A human Hox 1 homeobox gene exhibits myeloid-specific expression of alternative transcripts in human hematopoietic cells

P.Lowney+ J.Corral, K.Detmer, M.M.LeBeau[1], L.Deaven[2], H.J.Lawrence and C.Largman*
Department of Internal Medicine, University of California, Davis, CA 95616 and Department of Medicine, Veterans Affairs Medical Center, Martinez, CA 94553, [1]Section of Hematology/Oncology, The University of Chicago, Chicago, IL 60637 and [2]Life Sciences Division, Los Alamos National Laboratory, Los Alamos, NM 97545, USA

Received November 15, 1990; Revised and Accepted February 4, 1991

ABSTRACT

As part of a survey of the expression of homeobox-containing genes in human hematopoietic cells, we identified a novel gene (PL1) expressed only in cells of the myelomonocytic lineage (Shen *et al., Proc. Natl. Acad. Sci, USA* 86, 8536, 1989). On Northern gel analysis, major transcripts of 3.0 and 2.2 kb length are observed. Alternatively spliced homeobox-containing cDNAs, corresponding to the major transcripts, have been cloned from two myeloid leukemia cell libraries. The two cDNAs share the homeodomain and 3′ flanking region but have unique 5′ flanking regions. The longer transcript, would encode a 496 amino acid homeobox-containing protein, while the shorter message would encode a 94 amino acid homeobox-containing protein lacking the extended amino-terminal region. These two transcripts are differentially expressed in human leukemia cell lines. The larger transcript is exclusively expressed in cells with myelomonocytic features, while the smaller transcript is expressed in a variety of hematopoietic cell types. PL mRNA is also detectable in normal human bone marrow by RNAse protection. Neither transcript is expressed in uninduced teratocarcinoma cells or in the adult human tissues surveyed. The homeodomain is identical to the genomic sequence for Hox 1H, a newly identified member of the Hox 1 locus (Acampora *et al. Nucl. Acids Res.* 17, 10385, 1989). The PL1 gene was localized to chromosome 7 using chromosome specific blots and sublocalized to region pl4 – 21 by *in situ* hybridization of chromosomal spreads, confirming its location within the Hox 1 complex.

INTRODUCTION

Proteins containing the homeobox, a DNA binding domain, play a role in the transcriptional regulation of mammalian cell-specific genes in somatrophic cells (1, 2), B cells (3) and liver (4). Homeoproteins have also been been suggested to function as transcriptional activators or repressors in the determination of Drosophila embryological and neuronal development (5–7). Since the process of the differentiation of pluripotent blood stem cells to committed daughter cells has parallels to embryogenesis, we have evaluated the possible role of homeobox-containing genes in hematopoiesis. We have demonstrated the restricted expression of certain homeobox genes in the HOX 2 locus in pluripotent cell lines with erythroid features and the expression of a novel homeobox containing cDNA, previously referred to as PL1, in myeloid cell lines (8). Similar lineage restricted expression of murine HOX genes in murine leukemic cell lines has been reported by Kongsuwan and colleagues (9).

We have previously reported that the PL gene is expressed as two major transcripts of 3.0 and 2.2 kb in a limited of number of myeloid cell lines (8). The current work sought to characterize the structure of these transcripts and to further evaluate the specificity of their expression. Here we report the cloning and differential expression of two cDNAs, PL1 and PL2, which arise from the alternate splicing of two 5′ sequences to a common homeobox-containing 3′ region, and the chromosomal localization of the PL gene.

MATERIALS AND METHODS

Cell lines and culture

CESS (ATCC), MOLT (ATCC), human bladder GIBBS (Dr. B.Edwards, Dept. of Surgery, VAMC, Martinez, CA), HEL (Dr. T.Papayannopoulou, Dept. of Medicine, University of Washington School of Medicine, Seattle, WA), PLB985 (Dr. T.Rado, Dept. of Medicine, University of Alabama, Birmingham, AL) and U937 (Dr. E.Campbell, Washington University, St. Louis, MO) cell lines were maintained in RPMI media supplemented with 10% fetal calf serum (FCS, HyClone Labs, Logan, UT). KG1 (ATCC) cells were maintained in

*To whom correspondence should be addressed at Research 151, Martinez VA Medical Center, 150 Muir Road, Martinez, CA 94553, USA

+Present address: Department of Foods and Nutrition, Purdue University, West Lafayette, IN 47907, USA

Iscove's modified Dulbecco medium (IMDM) supplemented with 20% FCS. ML1, ML2 and ML3 cells (Dr. J.C.Pauly, Reswell Park Memorial Institute, Buffalo, NY) were maintained in RPMI media supplemented with 10% heat inactivated FCS. Neuroblastoma SKNSH, and human teratocarcinoma PA (Dr. P.Gumerlock, Dept. of Internal Medicine, University of California, Davis, CA) were maintained in Dulbecco's Modified Eagle's Media supplemented with 10% FCS. Cells were resuspended in fresh medium at a density of $0.2-0.3\times10^6$ cells/ml every third or fourth day and incubated at 37°C in 5% CO_2 atmosphere. Messenger RNA prepared from KU812 cells (10) was kindly provided by Dr. T.Papayannopoulou.

RNA isolation and Northern gel analyses

Total RNA was extracted from cells or tissue by the guanidinium thiocyanate method of Chirgwin and coworkers (11). Poly $(A)^+$ RNA was separated from poly $(A)^-$ by oligo (dT) cellulose chromatography. RNA (5 μg) was electrophoresed on 1.5% agarose–formaldehyde gels, transferred to nylon, cross-linked by UV irradiation for 2 min and hybridized in 50% formamide, 5×Denhardt's solution, 0.5% SDS, 1 M sodium chloride, 1 μg/ml single-stranded salmon sperm DNA, 50 mM Tris, pH 8.0 and 5% dextran sulfate. RNA ladders (BRL, Gaithersburg, MD) were run in a separate lane which was cut off prior to transfer and stained with ethidium bromide to estimate transcript sizes. Filters were hybridized with random ^{32}P CTP labelled DNA probes at 55°C overnight. After hybridization, filters were washed: 3 times for 10 minutes in 2×SSC, 0.1% SDS, 25°C; twice for 30 minutes in 2×SSC, 1.0% SDS, 55°C; twice for 30 minutes in 0.1×SSC, 0.1% SDS, 55°C. Filters were air dried and exposed to Kodak XAR film at −70°C for 4 h to 3 d. Filters were stripped in 0.1×SSC 0.1% SDS at 95°C and re-hybridized with a 2.2 kb human β-actin probe.

RNAse protection analysis

DNA fragments C and D, shown in Figure 1, were subcloned into Bluescript vectors and RNA probes synthesized with ^{32}P UTP as described (12). Probes were purified by electrophoresis in 4% polyacrylamide/urea gels, hybridized to poly $(A)^+$ RNA at 65°C, digested with RNAse A and T1, and the protected bands electrophoresed in 5% polyacrylamide/urea gels as described (12). Each protection experiment was repeated with at least two separate preparations of cytoplasmic mRNA from U937 cells, and from normal bone marrow.

cDNA library preparation and screening

A U937 cDNA library was synthesized using AMV Reverse Transcriptase (Life Sciences) for first strand synthesis and *E.coli* DNA polymerase and RNase H for second strand synthesis. Double stranded cDNA was end-filled, methylated with DNA methylase and ligated to EcoRI linkers. Complementary DNA was cloned into the EcoRI site of λgt11; titration indicated that the library contained $\sim5\times10^5$ independent clones. The ML3 cDNA library was synthesized using Moloney Murine Leukemia Virus Reverse Transcriptase (BRL, Gaithersburg, MD) for first strand synthesis and *E.coli* DNA polymerase and RNase H for second strand synthesis. Double stranded cDNA was end-filled and ligated to EcoRI linkers. Complementary DNAs greater than 350 bp were cloned into the EcoRI site of λZAPII (Stratagene, La Jolla, CA); titration indicated that the library contained 4×10^5 independent clones. The human genomic library in Charon 4A was the generous gift of Dr. C.Craik, Dept. of Pharmaceutical Chemistry, University of California, San Francisco, CA. PL1 is a partial cDNA clone initially isolated from the U937 library during a moderate stringency screen with a Hox 2.3 homeobox probe (8). The ML3 and U937 libraries were screened with a random primer labeled PL1 probe under stringent conditions (50% formamide, 5×SSC, 5×Denhardt's solution, 0.01 M Tris, pH 6.5; 1 μg/ml single-stranded salmon sperm DNA, 42°C) to isolate additional PL1-related clones.

Southern blot analysis

High molecular weight DNA was isolated from human placenta (13). DNA digestions were performed with commercially available restriction enzymes. Digested DNA was run on 0.8% agarose gels and transferred to nitrocellulose (13). Southern blot analysis was performed using a single-stranded MspI–MspI cDNA probe containing the 3′ region of PL (Figure 2; bases 1456–1936).

DNA Sequencing

Lambda gt11 cDNA and Charon 4A genomic clones were subcloned into m13 and sequenced in both orientations using the dideoxy-termination method according to manufacturer's instructions (Sequenase, US Biochemicals, Cleveland, OH). Bluescript clones obtained from the ML3 library were screened using double stranded sequencing according to suppliers instructions (Taq Track, Promega, Madison, WI) or subcloned into m13 for single stranded sequencing. Sequencing primers were synthesized by the UC Davis Protein Structure Laboratory or on a PCR-mate model 391 (Applied Biosystems, Foster City, CA).

In vitro transcription and translation

The longest PL1 cDNA was converted to mRNA using the T7 promoter system in a Bluescript plasmid (Stratagene, La Jolla, CA). RNA was translated in a reticulocyte lysate system (Stratagene, La Jolla, CA) in the presence of ^{35}S-methionine. The products were subjected to SDS-gel electrophoresis on a 12.5% polyacrylamide gel and detected by autoradiography. Low and high molecular weight markers (Amersham) run simultaneously were used to estimate the translation product size.

Chromosomal identification and sublocalization

Human chromosomes were isolated from several diploid or haploid cell lines and flow sorted (14, 15) onto 24 spots on nitrocellulose filters. Each spot, representing an individual human chromosome, contained 3×10^4 copies of that chromosome. In some cases, human chromosomes 9–12 were sorted together into spots containing 1.2×10^5 chromosomes. These spot blots were hybridized with ^{32}P-labeled probes corresponding to the 3′ flanking region of PL (Figure 1, Probe E).

In situ chromosomal hybridization

Human metaphase cells were prepared from phytohemaglutinin-stimulated peripheral blood lymphocytes. Radiolabeled PL2 cDNA and genomic (HG-19) probes were prepared by nick translation of the entire plasmid with all four ^3H-labeled deoxynucleoside triphosphates to a specific activity of $7.0\times10^7-1.0\times10^8$ dpm/μg. *In situ* hybridization was performed as described previously (16). Metaphase cells were hybridized at 0.5, 1.0, 5.0, and 20 ng of probe per ml of hybridization mixture. Autroradiographs were exposed for 11 days. All hybridizations were repeated three times and gave similar results.

RESULTS

Isolation of PL cDNAs

Preliminary screening of human leukemia cell lines suggested that the myeloid line ML3 expresses the 3.0 kb transcript as the predominant message (see below). To obtain clones corresponding to this transcript, ML3 poly $(A)^+$ RNA was used to construct a cDNA library in λZAPII. Five independent clones corresponding to the 3.0 kb transcript (see below) were obtained by screening the library with a partial homeobox containing a cDNA clone (PL1) (8). The clones were shown to be co-linear by restriction mapping and partial sequence determination. The longest clone, which was renamed PL1 (Figure 1), was sequenced and shown to contain a long open reading frame which began at the 5′ end of the clone and terminated immediately 3′ to the homeobox (Figure 2). This open frame encodes a putative 390 amino acid protein when the first in-frame methionine (residue 107, Figure 2) is used for translation initiation. This clone has been transcribed and translated *in vitro* to produce an approximately 40 kD protein closely corresponding to the 43.3 kD calculated size (results not shown).

Because the reading frame of the PL1 cDNA was open to the 5′end of the clone and there was a discrepancy between the length of PL1 (2.7 kb) and observed transcript size seen by Northern analysis, it appeared that the PL1 cDNA was not a full length clone and did not contain the actual starting methionine. To isolate a genomic subclone of the 5′ region of PL1, a human genomic library in Charon 4A was screened with a probe specific to the PL1 5′ flanking region (Figure 1, probe A). A 1.5 kb EcoRI genomic clone containing the PL 5′ flanking region was subcloned and sequenced. As shown in Figure 2, the reading frame remains open for an additional 237 bp 5′ to the cDNA clone to a methionine which is preceded by numerous stop codons in all three reading frames. The conceptual translation of the protein encoded by the combined genomic and cDNA sequences suggest that a full-length PL1 transcript would generate a 496 amino acid, 54.7 kD protein. The full-length PL1 protein does not possess regions enriched in glutamine, serine/threonine, or acidic amino acids, regions associated with transcriptional activation (17−19). There are however, regions enriched in preline (residues 215−238) and glycine (residues 181−197), amino acids which disrupt alpha helical structure. In addition to the large concentration of basic residues in the homeodomain, the remainder of the protein is enriched in arginine (12.4%) and there is a second particularly rich basic region (residues 127−147) which might function as a DNA binding moiety.

The 5′ flanking region of the PL1 clone (Figure 1, probe A) did not hybridize to the major 2.2 kb transcript observed in U937 cells (below), suggesting that this band represented an alternatively spliced transcript. To obtain full length clones corresponding to this message, a U937 cDNA library constructed in λgt11 was screened using a clone containing the PL homeobox. Three unique co-linear clones were obtained which appeared to differ in structure from PL1 based on restriction mapping. The longest cDNA (PL2) was sequenced (Figure 2) and shown to contain the PL homeobox. However, PL2 diverges from PL1 at a consensus splice acceptor and possesses a unique 5′ flanking region (Figure 1). Utilization of the first methionine in the reading frame which contains the homeobox predicts a 94 amino acid protein consisting predominantly of the homeodomain DNA binding region (Figure 2). There are only 14 amino acids 5′ to the homeodomain, three of which are unique to PL2 and the remainder common to PL1 and PL2.

In order to clarify the relationship between the PL transcripts, a series of overlapping Charon 4A PL genomic subclones were isolated. Restriction mapping and partial sequence analysis suggests that there are two introns between the homeobox and the start of the PL2 cDNA clone and that another intron of undefined size separates the PL1 specific exon from the first PL2 exon (Figures 1 and 2). Each intron-exon boundary as well as regions of nucleotide ambiguity in the cDNA clones were confirmed by sequencing. Thus, PL1 and PL2 represent

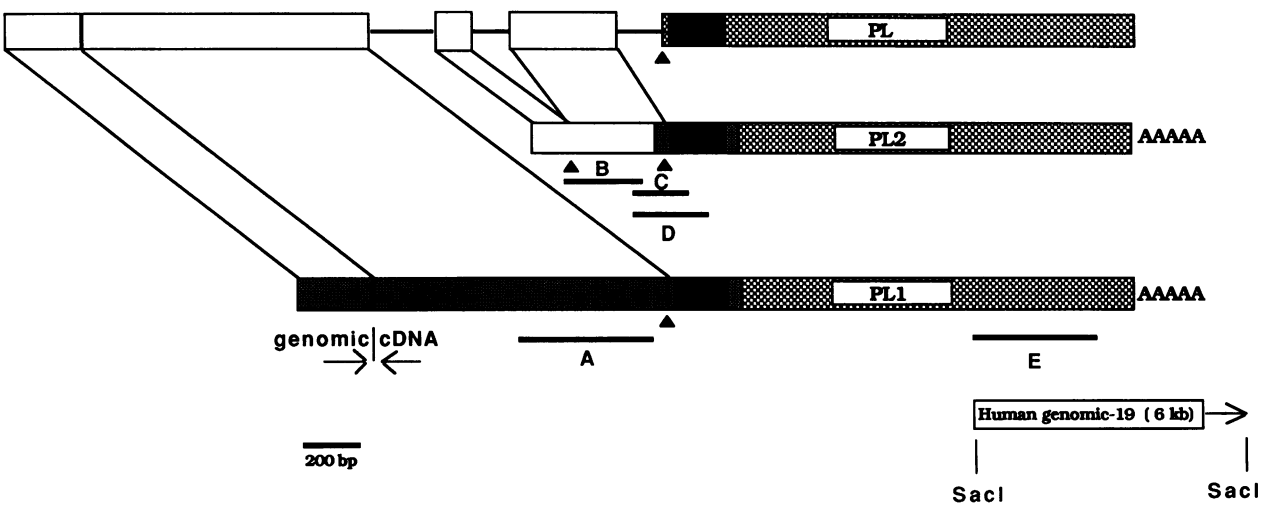

Figure 1. PL genomic and cDNA structures. The upper line represents the PL genomic structure with boxes indicating exons and thin lines representing introns of unspecified sizes. PL1 and PL2 represent the longest cDNA clones containing the intron splicing structures shown. An additional genomic fragment is appended to the 5′ side of the PL1 cDNA to indicate the structure of the proposed full-length PL1 transcript. The triangles represent splice junctions for which the genomic boundaries have been determined by sequence analysis. Black boxes represent the homeobox domain, while grey areas represent protein coding sequences. The cross-hatched boxes indicate that both cDNAs are co-linear with the gene in these areas. The human genomic 19 clone is a 6 kb SacI/SacI fragment which was used for in situ chromosomal hybridization. Restriction fragments depicted by A, B, and E represent specific probes for PL1, PL2, or a common PL probe, respectively, while fragments C and D represent probes used for RNAse protection experiments.

alternative transcripts of the PL gene which would code for two proteins, both of which would contain the homeodomain DNA binding region but possess either an extended N-terminal region (PL1) or a truncated N-terminal region (PL2).

Correspondence of PL cDNA clones to myeloid cell transcripts

To confirm the identity of the two PL transcripts observed in myeloid cells, probes corresponding to unique 5′ regions of PL1 and PL2 and the common 3′ flanking region were used to evaluate expression in U937 and ML3 cell lines. The 3′ flanking region probe hybridizes to both the 3.0 kb and a 2.2 kb transcripts in U937 and ML3 cells (Figure 3, lanes 1, 3 and 4). The PL1 specific probe binds only to the upper 3.0 kb transcript of both cell lines (Figure 3, lanes 5, 6) while the PL2 specific probe binds only the 2.2 kb transcript (Figure 3, lane 2). Figure 3 also illustrates the relative expression of the two transcripts in U937 and ML3 cell lines. Using the common 3′ probe, U937 cells are enriched for the 2.2 kb PL2 transcript (lane 1) while in contrast, ML3 cells are enriched for PL1 message (lane 4). It should be noted that the 2.2 kb band seen in U937 cells (lane 1) is broad and may contain additional uncharacterized transcripts. In addition, the 5′ PL1 specific flanking region probe appears to hybridize weakly to other transcripts (lane 5).

Myeloid specific expression of PL

We have previously reported that the expression of PL occurs predominantly in cells of the human myeloid lineage. However, PL2 is expressed in an Epstein-Barr virus infected B cell line, CESS (Figure 4A, lane 1) and a T cell line, MOLT (data not shown). Two cell lines with mixed erythroid/myeloid potential, HEL and KU812 (lanes 2, 3) also express PL2. In contrast, the PL1 transcript appears to be expressed exclusively in cells with a predominant myeloid phenotype (Figure 4). In these myeloid cell lines, both PL1 and PL2 transcripts are expressed. U937 and KG1 cells are relatively enriched for the PL2 transcript (lanes

```
      1                                                  20
PL1   MetProGlyProProProAlaThrLeuGlySerProHisSerSerProPheGlyProArg
  1   ATGCCAGGCCCCCCACCAGCCACGTTGGGCAGCCCCCACAGCTCCCCGTTCGGGCCAAGG
                                                         40
      CysArgGlyAlaSerProGlyProSerIleGlnIleThrTyrLeuTyrGlnSerArgAla
 61   TGTCGGGGTGCGTCTCCTGGCCCATCAATACAGATTACATATTTATATCAATCGCGGGCT
                                                         60
      LeuArgAlaProSerGluSerGlyProArgAlaTyrGluThrLysLeuGlyValValAla
121   CTGAGGGCGCCCTCGGAGAGCGGGCCCCGCGCCTACGAAACCAAACTGGGAGTGGTCGCG
                                                        *80
      ArgLysLeuTrpLeuGlyIleGlySerGlyArgArgProValArgGlyAspCysTyrArg
181   CGGAAACTCTGGCTCGGGATTGGCTCGGGGCGCCGGCCGGTGCGGGGGGATTGCTATCGT
                                                        100
      IleGlnHisValLeuHisLysLysCysGlnProGluGlyLeuSerAlaProPheAlaLys
241   ATTCAGCATGTTTTGCACAAGAAATGTCAGCCAGAAGGGCTATCTGCTCCCTTCGCCAAA
                ###                                     120
      LeuSerHisAsnAsnValMetLeuGlyGluProArgArgGluLeuPhePheGlyArgLeu
301   TTATCCCACAACAATGTCATGCTCGGAGAGCCCCGCCGCGAACTCTTTTTTGGTCGACTC
                                                        140
      AlaHisGlnLeuGlyGlnArgArgGlyArgArgArgTrpTrpTrpArgGlyGlyArgArg
361   GCTCATCAGCTCGGGCAGAGGCGAGGCAGGCGGCGGTGGTGGTGGCGCGCGGGGCGCCGCG
                                                        160
      ArgTrpArgLeuHisAlaHisAlaGlySerThrCysArgProProProThrCysProThr
421   CGGTGGCGGTTACACGCCCACGCGGGGTCTACCTGCCGCCCGCCGCCGACCTGCCCTACC
                                                        180
      TrpGlnSerCysGlyLeuPheProThrLeuGlyArgLysArgAsnGluAlaAlaSerPro
481   TGGCAGAGCTGCGGGCTCTTCCCCACGCTGGGGCGCAAGCGCAATGAGGCAGCGTCGCCG
                                                        200
      GlySerGlyGlyGlyGlyGlyGlyLeuGlyProGlyAlaHisGlyTyrGlyProSerPro
541   GGCAGCGGTGGCGGTGGCGGGGGTCTAGGTCCCGGGGCGCACGGCTACGGGCCCTCGCCC
                                                        220
      IleAspLeuTrpLeuAspAlaProArgSerCysArgMetGluProProAspGlyProPro
601   ATAGACCTGTGGCTAGACGCGCCCCGGTCTTGCCGGATGGAGCCGCCTGACGGGCCGCCG
                                                        240
      ProProProGlnGlnGlnProProProProGlnProProGlnProAlaProHisArg
661   CCGCCGCCCCCAGCAGCAGCCGCCGCCCCCGCCGCAACCACCCCAGCCAGCGCCGCACCGG
                                                        260
      ThrSerCysSerPheAlaGlnAsnIleLysGluGluSerSerTyrCysLeuTyrAspSer
721   ACCTCGTGCTCTTTCGCGCAGAACATCAAAGAAGAGAGCTCCTACTGCCTCTACGACTCG
                                                        280
      AlaAspLysCysProLysValSerAlaThrAlaProAsnTrpLeuProSerAlaGlyPro
781   GCGGACAAATGCCCCAAAGTCTCGGCCACCGCGCCCGAACTGGCTCCCTTCCGCGGGCCCG
                                                        300
      ProProAspGlyAlaAlaProTrpAlaProProAlaGlyCysSerAlaTrpLeuLeuArg
841   CCGCCCGACGGCGCTGCGCCCTGGGCACCTCCAGCGGGGTGCAGTGCCTGGCTACTTCGC
                                                        320
      LeuLeuArgProThrArgProProArgAlaMetAlaAlaGlyGlyCysTrpArgAlaArg
901   CTTCTCAGGCCTACGCGACCGCCAAGGGCTATGGCAGCTGCGCGCTGCTGGCGCGCACGA
                                                        340
      ThrArgGlyTrpProSerProArgSerProArgGlyGlyGlyPheAspLeuProProAla
961   ACTCGGGGCTGGCCGTCCCCGCCGCAGCCCCCGGGGCCGCCGTTTCGATCTCCCGCCCGCG
                                                        360
      LeuAlaSerGlySerAlaArgMetArgProGlyArgSerGluProSerIleArgArgArg
1021  CTAGCCTCCGGCTCGGCTCGGATGCGGCCCGGAAGGAGCGAGCCCTCGATTCGCCGCCGC
```

```
                                                        380
      ProProArgTrpLeuAlaAlaAlaAlaGlyAlaArgArgAlaThrArgArgAlaThrArg
1081  CCCCCACGCTGGCTTGCGGCAGCGGCGGGGGCTCGCAGGGCGACGAGGAGGGCCACGCGT
                                                        400
      ArgSerSerAlaAlaGluGluLeuSerProAlaProSerGluSerSerLysAlaSerPro
1141  CGTTCCTCGGCCGCGGAGGAGCTCTCCCCGGCCCCTTCCGAGAGCAGCAAAGCCTCGCCG
                 ***                                    420
      GluLysAspSerLeuGlyAsnSerLysGlyGluAsnAlaAlaAsnTrpLeuThrAlaLys
1201  GAGAAGGATTCCCTGGGCAATTCCAAAGGTGAAAACGCAGCCAACTGGCTCACGGCAAAG
                                                        440
      SerGlyArgLysLysArgCysProTyrThrLysHisGlnThrLeuGluLeuGluLysGlu
1261  AGTGGTCGGAAGAAGCGCTGCCCCTACACGAAGCACCAGACACTGGAGCTGGAGAAGGAG
                                                        460
      PheLeuPheAsnMetTyrLeuThrArgGluArgArgLeuGluIleSerArgSerValHis
1321  TTTCTGTTCAATATGTACCTTACTCGAGAGCGGCGCCTAGAGATTAGCCGCAGCGTCCAC
                                                        480
      LeuThrAspArgGlnValLysIleTrpPheGlnAsnArgArgMetLysLeuLysLysMet
1381  CTCACGGACAGACAAGTGAAAATCTGGTTTCAGAACCGCAGGATGAAACTGAAGAAAATG
                                                  496
      AsnArgGluAsnArgIleArgGluLeuThrAlaAsnPheAsnPheSerTRM
1441  AATCGAGAAAACCGGATCCGGGAGCTCACAGCCAACTTTAATTTTTCCTGATGAATCTCC
1501  AGGCGACGCGGTTTTTTCACTTCCCGAGCGCTGGTCCCCTCCCTCTGTCTTCAGGCTCTG
1561  CCAGGAACTCCGCACCTGTGCTGGAGCCCTGTTCCTCCCTCCCACACTCGCCATCTCCTGG
1621  GCCGTTACATCTGTGCAGGGCTGGTTTGTTCTGACTTTTTGTTTCTTTGTGTTTGCTTGG
1681  TGCTGGTTTATTTGTTGTTTTCTGGGGGAAAAAGCCATATCATGCTAAAATTCTATAGAG
1741  ATAGATATTGTCCTAAGTGTCAAGTCCTGACTGGGCTGGGTTTGCTGTCTTGGGGTCCCA
1801  CTGCTCGAAATGGCCCCTGTCTTCGGCCGAGCTGGTTTCCTGCCCAGCCTGGGCGAAACC
1861  TAGCCGAAGGCCGAGGTCCCATTGTTGGCGCTGAGGTGTCTGGCCTGAGGTCAATGGTGC
1921  AAAGGAGCCGCCACCGGGCATGTCTGCCTGGAGTGCTGTGCTGTGTTTAATCAGGGGATA
1981  CAGGCCCCTGGGTTTCTTTTTTCTTTCTTCCTTTCTTCCTTGGCCAAGAGAAGGGCTTAC
2041  AGGCATGGACATGCAGGTTGGCAAACGGGCTTGACTTTGGCTGATTTAAAAAGTGAGAAA
2101  GAAAGTAAAAAAGGTTAATTTTTCCTTCCTCTGTAAGATATCCCAGCTTTAAAAAGAAAA
2161  AAAAAAAGAATTACCAAGAGAAGGGGACTTCTCTTCCAGTTTCTGTAAGGTCTTACATTG
2221  CCTGACTAAAATGTTTCATTTACCTCTAAATTTCCATATCCTTCTGGCTGTAGATAAATA
2281  ATGTAGTTTTGTTTATGCATTTGGAATTAGTGGATTTTTTTGTCATTAAAATTGTTACCA
2341  CTGGTAACATGTGACAAGCACACCACAATTCTCCCTATCTTGTGAAGTTGTTTTTTTAAA
2401  TCGCCTTGAACAAAAAGTTTTTTTTTTTGTTTGTTTTTGCTTTCTGAAATTCACAGAACC
2461  CTAGGAGGACTGGGGTAAGCGGAATAAACTAGAGAAGGGAGACATTGTTTGGATTTCCTT
2521  CCTATAAATACAAATCTGTATAAATGTCTATTATTATGAAGAATTGCCAATCTTGTTTTA
2581  AGCAAATGCATTCTATCGTTATTATAAATGTTAGTTCTAGCTCTATTTACTTCTAATCTT
2641  AAATCAGAATAAATTAATATTGTATTGCTGCTGTGCGTGGAAAAAGACGATGTTTATGTT
2701  CTTATAGAATAAAAGCTGTGGAATGAAAAAAAAAAAAAAAAAAAA

PL2: 1  GTAGGGGTGGAGACCTAATTGGGCTGATTTGCCTGCTGCTGCTAGGAGGAGGCCTAGTAG
    61  TGGGGTGAGGCTTGGATTAGCGTTTAGAAGGGCTATTTGTTGTGGGTCTCATGATTTATC
   121  CTCGGAATTCCGGGAGAGGATGTGCCCAGAGCTGGTTTCGCTCTCTCGGCTCGGGCTGG
   181  AACTCGGGCCCAACCTAGGCCCCCACCGCGACGACATCGCGCACTTCCGATCAATGTCAA
   241  AGCCGCCGGGGAGCCGGGAACCCCAGCATGATTCTTGGCCTTTGTTCGCTTCTGATACTA
   301  AGAGCAGCACGGTACATTATTTCACTTGTCCCGCTCCCCTTCATAACAGAAAAAGGGGAC
   361  TCACCCTCAAGAAGTGATTGGTATGGTAATTTAAAGCAACGCGCATTCGCTAGGCCTCGC
                                                  1      ***
                                                MetCysGlnGly...
   421  GAGCGTCGCCGCGCGGAGAAGCCAGCTGTCCCTTGGCAGTGATTTCGGAAATGTGTCAAGGC...
```

Figure 2. Nucleotide and deduced amino acid sequences of the PL cDNAs and 5′ genomic region. The PL1 cDNA extends from nucleotide 238 (*) to the poly A tail. Met-107 (# # #) is apparently used during *in vitro* translation of this cDNA to synthesize a 43,300 dalton protein (see results). The open reading frame in the 5′ direction extends from Met-107 through the start of the cDNA clone and for an additional 237 bp of upstream genomic sequence to Met-1. The PL2 cDNA is shown below the PL1 sequence and extends from nucleotide 1 through nucleotide 480 of unique sequence followed by a region common to PL1 which contains the homeodomain, 3′ untranslated region, and poly A tail. Gly-3 of PL2 (***) denotes the start of the common nucleotide and amino acid sequence shared with Gly-406 (***) of PL1. Splice sites in the PL cDNAs are underlined. The common homeodomain is boxed, while three possible polyadenylation signals are overlined.

4, 5) while ML1, ML2, ML3 and PLB985 cells are relatively enriched for PL1 transcript (lanes 6–9). Blood cell specificity of PL expression is also suggested by the failure to detect PL transcripts in Northern gel analysis of human neuroblastoma, bladder carcinoma, uninduced teratocarcinoma, and lung carcinoma cell lines or in normal human testis (Figure 4B).

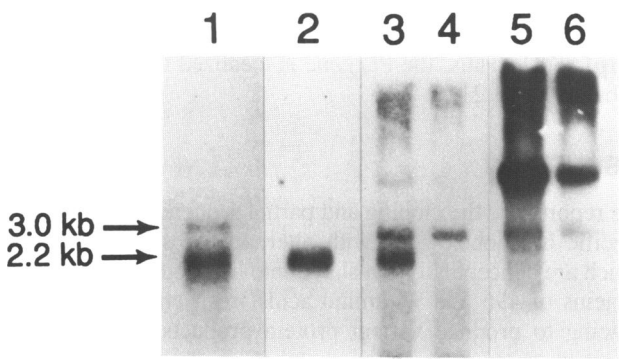

Figure 3. Correspondence of PL1 and PL2 cDNAs in myeloid cell lines. Expression of PL homeobox-containing transcripts was detected using a common 3′ flanking region probe, (Figure 1, probe E), in U937 poly A^+ RNA (lanes 1 and 3) and ML3 poly $(A)^+$ RNA (lane 4). The lower 2.2 kb transcript is visible in lane 4 upon extended autoradiogaph exposure (data not shown). Re-probing the blot of lane 1 with a PL2 5′ specific probe (Figure 1, probe B) demonstrates that the 2.2 kb message corresponds to PL2 (lane 2). Re-probing the blot of lanes 3 and 4 with a PL1 specific 5′ flanking region probe (Fiigure 1, probe A) demonstrates that the 3.0 kb transcript corresponds to PL1 (lanes 5 and 6). The heavy band in lanes 5 and 6 at approximately 4.4 kb co-migrates with residual ribosomal RNA remaining in the sample and probably represents non-specific hybridization of this material with the probe. To confirm this possibility, a sample of poly $(A)^-$ RNA from ML3 cells was run in a parallel Northern gel experiment. Probe A was shown to hybridize under stringent conditions to a 4.4 kb band in the poly $(A)^-$ RNA (results not shown).

Expression of PL transcripts in normal human bone marrow

RNAse protection of mRNA from bone marrow or U937 cells was performed with two probes derived from PL2 which span the splice junction (Figure 1, C and D). As shown in Figure 5 lanes 1 and 2, mRNA from U937 cells protects two bands with both probes. In each case, the top band represents full protection of the probe afforded by transcripts corresponding to PL2, while the lower band represents protection of the 3′ common region of the probe and thus is attributed to the presence of PL1 transcripts. RNA isolated from normal human bone marrow protects only the lower band of message, suggesting that the PL2 transcript is quite minor in normal marrow (Figure 5, lane 3).

Chromosomal localization and structure

The PL gene is present in the mammalian genome as a single copy gene as indicated by Southern analysis of enzymatic digests of human genomic DNA probed with a PL homeobox containing cDNA probe (Figure 6). Hybridization to chromosome specific blots with a probe corresponding to the 3′ flanking region common to PL1 and PL2 (probe E, Figure 1) placed the PL gene on chromosome 7.

To sublocalize the PL gene, cDNA and genomic probes were hybridized to normal metaphase chromosomes, resulting in specific labeling only of chromosome 7. Of 100 metaphase cells examined from the hybridization of the HG-19 genomic probe, 33 (33%) were labeled on regions p1 or p2 of one or both chromosome 7 homologues. The distribution of labeled sites on this chromosome is illustrated in Figure 7; of 227 labeled sites observed, 63 (27.8%) were located on this chromosome. These sites were clustered at bands p13–21, and this cluster represented 20.3% (46/227) of all labeled sites (cumulative probability for the Poisson distribution is $<<0.0005$). The majority of the labeled sites were located at 7p14–21. To examine the possibility that the specific labeling of 7p, and in particular, the labeling

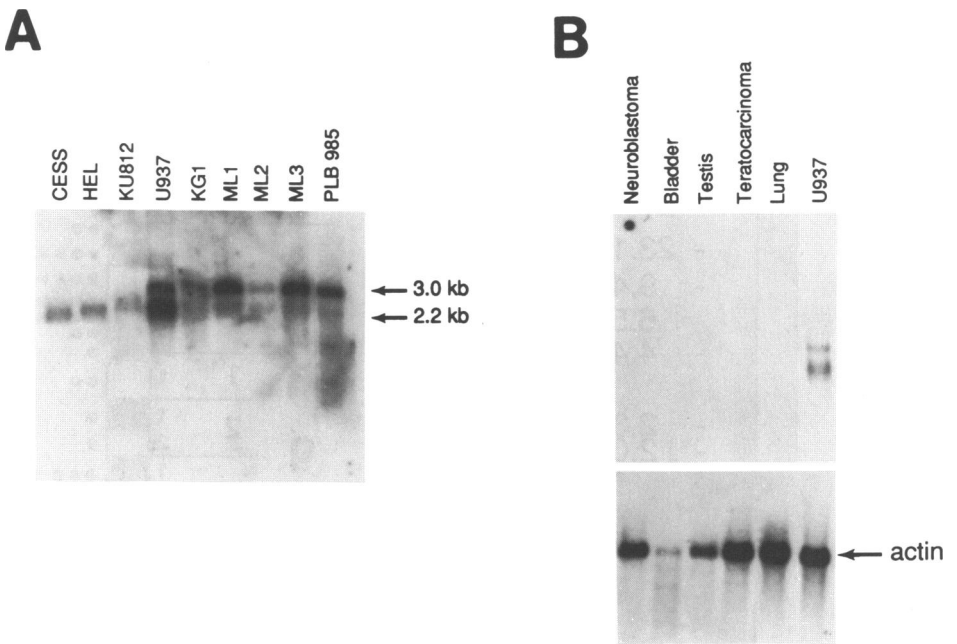

Figure 4. Myeloid specific expression of PL. Northern gel analysis of PL expression in human hematopoietic cell lines (**A**) and non-hematopoietic cell lines and tissue (**B**). Both blots were hybridized with probe E which contains 3′ flanking sequence which is common to PL1 and PL2. The cell phenotypes are described in the text. U937 cells were included on blot B to act as a positive PL signal. An actin probe was used to demonstrate equivalent mRNA loading.

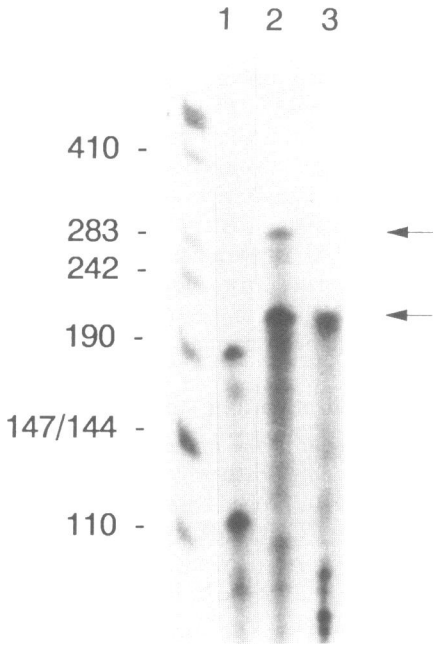

Figure 5. Expression of PL in bone marrow cells. RNAse protection was used to assess PL expression in normal marrow and in U937 cells. Lane 1 shows protection of probe C by U937 mRNA. Lanes 2 and 3 show protection of probe D by mRNA from U937 cells and bone marrow, respectively. Probe C is 217 nt and contains 90 nt of PL2 specific sequence and 127 nt of sequence common to PL1 and PL2. Probe D is 328 nt and contains 90 nt of PL2 specific sequence and 238 nt of sequence common to PL1 and PL2. The top arrow shows the 328 nt band representing protection of the entire D probe by PL2 transcripts, and the bottom arrow shows the 238 nt band representing protection of the common PL region of probe D by transcripts thought to represent PL1.

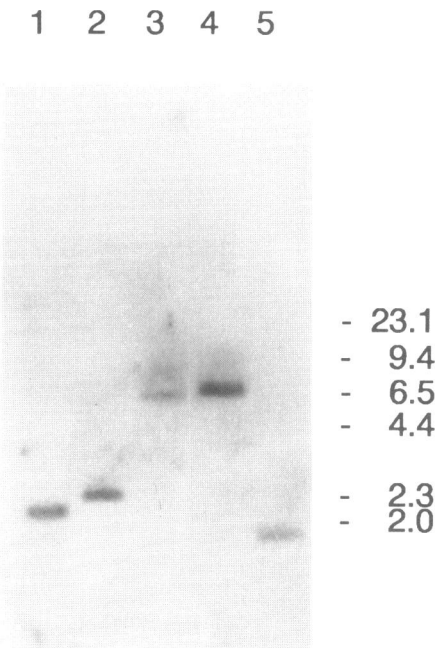

Figure 6. Southern analysis of PL gene copy number. Restriction digests of human genomic DNA were hybridized to a PL homeobox containing single stranded DNA probe (probe E). Lane 1: ApaI; Lane 2: BamHI; Lane 3: HindIII; Lane 4: EcoRI; Lane 5: StyI.

at 7p14−21 observed in these hybridizations was due to cross hybridization to the other homeobox genes located on 7p, hybridization experiments were performed using a PL2 3′ flanking region probe (B, Figure 1). In this case, also, we noted specific labeling only of chromosome 7. Of 100 metaphase cells examined, 25 cells were labeled on regions p1−p2 of one or both chromosome 7 homologues. A total of 44 grains were observed on this chromosome; of these, 34 (77%) were clustered at bands 7p13−22 and represented 16% (34/212) of all labeled sites ($P << 0.0005$). The largest cluster of grains was observed at 7p14−21. Thus, the PL gene is localized to chromosome 7, at bands p14−21.

DISCUSSION

We report here the cloning and partial sequencing of a myeloid-specific homeobox gene with alternatively spliced transcripts which are conceptually translated into homeodomain-containing proteins of 496 and 94 amino acids, respectively. Alternative splicing to produce variant protein products appears to be a frequent event in homeobox containing genes. For example, an Abd-B homeoprotein which contains a long amino terminal region is thought to function as a transcriptional activator while a truncated Abd-B protein, which lacks the activating region, functions as a repressor (20). In contrast, the ERA-1 gene which contain the Hox 1.6 homeobox (21) is expressed as two transcripts one of which encodes a protein which contains the homeodomain while the other encodes a truncated protein containing only the amino terminal region.

The products of PL gene expression reported here are similar to Abd-B and the converse of ERA-1 expression products. While the homeodomain has been shown to bind to DNA (22), the flanking regions are thought to confer, in part, the properties of transcriptional activation or repression on the homeoprotein. The two putative PL proteins described here both contain the homeobox and would be likely to compete for the same target DNA. Although the PL1 protein lacks several of the known activation motifs such as acidic (23), glutamine (24) or serine/threonine rich regions (25), there is a proline rich domain (residues 215−238) which may function as an activation region (26). In addition, there are islands of arginine which could form

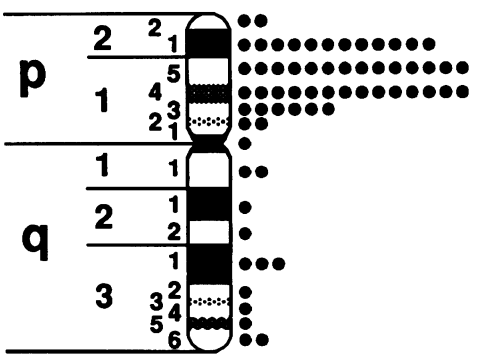

Figure 7. In situ localization of the PL gene to chromosome 7. Distribution of labeled sites on chromosome 7 in 100 normal human metaphase cells from phytohemagglutinin-stimulated peripheral blood lymphocytes that were hybridized with the PL HG-19 genomic probe. The labeled sites observed in the hybridization were clustered at 7p14−21.

a second DNA binding site, as well as glycine rich regions which should yield an extended protein structure. Since the putative PL2 protein lacks this amino terminal element, it may function as a non-regulated DNA binding protein interacting with a number of DNA binding sites. Expression of the two PL proteins will permit analysis of their respective roles in transcriptional regulation.

Knowledge of the chromosomal localization of the PL gene may also provide insight into its interaction with other homeobox-containing genes. The PL homeodomain is a member of the Antennapedia class I homeobox proteins (27). While this work was in progress, Boncinelli and coworkers described the cloning of several new human homeobox genes which are arranged along the chromosome in patterns which parallel homeobox gene patterns in Drosophila. One of these genes, 1H, is identical to PL within the homeodomain (28). These authors assigned the 1H homeobox to the 5′ end of the Hox 1 cluster on chromosome 7 and our localization of the PL gene to 7p14−21 is consistent with its placement within the HOX 1 locus (29). On the basis of the nomenclature of Acampora *et al.* (28) and the tentative designations of the murine Hox 1 locus by Kessel and Gruss (30), we suggest that the human PL homeobox gene be designated HOX 1.8. Boncinelli and colleagues also report that a previously described conserved peptide YPWM, is contained in the limited number of human class I homeobox proteins for which the flanking regions have been determined (28). However, neither the PL1 nor PL2 protein contains this sequence.

PL expression appears to be limited to hematopoietic cells and is predominantly localized to cells with myeloid potential. Although a range of human hematopoietic cell lines express varying amounts of the PL1 and PL2 transcripts, bone marrow apparently contains predominantly the PL1 transcript, which would encode a full-length homeobox protein. We hypothesize that PL proteins play a role(s) in control of myeloid cell phenotypic differentiation in a manner similar to that of MyoD in muscle cells (31). Transfection of sense and antisense PL1 or PL2 expression plasmids into human hematopoietic cell lines should help to elucidate the role of PL in hematopoiesis.

ACKNOWLEDGEMENTS

We would like to thank T.A.S.Eason, R.Espinosa,III and A.Fernald for expert technical assistance and F.Hack for maintenance of cell lines, Drs E.Campbell, B.Edwards, P.Gumerlock, T.Papayannopoulou, T.Rado for generous donation of cell lines, Dr C.S.Craik for donation of the Charon 4A genomic library, and Dr T.Papayannopoulou for a gift of KU812 mRNA. Supported in part by the Veterans Affairs Research Service (C.L. and H.J.L.) and American Cancer Society Grant #CH-459 (C.L.), NIH Grant #CA 47866 (H.J.L.), and PHS Grant #CA 40064 (M.M.L.). J.C. is a University of California, Davis, Medical Scholar; K.D. is a University of California Cancer Research Coordinating Committee Fellow, and M.M.L. is a Scholar of the Leukemia Society of America.

REFERENCES

1. Bodner,M., J.-L.Castrillo, L.E.Theill, T.Deerinck, M.Ellisman and M.Karin (1988) *Cell* **55**, 505−518.
2. Ingraham,H.A., R.Chen, H.J.Mangalam, H.P.Elsholtz, S.E.Flynn, C.R.Lin, D.M.Simmons, L.Swanson and M.G.Rosenfeld (1988) *Cell* **55**, 519−529.
3. Scheidereit,C., J.A.Cromlish, T.Gerster, K.Kawakami, C.-G.Balmaceda, R.A.Currie and R.G.Roeder (1988) *Nature* **336**, 551−557.
4. Frain,M., G.Swart, P.Monaci, A.Nicosia, S.Stampfli, R.Frank and R.Cortese (1989) *Cell* **59**, 145−157.
5. Biggin,M.D. and R.Tjian (1989) *Cell* **58**, 433−440.
6. Doe,C.G., Y.Hiromi, W.J.Gehring and C.S.Goodman (1988) *Science* **239**, 170−175.
7. Han,K., M.S.Levine and J.L.Manley (1989) *Cell* **56**, 573−583.
8. Shen,W.-F., C.Largman, P.Lowney, C.Hauser A., T.A.Simonitch, F.M.Hack and H.J.Lawrence (1989) *Proc. Natl. Acad. Sci. USA* **86**, 8536−8540.
9. Kongsuwan,K., E.Webb, P.Housiaux and J.M.Adams (1988) *EMBO J.* **7**, 2131−2138.
10. Kishi,K (1985) *Leuk. Res.* **9**, 381−390.
11. Chirgwin,J.M., A.E.Przybyla, R.J.MacDonald and W.J.Rutter (1979) *Biochem.* **18**, 5294−5299.
12. *Current Protocols in Molecular Biology* (1987) Edited by F.M.Ausubel. Greene Pub. Assoc. and Wiley Interscience.
13. *Molecular Cloning*. Maniatis,T., E.F.Fritsch and J.Sambrook (1982) Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
14. Lebo,R.V., F.Gorin, R.J.Fletterick, F.T.Kao, M.C.Cheung, B.D.Bruce and Y.W.Kan (1985) *Hum. Genet.* **69**, 316−320.
15. Deaven,L.L., M.A.V.Dilla, M.F.Bartholdi, A.V.Carrano, L.S.Cram, J.C.Fuscoe, J.W.Gray, C.E.Hildebrand, R.K.Moyzis and J.Perlman (1986) *Cold Spring Harbor Symp. Quant. Biol.* **51**, 159−167.
16. Le Beau,M.M., C.A.Westbrook, M.O.Diaz and J.D.Rowley (1984) *Nature* **312**, 70−71.
17. Kuziora,M.A. and W.McGinnis (1989) *Cell* **59**, 563−571.
18. Gay,N.J., S.J.Poole and T.B.Kornberg (1988) *Nucl. Acids Res.* **16**, 6637−6647.
19. Ptashne,M. (1988) *Nature* **335**, 683−689.
20. Zavortink,M. and S.Sakonju (1989) *Gene Dev.* **3**, 1969−1981.
21. LaRosa,G.J. and L.J.Gudas (1988) *Mol. Cell Biol.* **8**, 3906−3917.
22. Hoey,T. and M.Levine (1988) *Nature* **332**, 858−861.
23. Ma,J. and M.Ptashne (1987) *Cell* **51**, 113−119.
24. Courey,A.J., D.A.Holtzman, S.P.Jackson and R.Tjian (1989) *Cell* **59**, 827−836.
25. Theill,L.E., J.-L.Castrillo, D.Wu and M.Karin (1989) *Nature* **342**, 945−948.
26. Mermod,N., E.D.O'Neill, T.J.Kelly and R.Tjian (1989) *Cell* **58**, 741−753.
27. Gehring,W.J. (1987) *Science* **236**, 1245−1252.
28. Acampora,D., M.D'Esposito, A.Faiella, M.Pannese, E.Migliaccio, F.Morelli, A.Stornaiuolo, V.Nitro, A.Simeone and E.Boncinelli (1989) *Nucl. Acids Res.* **17**, 10385−10402.
29. Tsui,L.-C., M.Farrall and H.Donis-Keller (1989) *Cytogenet. Cell Genet.* **51**, 166−201.
30. Kessel,M. and P.Gruss (1990) *Science* **249**, 374−379.
31. Davis,R.L., H.Weintraub and A.B.Lassar (1987) *Cell* **51**, 987−1000.