranded DNA oligonucleotides containing 30 random internal nucleotides flanked by two 20-nucleotide sequences containing MCSs was synthesized by PCR as described in Materials and Methods. Immobilized GST-ΔPbx1b fusion protein was used in a solution binding protocol (Materials and Methods) to select oligonucleotides that bound Pbx1. After six cycles of selection and amplification, the DNA was cloned and sequenced, and these sequences are represented as the F series (F1 to F18) in panel B. GST-ΔPbx1b was used in EMSA to shift a subset of the pool of PRSs selected by affinity to GST-ΔPbx1b in solution, and the sequences of these oligonucleotides comprise the S series (S1 to S22) in panel B. (B) Comparison of DNA sequences comprising the random core of Pbx1-bound oligonucleotides that contained the consensus binding sequence (BS) or single mismatches with this sequence.

trifugation at 20,000 rpm for 15 min at 4°C in an SW28 rotor. Precipitated proteins were dissolved in buffer D (20 mM HEPES [pH 7.9], 20% glycerol, 0.1 M KCl, 0.2 mM EDTA, 0.5 mM dithiothreitol, 0.5 mM phenylmethylsulfonyl fluoride), dialyzed against 1 liter of buffer D at 4°C overnight, and clarified by centrifugation. Nuclear extracts were stored at −70°C.

DNase I protection assays. DNase I footprinting (16) was performed with 10 fmol of probe, 1 μg of poly(dI-dC) when nuclear extracts were used, and no poly(dI-dC) when recombinant Pbx1 was used; 15 μg of nuclear extracts or 4 to 100 ng of one of the thrombin-cleaved recombinant Pbx proteins was used for each reaction. Reaction mixtures were subjected to digestion with 5 to 15 ng of DNase I on ice for 1 min.

Transfections and CAT assays. A mixture of pCAT or PRS7-pCAT reporter plasmid (0.5 μg) and pGD-derived expression vector (3 μg) was introduced into 10^7 S194 B cells by DEAE-dextran-mediated transfection (9) or into 10^6 COS cells by calcium phosphate coprecipitation (40). The cells were cultured for 36 h in Dulbecco modified Eagle medium containing 10% fetal bovine serum. Preparation of cell extracts and CAT assays were performed as previously described (8). Transfections were performed twice for COS cells and three times for S194 B cells, and the standard deviation between independent measurements varied by less than 20% of the absolute value in all cases. Transfection of 293 epithelial cells yielded the same results as did transfection of S194 B cells and COS cells.

Biosynthetic labeling and immunoprecipitations. Biosynthetic labeling using [^{35}S]methionine and immunoprecipitations using a polyclonal rabbit anti-Pbx1 serum were performed as described previously (18).

pGD retroviral expression vectors. pGD is a murine retroviral vector that expresses the neomycin phosphotransferase

cDNA from an internal simian virus 40 early promoter. Pbx1 and E2A-Pbx1 cDNAs were cloned into a *BclI* site immediately 3′ of the 5′ long terminal repeat as described previously (18).

RESULTS

Pbx1 selectively binds (A/C)ATCAATCAA(A/T). A double-stranded DNA library of possible Pbx1 recognition sites was synthesized from a population of single-stranded oligonucleotides containing a central segment of 30 random nucleotides flanked on each side by 20 defined nucleotides (Fig. 2A; Materials and Methods). In this method, two cycles of PCR are used to convert the single-stranded oligonucleotides to the double-stranded form, using primers that anneal to each of the defined 5′ and 3′ MCS sequences. The double-stranded oligonucleotides were subjected to affinity selection, using a fusion protein between GST and the Pbx1 sequences found in E2A-Pbx1b (termed GST-ΔPbx1b [18]), and amplified by PCR. Oligonucleotides purified by six selection cycles were cloned, and 25 were sequenced. These sequences are designated the F series (F1 to F18) in Fig. 2B. The consensus, ATCAATCAA, was found in 25% (F1 to F6), while 40% contained the consensus having a single mismatch (F7 to F18). The remaining 35% contained two or more mismatches with the consensus sequence but always contained the sequence ATCA. Two oligonucleotides contained duplicate ATCAATCAA sequences, one as a tandem direct repeat (F6) and the second as an inverted repeat spaced by 3 bp (F1). Many oligonucleotides containing an ATCAATCAA sequence also contained shorter segments of this sequence, particularly ATCA and ATCAA (underlined in Fig. 2B). These results suggested that Pbx1 bound most specifically to the canonical sequence, ATCAAT CAA, and exhibited a lesser affinity for variants of this sequence.

A final seventh round of selection by native gel electrophoresis using GST-ΔPbx1b to shift a fraction of the amplified oligonucleotide pool from the sixth round of affinity and PCR purification resulted in 50% of the selected sequences containing ATCAATCAA, which represented an additional twofold enrichment for this sequence. The sequences of 34 of these oligonucleotides are represented by the S series in Fig. 2B. Enrichment for ATCAATCAA suggested that Pbx1 has a stronger affinity for this sequence than it does for either ATCA or ATCAA alone. Subsequent measurement of K_ds (see below) confirmed that Pbx1 binds ATCAA with a somewhat lower affinity than ATCAATCAA. Comparison of all AT CAATCAA sites also revealed that A or C and A or T are preferred at the immediate 5′ and 3′ ends, respectively, of the ATCAATCAA consensus. Analysis of variant binding motifs suggested that the first C and the second ATC (ATCAAT CAA) were invariant and that changes at A in the fourth position strongly favored C.

The isolated homeodomains of Pbx1, Pbx2, and Pbx3 select similar DNA motifs from populations of degenerate oligonucleotides and bind ATCAATCAA with comparable affinities. To determine whether specific DNA binding was achieved through the homeodomain and to compare the sequence specificities of the homeodomains of Pbx1, Pbx2, and Pbx3, GST fusion proteins encoding only the homeodomain from each Pbx protein were constructed (Fig. 1B; Materials and Methods), purified, and used in the same affinity selection procedure described for Fig. 2A. The final selection step was accomplished by isolation of the gel shift complex formed between the sixth round of affinity-selected oligonucleotides and each of the three purified Pbx homeodomains, which had been cleaved from adjacent GST sequences by incubation with thrombin. These homeodomains are designated Pbx1-HD, Pbx2-HD, and Pbx3-HD. Twelve oligonucleotides selected by each homeodomain were sequenced, and a subset is shown in Fig. 3. In all cases, the recurring selected sequences were ATCAATCAA and portions of this sequence, usually containing ATCA or ATCAA. Similar to Pbx1, the homeodomains of Pbx2 and Pbx3 also exhibited a preference for a 5′ C residue immediately preceding the ATC sequence (50%). This analysis suggested that the two amino acid differences in the homeodomains of Pbx2 and Pbx3 did not affect their DNA-binding specificities.

GST-PBX1-HD		
1-1	GAG	ACC**ATCAATCA**TTATTAAAATCGC**TGATGG** GTC
1-2	GAG	CTAAACTGTTTAC**TTGATAAGTTGATGG** GTC
1-3	GAG	TTGGAT**TTGAT**ATGTCTACGAGTTTGACGA GTC
1-4	GAG	TACTC**ATCAA**ACTTCC**TT**TC**TTGAT**GTCGA GTC
1-5	GAG	**TGAT**GTGTGACCTC**ATCA**TATA**AGTGATT**A GTC
1-6	GAG	G**TGAT**AATAAT**TTGAT**TAGTTTGAC**GATTG** GTC
1-7	GAG	AATTACATCCA**ATCAATCA**TATTTAGTTAA GTC
1-8	GAG	AGTTGAATAGTACTG**TGAT**ATCT**TTGATTG** GTC
1-9	GAG	GAGAC**ATCA**T**ATCAA**ATGAGATGTCATTTA GTC
1-10	GAG	**TC**C**ATCAA**ACGT**ATCA**TGCGTCACTCACTA GTC
GST-PBX2-HD		
2-1	GAG	**AT**AT**ATCAA**ATACT**TTGAT**TGGTATCGGAT GTC
2-2	GAG	TATGCGATCC**ATCAATCAA**CAT**ATTGAT**GA GTC
2-3	GAG	**TGATTGAT**GGAAGT**TTGA**G**TGAT**CGCAGTA GTC
2-4	GAG	**GA**G **TGAT**ATGTCAAATTTTGTTATTT**TTGATTG** GTC
2-5	GAG	AACTATTTAACA**ATC**C**ATCA**CACTC**ATCA**G GTC
2-6	GAG	CGATGAGCC**ATCAA**AGTTAC**ATCA**GAAACG GTC
2-7	GAG	**TTGATTGA**AGT**TTGATTAA**GGTTATGTCAA GTC
2-8	GAG	GAC**ATCA**TT**TCAA**ATTATG**ATCA**CG**TGATGG** GTC
GST-PBX3-HD		
3-1	GAG	AACGAG**ATCAATCAATCAA**CTATGTCATGA GTC
3-2	GAG	GTAGGGTAATC**ATCAA**AC**ATCAATCAATCA** GTC
3-3	GAG	GTC**ATCA**CTC**ATCAATCAA**ATTTTTTACTA GTC
3-4	GAG	TCA**TTGATT**CTACTC**TGATTGA**CACGCTGA GTC
3-5	GAG	TACATCTACGGAGTCT**TTGATTGATTGATG** GTC
3-6	GAG	TCATTAGTGAG**TGATTTAT**GACGGACATTG GTC
3-7	GAG	ATAAG**TGAT**GGATTTA**TGATT**AAC**ATCA**GG GTC
3-8	GAG	TACCACA**ATCAATCAATCAA**C**ATCA**CCCCG GTC
3-9	GAG	TACGTCAATATTGTTCCATG**TTGATTGAT**C GTC
3-10	GAG	ATACT**TTGATTGAT**GTCTC**TGAT**ACGTCTG GTC

FIG. 3. Oligonucleotide sequences selected by GST fusion proteins containing the homeodomains of Pbx1, Pbx2, and Pbx3. Sequences obtained from cloned oligonucleotides selected by GST-Pbx1-HD, GST-Pbx2-HD, and GST-Pbx3-HD are shown. The first and last three nucleotides in each sequence represent conserved bases flanking the 30 degenerate positions. Boldface and underlined sequences represent portions of the consensus DNA sequence, ATCAATCAA (Fig. 2).

The K_ds of Pbx1-HD, Pbx2-HD, and Pbx3-HD for oligonucleotides containing CATCAATCAA, CATCAA, or CGTGA were next calculated to determine whether the various homeodomains exhibited similar affinities for the same DNA motifs and to determine which sequence exhibited the highest affinity for Pbx1. The affinity of each homeodomain for each DNA probe was measured by EMSA, and the K_d represents the concentration of Pbx homeodomain required to shift 50% of the DNA probe (Materials and Methods). In this analysis, low concentrations of homeodomain are required to shift a high-affinity binding site, whereas higher concentrations of the homeodomain are required to shift lower-affinity sites. The profile of binding between increasing concentrations of Pbx1-HD and a probe containing CATCAATCAA is illustrated in Fig. 6B. The first complex to form at low concentrations of Pbx1-HD is designated C1. Higher concentrations of Pbx1-HD, as well as of Pbx2-HD and Pbx3-HD, resulted in the

TABLE 1. K_ds for Pbx1-HD, Pbx2-HD, and Pbx3-HD

Sequence[a]	K_d (nM) ± 1 SD[b]		
	Pbx1-HD	Pbx2-HD	Pbx3-HD
TACATCAATCAAATGG	6.5 ± 0.8 (5)	5.8 ± 1.7 (5)	7 ± 2.8 (5)
TACATCAA ATGG	14 ± 2.5 (4)	19 ± 6.0 (4)	15 ± 2.3 (4)
TACGTGAA ATGG	217 ± 107 (2)	70 ± 10 (3)	152 ± 38 (4)

[a] Nucleotides containing identities with the PRS are underlined.
[b] The number of repetitions used to determine the average K_d is given in parentheses.

formation of larger complexes, the predominant one designated C2. These larger complexes represent the nonspecific binding of Pbx1-HD with the DNA probe because a point mutant substituting serine for asparagine 51 in the Pbx1 homeodomain that abolishes specific DNA-binding does not affect formation of the larger complexes (see Fig. 6B; this mutation is discussed in greater detail below). Therefore, K_d measurements were based on the formation of the C1 complex. Little difference was observed in the K_ds of different Pbx-HD proteins for the same DNA probe (Table 1). Likewise, the relative affinities of each Pbx-HD protein for different DNA sequences were similar, being the highest for CATCAATCAA, 2- to 3-fold lower for CATCAA, and a 10- to 30-fold lower for the double-mutant sequence CGTGAA (Table 1). These data supported the possibility that Pbx1, Pbx2, and Pbx3 also bind the same motif in vivo. We conclude that CATCAATCAA represents the highest-affinity DNA sequence that we have isolated for Pbx proteins, and we have designated the highly conserved ATCAATCAA sequence the Pbx1 recognition sequence (PRS). This PRS was used in subsequent footprinting, gel shift, and transactivation experiments.

Both recombinant Pbx1 and nuclear factors protect the PRS in DNA footprinting analysis. Footprinting analysis was used to reaffirm the specificity of recombinant Pbx1 for the PRS and to investigate whether normal nuclear factors, possibly normal Pbx proteins, recognized this motif (Fig. 4). The probe for this experiment contained two PRSs separated by 22 nucleotides, whose positions were determined relative to the control DNase I digestion pattern (lanes 2 and 9) by comparison with adjacent A+G sequencing reactions (lanes 1 and 10). Both PRS sites were fully protected by 10 and 20 ng of ΔPbx1b (lanes 4 and 5). This finding confirmed the specificity of Pbx1 for the PRS.

Factors in nuclear extracts also protected the PRS, regardless of whether the nuclear extracts contained E2A-Pbx1. Nuclear extracts were derived from NIH 3T3 cells and mouse myeloblasts transformed by E2A-Pbx1 (Fig. 4, lanes 6 and 7, respectively) as well as by NIH 3T3 cells transformed by the $p160^{gag\text{-}abl}$ tyrosine protein kinase (lane 8). Nuclear extracts from normal NIH 3T3 fibroblasts, which express *PBX1*, the human B-cell line EW, which does not express *PBX1*, and the human fetal kidney cell line 293Tag also strongly protected the PRS (not shown). In the upper footprint, the boundaries of sequences protected by nuclear factors were identical to those protected by ΔPbx1b; however, in the lower footprint, the upper boundary was extended by four nucleotides. These data demonstrated that normal cellular proteins bound the PRS but did not prove that these proteins were related to Pbx proteins.

A Pbx1-related protein(s) binds ATCAATCAA in nuclear extracts. EMSA was used to determine whether Pbx1-related proteins bound the PRS specifically. Observation of gel shift complexes with Pbx1-related proteins would support a model in which E2A-Pbx1 might compete with other Pbx proteins in vivo for a common transcriptional regulatory sequence. A major complex, denoted by an arrow in Fig. 5, was found in Nalm-6 pre-B cells, which lack the t(1;19) translocation (Fig. 5, lanes 1 to 5), as well as in 697 pre-B cells, which contain the t(1;19) translocation (lanes 6 to 10), and in 293Tag epithelial cells (lanes 11 to 15). Unexpectedly, nuclear extracts from 697 cells did not exhibit an additional gel shift complex with the PRS probe, even though they contained E2A-Pbx1 by Western blotting (immunoblotting). Nuclear extracts from E2A-Pbx1-transformed NIH 3T3 fibroblasts also failed to produce a unique complex with the PRS (data not shown). While a unique complex containing E2A-Pbx1 was not apparent, the major gel shift complex behaved as though it contained a Pbx

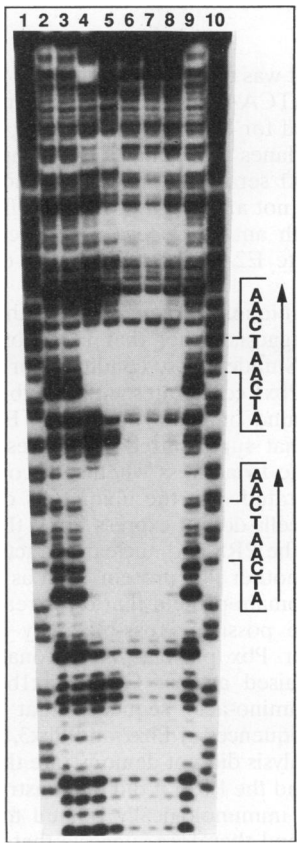

FIG. 4. Footprinting analysis with recombinant ΔPbx1b and nuclear extracts. A radioactive probe containing two PRSs separated by 22 intervening nucleotides was used for this footprinting assay. Lanes: 1 and 10, G-plus-A sequencing reaction; 2 and 9, BSA control; 3 to 5, 5, 10, and 20 ng of recombinant ΔPbx1b, respectively; 6 and 7, 15 μg of nuclear extract from E2A-Pbx1-transformed NIH 3T3 fibroblasts and marrow myeloblasts, respectively; 8, 15 μg of nuclear extract from NIH 3T3 fibroblasts transformed by the $p160^{gag\text{-}abl}$ tyrosine protein kinase.

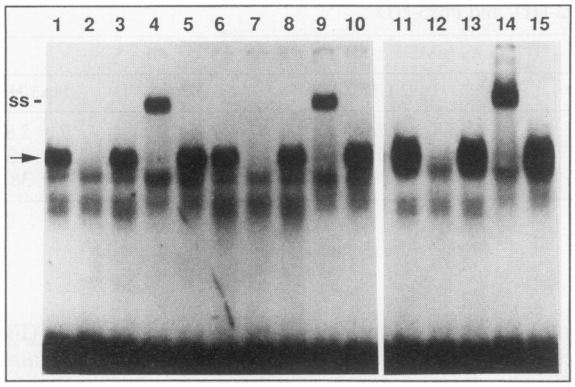

FIG. 5. Gel mobility shift assay with nuclear extracts. Nuclear extracts from Nalm-6 pre-B cells (lanes 1 to 5), 697 pre-B cells (lanes 6 to 10), and 293Tag epithelial cells (lanes 11 to 15) were incubated with the same CATCAATCAA oligonucleotide probe used for K_d measurements. Oligonucleotide competitors or antisera were added as follows: lanes 1, 6, and 11, no additions; lanes 2, 7, and 12, 15 ng of the CATCAATCAA oligonucleotide; lanes 3, 8, and 13, 15 ng of the CATCAA oligonucleotide; lanes 4, 9, and 14, 1 μl of a polyclonal rabbit anti-Pbx1 serum; lanes 5, 10, and 15, 1 μl of a polyclonal rabbit antiserum to the catalytic subunit of protein phosphatase IIA. ss indicates the position of the complex supershifted by antibodies against Pbx1.

protein because it was fully competed for by inclusion of 15 ng of the CATCAATCAA oligonucleotide (lanes 2, 7, and 12), was not competed for by inclusion of 15 ng of the CATCAA oligonucleotide (lanes 3, 8, and 13), was supershifted by two different anti-Pbx1 sera (one of which is shown in lanes 4, 9, and 14), and was not affected by a nonspecific serum (lanes 5, 10, and 15). Both anti-Pbx1 sera were previously shown to immunoprecipitate E2A-Pbx1 and bind it on Western blots (18, 19).

The lack of a unique complex formed with extracts containing E2A-Pbx1 suggests either that E2A-Pbx1 does not bind tightly to the PRS under these conditions or that the complex containing E2A-Pbx1 comigrates with the ubiquitous complex. Antisera that bound only the portion of E2A contained in E2A-Pbx1 and that supershifted complexes containing E2A were not available to address whether a complex containing E2A-Pbx1 comigrated with the ubiquitous complex. Because Nalm-6 and 697 cells do not express Pbx1, the major complex formed between the PRS and nuclear extracts from these cells contains either another Pbx protein, such as Pbx2 or Pbx3, or another homeodomain protein that cross-reacts with antibodies to Pbx1. The possible cross-reactivity of the anti-Pbx1 serum with other Pbx proteins is reasonable because this antiserum was raised against GST-ΔPbx1b (Fig. 1), which contains a 266-amino-acid sequence that is 92 and 94% identical to the sequences of Pbx2 and Pbx3, respectively (28). Although this analysis did not demonstrate that either Pbx1 or E2A-Pbx1 can bind the PRS, it did demonstrate that a normal nuclear factor(s) immunologically related to Pbx1 binds the PRS specifically and therefore suggests that E2A-Pbx1 might need to compete with such a factor(s) in vivo to exhibit an effect on gene transcription.

E2A-Pbx1 activates transcription through the PRS, while Pbx1 does not. While the binding of E2A-Pbx1 to the PRS was not evident by EMSA, the transcriptional effects of E2A-Pbx1 through the PRS in vivo were dramatic. To investigate the transcriptional function of E2A-Pbx1 in vivo and to determine whether fusion with E2A sequences altered the transcriptional activity of Pbx1, we cotransfected COS epithelial cells and S194 B cells with a reporter CAT construct driven by seven tandem PRS sequences (PRS7-pCAT) and a second vector expressing E2A-Pbx1 or Pbx1 proteins (Fig. 6A). While neither Pbx1a nor Pbx1b activated transcription of PRS7-pCAT in this assay (lanes 7 and 9), both E2A-Pbx1a and E2A-Pbx1b strongly activated its transcription (lanes 8 and 10). The PRSs mediated transactivation because the same vector lacking the seven PRSs (pCAT) exhibited no transactivation by either E2A-Pbx1a or E2A-Pbx1b (lanes 3 and 5). Lack of transactivation by Pbx1 proteins was not due to lack of expression of Pbx1, because Pbx1 and E2A-Pbx1 were expressed at approximately equal levels in transfected COS cells (Fig. 6C).

To strengthen the hypothesis that E2A-Pbx1 activates transcription directly through its binding to the PRS, a mutant form of E2A-Pbx1b whose homeodomain did not bind the PRS specifically was tested for its ability to activate transcription through the PRS. The mutation converted asparagine 51, an invariant residue found in all homeodomains, to serine. In the crystal structures of both the Engrailed (En) (21) and yeast α2 (41) homeodomains, Asn-51 forms two hydrogen bonds with the N-6 and N-7 positions of an adenine base that resides near the center of the DNA recognition sequence. We chose to replace Asn-51 with serine because the R group of serine ($-CH_2-OH$) should be too small to form the same stable hydrogen bonds formed by the R group of asparagine ($-CH_2-CO-NH_2$). However, like asparagine, the R group of serine is small, polar, uncharged, and compatible with an alpha-helical structure and therefore should not alter the tertiary structure of the Pbx1 homeodomain. The purified mutant homeodomain, designated Pbx1-N51S-HD, did not bind the PRS specifically, as evidenced by a total lack of C1 complex formation (Fig. 6B). Likewise, E2A-Pbx1b-N51S did not activate transcription of the PRS7-pCAT vector in the cotransfection assay (Fig. 6A, lane 11), even though it was as abundant as wild-type E2A-Pbx1b in transfected COS cells (Fig. 6C, lane 6). This finding strongly suggests that E2A-Pbx1 proteins activate transcription in vivo through direct binding to the PRS and that normal Pbx1 proteins are qualitatively different from E2A-Pbx1, as observed by their inability to induce transcriptional activation.

DISCUSSION

In this report, we establish a fundamental biochemical difference between the Pbx1 and E2A-Pbx1 proteins that may explain how E2A-Pbx1 functions as an oncoprotein. Having identified ATCAATCAA as the best in vitro DNA-binding sequence for ΔPbx1b, we demonstrate that a transcription reporter construct driven by this sequence is efficiently transactivated by both E2A-Pbx1a and E2A-Pbx1b but not by Pbx1a or Pbx1b. Therefore, fusion with E2A modifies the activity of Pbx1 and causes it to behave as a constitutive transcriptional activator in pre-B cells containing the t(1;19) translocation. A mutant form of E2A-Pbx1 that could not bind specifically to the PRS did not activate transcription of the PRS7-pCAT reporter construct (Fig. 6), strongly suggesting that E2A-Pbx1 activates transcription through its direct physical contact with the PRS. Footprinting (Fig. 4) and gel shift (Fig. 5) experiments indicated that E2A-Pbx1 coexists in transformed 697 pre-B cells with other transcription factors that also bind the PRS and that the major binding protein(s) in these cells is immunologically related to Pbx1 (Fig. 5). In vitro, the homeodomains of Pbx2 and Pbx3 were also found to bind the PRS with approximately the same affinity as did the homeodomain of Pbx1 (Table 1). Therefore, because neither Nalm-6 nor 697

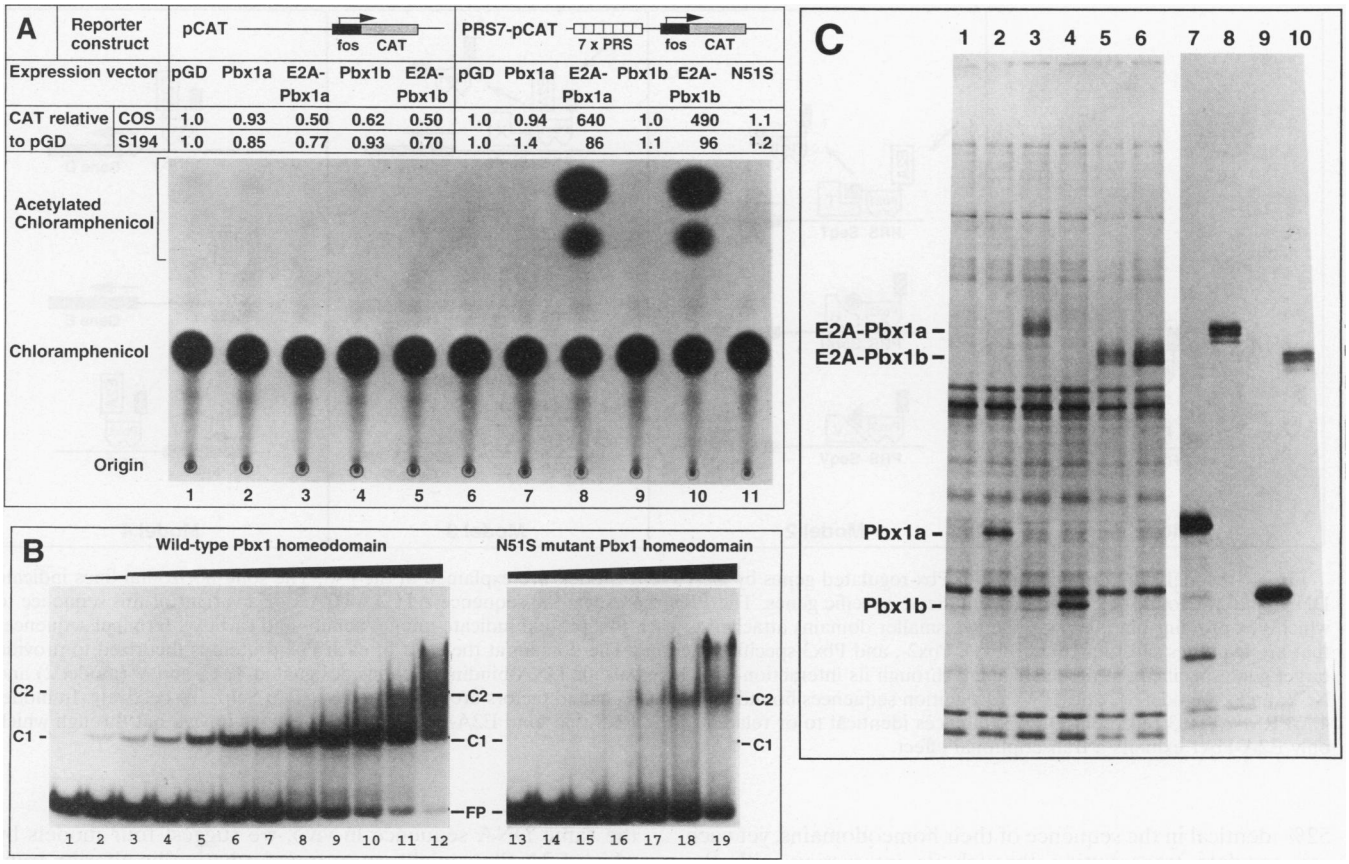

FIG. 6. Comparison of in vivo transcriptional functions of Pbx1, E2A-Pbx1, and a DNA-binding mutant of E2A-Pbx1b. (A) The reporter constructs pCAT (lanes 1 to 5) and PRS7-pCAT (lanes 6 to 11) were transfected into COS epithelial cells or S194 B cells together with the pGD retroviral vector expressing no additional cDNA (lanes 1 and 6) or a cDNA encoding Pbx1a (lanes 2 and 7), E2A-Pbx1a (lanes 3 and 8), Pbx1b (lanes 4 and 9), E2A-Pbx1b (lanes 5 and 10), or E2A-Pbx1b-N51S (lane 11), and the amount of CAT activity was measured after 48 h. The pCAT vector contained the CAT cDNA downstream of a minimal c-*fos* promoter extending 55 bases 5′ of the transcriptional initiation site of c-*fos*. PRS7-pCAT contained seven tandem repeats of the PRS inserted 5′ of the minimal c-*fos* promoter in pCAT. Numbers shown above lanes indicate fold stimulation of CAT activity mediated by Pbx1 or E2A-Pbx1 protein compared with the level for the pGD control (lanes 1 and 6). Numerical values represent averages of two measurements for COS cells and three measurements for S194 cells. The autoradiogram shown illustrates results for COS cells. (B) Effect of substituting serine for asparagine 51 on the DNA-binding affinity of the Pbx1 homeodomain. The DNA probe contained a single cloned ATCAATCAA sequence. Binding reaction mixtures contained either no added protein (lane 1) or increasing amounts of wild-type Pbx1-HD (lanes 2 to 12) or Pbx1N51S-HD, which contains the N51S mutation (lanes 13 to 19). A twofold increase of homeodomain protein was added to consecutive binding reactions, beginning with 0.1 ng of wild-type Pbx1-HD in lane 2 and 1.6 ng of the N51S mutant in lane 13. (C) Abundance of Pbx1 and E2A-Pbx1 in transfected COS cells was measured by immunoprecipitation. Parallel plates of cells used for measurement of CAT activities (A) were subjected to metabolic labeling by addition of [^{35}S]methionine for 6 h, beginning 40 h posttransfection, and immunoprecipitation with antibodies against Pbx1. The expression vectors used in the transfections were as follows: lane 1, pGD; lane 2, pGD encoding Pbx1a; lane 3, pGD encoding E2A-Pbx1a; lane 4, pGD encoding Pbx1b; lane 5, pGD encoding E2A-Pbx1b; and lane 6, pGD encoding the N51S mutant of E2A-Pbx1b. In vitro-transcribed and -translated proteins serving as size markers were as follows: lane 7, Pbx1a; lane 8, E2A-Pbx1a; lane 9, Pbx1b; and lane 10, E2A-Pbx1b.

pre-B cells express Pbx1, the specific gel shift complex formed in nuclear extracts from these cells may contain Pbx2 or Pbx3. Together, these data suggest that normal Pbx proteins bind a motif similar to the PRS in vivo and that E2A-Pbx1 may compete with one or more Pbx proteins to bind this motif, activate transcription of adjacent genes, and induce transformation.

While ATCAATCAA represents the best in vitro Pbx-binding motif, as well as an apparent DNA-binding sequence for E2A-Pbx1 in vivo, only a portion of ATCAATCAA may comprise the binding motif for normal Pbx1, Pbx2, and Pbx3 in vivo. The fushi tarazu (Ftz) homeodomain binds to a consensus CAATTA in vivo, and random selection of binding sites in vitro using the Ftz homeodomain demonstrated that the affinity of its binding to DNA can be increased another threefold by the addition of 5′ AAG and 3′ AG flanking sequences (**AAGCAATTAAG** [7]). Therefore, by analogy with Ftz, Pbx proteins may bind a core motif, such as ATCAA, in vivo and exhibit an elevated affinity in vitro by addition of C to the 5′ end and TCAA to the 3′ end.

The similarity in DNA sequences recognized by the homeodomains of Pbx1, Pbx2, and Pbx3 in vitro suggests that the Pbx1, Pbx2, and Pbx3 may also bind similar sequences in vivo. It is not unusual to find such similarity in the DNA-binding specificities of homeodomain proteins, even among members of different homeodomain families. For example, the homeodomain proteins encoded by the *Drosophila* genes even-skipped (*eve*), engrailed (*en*), and fushi tarazu (*ftz*) are only

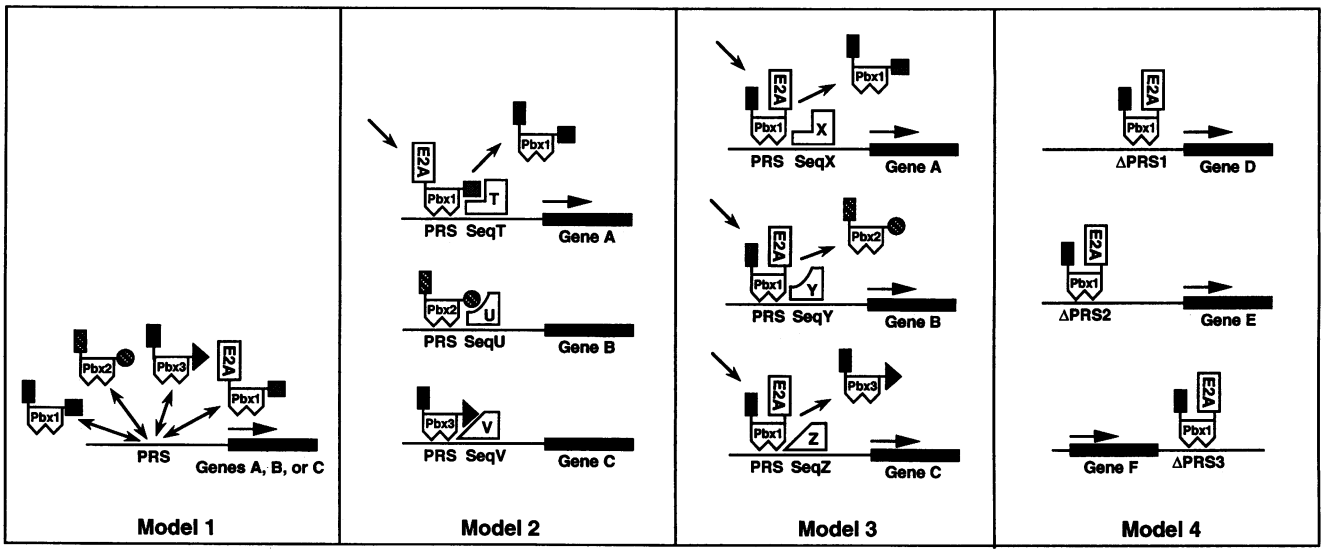

FIG. 7. Models for transactivation of Pbx-regulated genes by E2A-Pbx1. Models are explained in the text. The solid horizontal lines indicate DNA, and horizontal filled rectangles indicate specific genes. The PRS represents the sequence ATCAATCAA or a variant of this sequence to which Pbx proteins bind in vivo. The two smaller domains attached to each Pbx protein indicate unique amino- and carboxyl-terminal sequences that are hypothesized to mediate Pbx1-, Pbx2-, and Pbx3-specific functions. The domain at the right of each Pbx protein is theorized to provide target gene specificity in models 2 and 3 through its interaction with gene-specific DNA-binding proteins designated T, U, and V (model 2) and X, Y, and Z (model 3). The DNA recognition sequences bound by these associated factors are designated SeqT to SeqZ, respectively. In model 4, ΔPRS1-ΔPRS3 represent DNA sequences identical to or related to the PRS that bind E2A-Pbx1 and Pbx proteins in vivo but through which only E2A-Pbx1 exhibits a transcriptional effect.

52% identical in the sequence of their homeodomains, yet each can regulate transcription through its interaction with the sequence TCAATTAAAT (2, 15). While Ftz functions as a transcriptional activator, both En (11) and Eve (10) are transcriptional repressors. Therefore, individual Pbx proteins could have different biochemical functions and therein differentially regulate transcription of the same target genes.

While Pbx1, Pbx2, and Pbx3 may bind the same DNA motifs in vivo, they could regulate transcription of different subsets of genes through their specific association with other transcription factors. This possibility is raised by the biochemical behavior of the Pbx1-related yeast transcription factors ***a***1 and α2. The homeodomain of Pbx1 exhibits a remarkable identity of 19 of 27 and 17 of 27 amino acids with the DNA-binding helices of the yeast mating factors **a**1 and α2, respectively (19). The PRS also contains a majority of the yeast **a**1 DNA-binding motif CATCA, indicating that the sequence identity between the homeodomains of Pbx1 and yeast **a**1 may reflect a degree of functional similarity. In the diploid yeast cell, α2 cooperates with **a**1 to repress transcription of haploid-specific genes that contain unique **a**1:α2 DNA recognition sequences (22, 31). In haploid α cells, which do not express ***a***1, α2 cooperates with the MCM1 transcription factor to repress transcription of ***a***-specific genes, which contain a unique α2:MCM1 recognition sequence that is different from the **a**1:α2 recognition sequence (14, 27). Because of the adjacent binding of MCM1, the affinity of α2 for its portion of the composite DNA-binding sequence increases 50-fold (14). Cooperative binding also occurs between the transcription factors HMG1(Y) and NF-κB and between HMG1(Y) and ATF-2 and is one mechanism of controlling the selection of target genes transactivated by NF-κB and ATF-2 (3, 38). Cooperative interactions with other transcription factors might also be used by Pbx proteins to provide target gene specificity.

Considering the possibility that Pbx1, Pbx2, and Pbx3 bind the same DNA sequence in vivo, we suggest four models by which E2A-Pbx1 might cause transformation by altering transcription of cellular genes that are normally regulated by the binding of Pbx proteins to the PRS (Fig. 7). Each model is based on the hypothesis that E2A-Pbx1 coexists in the transformed cell with other transcription factors that also bind the PRS, some of which are members of the Pbx protein family. This hypothesis is supported by footprint experiments, which show that normal cellular factors protect the same region of DNA as does Pbx1 (Fig. 4), and by EMSAs, which show that a protein(s) in nuclear extracts from all cells examined is immunologically related to Pbx1 and binds specifically and with high affinity to the PRS (Fig. 5). In each model, E2A-Pbx1 competes with other Pbx proteins for the PRS, displaces a portion of them, and activates transcription of the adjacent gene.

Model 1 proposes that all normal Pbx proteins regulate transcription of a gene family (designated A to C) by their competitive binding to a common PRS. The extent to which a single Pbx protein contributes to net gene regulation would be determined by the relative abundance of that Pbx variant and its specific transcriptional activity, which could be modulated by external signals through posttranslational modifications such as phosphorylation. In this model, E2A-Pbx1 competes with all other Pbx proteins for binding to the PRS and activates transcription of all members of the gene family.

Models 2 and 3 take into account the possibility that Pbx1 cooperates with other transcription factors in regulating target gene expression. The possibility that Pbx1 cooperates with other transcription factors is raised not only by the biochemical properties of the Pbx1-related yeast transcription factors **a**1 and α2 but also by genetic data from studies of *Drosophila melanogaster*. In *D. melanogaster*, expression of the Pbx1 homolog extradenticle (Exd) is required for the appropriate expression of target genes of other homeodomain proteins, suggesting that Exd and other homeodomain proteins may

coregulate expression of single genes (33, 34). Model 2 proposes that specific Pbx proteins regulate transcription of different subsets of genes through their cooperative interactions with other DNA-binding factors, designated T, U, and V. In this model, E2A-Pbx1 retains specificity for Pbx1 target genes because it retains the Pbx1 domain that binds factor T. E2A-Pbx1 does not effectively compete with Pbx2 and Pbx3 proteins because it cannot interact with factors U and V. E2A-Pbx1 causes transformation by activating transcription of Pbx1 target genes.

Model 3 proposes that different Pbx proteins regulate transcription of different target genes through the interaction of their unique amino-terminal domains with factors X, Y, and Z. Binding of X, Y, and Z restricts the ability of specific Pbx proteins to bind the adjacent PRS but does not enhance their affinity for the PRS. In this model, fusion with E2A eliminates the amino-terminal restriction domain of Pbx1 and allows E2A-Pbx1 to bind all PRS sequences, compete with Pbx1, Pbx2, and Pbx3 for their individual sites, and activate transcription of each of their target genes, therein causing transformation.

Model 4 proposes that E2A-Pbx1 could cause transformation by activating transcription of genes D to F, which contain either the PRS or a variant of the PRS located in sequences that normally do not regulate their expression (sites designated ΔPRS1, ΔPRS2, and ΔPRS3). Transcription of genes D to F would not normally be affected by Pbx proteins either because the abundance of Pbx proteins is insufficient to substantially bind these sequences or because the binding of Pbx proteins to ΔPRS1, ΔPRS2, and ΔPRS3 does not interfere with the normal transcriptional regulation of genes D to F. Noting that a majority of the PRS (ATCAATCAA) is closely related to the common DNA motif bound by En, Ftz, and Eve (TCAATTAAAT [2]), high-level expression of E2A-Pbx1 could activate transcription of genes regulated by human homologs of *ftz* (C1, C8, HU2, HuHOX12, and HuHOX22 [35]), *en* (*en1* and *en2* [24]), or *eve* (*EVX1* and *EVX2* [6]).

Finally, it is possible that E2A-Pbx1 can alter transcription indirectly by binding and titrating out a cellular factor(s) that normally associates with either E2A or Pbx proteins and whose activity is required for normal gene regulation by E2A or Pbx proteins. Aberrant regulation of genes that are normally regulated by E2A or by Pbx proteins would result. This model for transformation by E2A-Pbx1 might be independent of the DNA-binding or transactivation functions of E2A-Pbx1. Distinguishing among these five models may be possible by cloning genes whose transcription is activated by E2A-Pbx1, identifying the DNA sequences that mediate transactivation, and determining whether one or more Pbx proteins can bind these DNA elements and regulate transcription. This approach may lead to the identification of genes whose normal transcription is regulated by Pbx1 and thereby provide a model to dissect the biochemical functions of Pbx1.

ACKNOWLEDGMENTS

We thank Toshiya Inaba, Tetsunori Funabiki, and Thomas Look (St. Jude Children's Research Hospital) for providing the oligonucleotides and protocol for selection of DNA-binding sites and for helpful discussions concerning its application to Pbx1. We also thank Mike Waterman for providing a cDNA containing the 5′ sequences of Pbx1. Q. Lu thanks Wei Wen for thoughtful and stimulating discussions.

This work was supported by Public Health Services grants CA56876-01 and 2 PO1 CA50528-04 and by a grant from the Pew Charitable Trust in the Biomedical Sciences.

ADDENDUM

While the manuscript was being reviewed, Van Dijk et al. (39) reported that a GST fusion protein containing residues 105 to 309 of Pbx1 binds ATCAATCAA in vitro and that E2A-Pbx1a but not Pbx1a functions as a transcriptional activator.

REFERENCES

1. **Dedera, D. A., E. K. Waller, D. P. LeBrun, A. Sen-Majumdar, M. E. Stevens, G. S. Barsh, and M. L. Cleary.** 1993. Chimeric homeobox gene E2A-Pbx1 induces proliferation, apoptosis, and malignant lymphomas in transgenic mice. Cell **74:**833–843.
2. **Desplan, C., J. Theis, and P. O'Farrell.** 1988. The sequence specificity of homeodomain-DNA interaction. Cell **54:**1081–1090.
3. **Du, W., D. Thanos, and T. Maniatis.** 1993. Mechanisms of transcriptional synergism between distinct virus-inducible enhancer elements. Cell **74:**887–898.
4. **Dube, I. D., S. Kamel-Reid, C. C. Yuan, M. Lu, X. Wu, G. Corpus, S. C. Raimondi, W. M. Crist, A. J. Carroll, and J. Minowada.** 1991. A novel human homeobox gene lies at the chromosome 10 breakpoint in lymphoid neoplasias with chromosomal translocation t(10;14). Blood **78:**2996–3003.
5. **Faiella, A., M. D'Esposito, M. Rambaldi, D. Acampora, S. Balsofiore, A. Stornaiuolo, A. Mallamaci, E. Migliaccio, M. Gulisano, and A. Simeone.** 1991. Isolation and mapping of EVX1, a human homeobox gene homologous to even-skipped, localized at the 5′ end of HOX1 locus on chromosome 7. Nucleic Acids Res. **19:**6541–6545.
6. **Fitzpatrick, V. D., and C. J. Ingles.** 1989. The Drosophila fushi tarazu polypeptide is a DNA-binding transcriptional activator in yeast cells. Nature (London) **337:**666–668.
7. **Florence, B., R. Handrow, and A. Laughon.** 1991. DNA-binding specificity of the fushi tarazu homeodomain. Mol. Cell. Biol. **11:**3613–3623.
8. **Gorman, C. M., L. F. Moffat, and B. H. Howard.** 1982. Recombinant genomes which express chloramphenicol acetyltransferase in mammalian cells. Mol. Cell. Biol. **2:**1044–1051.
9. **Grosschedl, R., and D. Baltimore.** 1985. Cell-type specificity of immunoglobulin gene expression is regulated by at least three DNA sequence elements. Cell **41:**885–897.
10. **Han, K., and J. L. Manley.** 1993. Transcriptional repression by the Drosophila even-skipped protein: definition of a minimal repression domain. Genes Dev. **7:**491–503.
11. **Han, K., and J. L. Manley.** 1993. Functional domains of the Drosophila Engrailed protein. EMBO J. **12:**2723–2733.
12. **Hatano, M., C. W. Roberts, M. Minden, W. M. Crist, and S. J. Korsmeyer.** 1991. Deregulation of a homeobox gene, HOX11, by the t(10;14) in T cell leukemia. Science **253:**79–82.
13. **Henthorn, P., M. Kiledjian, and T. Kadesch.** 1990. Two distinct transcription factors that bind the immunoglobulin enhancer micro E5/Kappa 2 motif. Science **247:**467–470.
14. **Herskowitz, I.** 1989. A regulatory hierarchy for cell specialization in yeast. Nature (London) **342:**749–757.
15. **Hoey, T., and M. Levine.** 1988. Divergent homeobox proteins recognize similar DNA sequences in Drosophila. Nature (London) **332:**858–861.
16. **Jones, K. A., K. R. Yamamoto, and R. Tijan.** 1985. Two distinct transcription factors bind to the HSV thymidine kinase promoter *in vitro*. Cell **42:**559–572.
17. **Kamps, M. P., and D. Baltimore.** 1993. E2A-Pbx1, the t(1;19) translocation protein of human pre-B-cell acute lymphocytic leukemia, causes acute myeloid leukemia in mice. Mol. Cell. Biol. **13:**351–357.
18. **Kamps, M. P., A. T. Look, and D. Baltimore.** 1991. The human t(1;19) translocation in pre-B ALL produces multiple nuclear E2A-Pbx1 fusion proteins with differing transforming potentials. Genes Dev. **5:**358–368.
19. **Kamps, M. P., C. Murre, X. H. Sun, and D. Baltimore.** 1990. A new homeobox gene contributes the DNA binding domain of the t(1;19) translocation protein in pre-B ALL. Cell **60:**547–555.
20. **Kennedy, M. A., R. Gonzalez-Sarmiento, U. R. Kees, F. Lampert,**

N. Dear, T. Boehm, and T. H. Tabbitts. 1991. Hox11, a homeobox-containing T-cell oncogene on human chromosome 10q24. Proc. Natl. Acad. Sci. USA **88:**8900–8904.

21. **Kissinger, C. R., B. Liu, E. Martin-Blanco, T. B. Kornberg, and C. O. Pabo.** 1990. Crystal structure of an engrailed homeodomain-DNA complex at 2.8 A resolution: a framework for understanding homeodomain-DNA interactions. Cell **63:**579–590.
22. **Klar, A. J. S., J. N. Strathern, J. R. Broach, and J. B. Hicks.** 1981. Regulation of transcription in expressed and unexpressed mating type cassettes of yeast. Nature (London) **289:**239–244.
23. **Levine, M., and T. Hoey.** 1988. Homeobox proteins as sequence-specific transcription factors. Cell **55:**537–540.
24. **Logan, C., M. Hanks, S. Noble-Topham, D. Nallainathan, N. Provart, and A. Joyner.** 1992. Cloning and sequence comparison of the mouse, human, and chicken engrailed genes reveal potential functional domains and regulatory regions. Dev. Genet. **13:**345–358.
25. **McGinnis, W., R. L. Garber, J. Wirz, A. Kuroiwa, and W. J. Gehring.** 1984. A homologous protein-coding sequence in Drosophila homeotic genes and its conservation in other metazoans. Cell **37:**403–408.
26. **McGinnis, W., M. S. Levine, E. Hafen, A. Kuroiwa, and W. J. Gehring.** 1984. A conserved DNA sequence in homeotic genes of the Drosophila Antennapedia and bithorax complexes. Nature (London) **308:**428–433.
27. **Miller, A. M., V. L. Mackay, and K. A. Nasmyth.** 1985. Identification and comparison of two sequence elements that confer cell-type specific transcription in yeast. Nature (London) **314:**598–603.
28. **Monica, K., N. Galili, J. Nourse, D. Saltman, and M. L. Cleary.** 1991. *PBX2* and *PBX3*, new homeobox genes with extensive homology to the human proto-oncogene *PBX1*. Mol. Cell. Biol. **11:**6149–6157.
29. **Murre, C., P. McCaw, and D. Baltimore.** 1989. A new DNA binding and dimerization motif in immunoglobulin enhancer binding, daughterless, MyoD, and Myc proteins. Cell **56:**777–783.
30. **Murre, C., P. McCaw, H. Vaessin, M. Caudy, L. Jan, Y. Jan, C. Cabrera, J. Buskin, S. Hauschka, A. Lassar, H. Weintraub, and D. Baltimore.** 1989. Interactions between heterologous helix-loop-helix proteins generate complexes that bind specifically to a common DNA sequence. Cell **58:**537–544.
31. **Nasmyth, K. A., K. Tatchell, B. D. Hall, C. Astell, and M. Smith.** 1981. Physical analysis of mating-type loci in S. cerevisiae. Nature (London) **289:**244–250.
32. **Nourse, J., J. D. Mellentin, N. Galili, J. Walkinson, E. Stanbridge, S. D. Smith, and M. L. Cleary.** 1990. Chromosomal translocation t(1;19) results in synthesis of a homeobox fusion mRNA that codes for a potential chimeric transcription factor. Cell **60:**535–545.
33. **Peifer, M., and E. Wieschaus.** 1990. Mutations in the Drosophila gene extradenticle affect the way specific homeo domain proteins regulate segmental identity. Genes Dev. **4:**1209–1223.
34. **Rauskolb, C., M. Peifer, and E. Wieschaus.** 1993. Extradenticle, a regulator of homeotic gene activity, is a homologue of the homeobox-containing human proto-oncogene pbx1. Cell **74:**1101–1112.
35. **Scott, M. P., J. W. Tamkun, and G. W. Hartzel III.** 1989. The structure and function of the homeodomain. Biochim. Biophys. Acta **989:**25–48.
36. **Scott, M. P., and A. J. Weiner.** 1984. Structural relationships among genes that control development: sequence homology between the Antennapedia, Ultrabithorax, and fushi tarazu loci of Drosophila. Proc. Natl. Acad. Sci. USA **81:**4115–4119.
37. **Smith, D. B., and K. S. Johnson.** 1988. Single-step purification of polypeptides expressed in Escherichia coli as fusions with glutathione S-transferase. Gene **67:**31–40.
38. **Thanos, D., and T. Maniatis.** 1992. The high mobility group protein HMG 1(Y) is required for NF-κB-dependent virus induction of the human IFN-B gene. Cell **71:**777–789.
39. **Van Dijk, M., P. Voorhoeve, and C. Murre.** 1993. Pbx1 is converted into a transcriptional activator upon acquiring the N-terminal region of E2A in pre-B-cell acute lymphoblastoid leukemia. Proc. Natl. Acad. Sci. USA **90:**6061–6065.
40. **Wigler, M., R. Sweet, G. K. Kim, B. Wold, A. Pellicer, E. Lacy, T. Maniatis, S. Silverstein, and R. Axel.** 1979. Transformation of mammalian cells with genes from prokaryotes and eukaryotes. Cell **16:**777–785.
41. **Wolberger, C., A. Vershon, B. Liu, A. Johnson, and C. Pabo.** 1991. Crystal structure of a MATα2 homeodomain-operator complex suggests a general model for homeodomain-DNA interactions. Cell **67:**517–528.
42. **Wright, C. V. E., K. W. Y. Cho, G. Oliver, and E. M. DeRobertis.** 1989. Vertebrate homeodomain proteins: families of region-specific transcription factors. Trends Biochem. Sci. **14:**52–56.