
Michael J. Klemsz, Scott R. McKercher, Antonio Celada, Charles Van Beveren and Richard A. Maki

J Immunol 2008; 181:1597-1608; ;
http://www.jimmunol.org/content/181/3/1597.citation
The Macrophage and B Cell–Specific Transcription Factor PU.1 Is Related to the ets Oncogene

Michael J. Klemasz, Scott R. McKercher, Anthony Pelada, Charles Van Beveren, and Richard A. Maki

Cancer Research Center
La Jolla Cancer Research Foundation
10901 North Torrey Pines Road
La Jolla, California 92037

Summary

We have isolated a cDNA clone, PU.1, that codes for a new tissue-specific DNA binding protein. Analysis of the binding site by methylation interference and DNAase I protection revealed that the PU.1 protein recognized a purine-rich sequence, 5′-GAGGAA-3′ (PU box). The PU.1 protein was shown to be a transcriptional activator that is expressed in macrophages and B cells. cDNA constructions used to generate proteins lacking portions of either the amino- or carboxy-terminal ends of the PU.1 protein placed the DNA binding domain in the highly basic carboxy-terminal domain of the protein. The amino acid sequence in the binding domain of PU.1 has considerable identity with proteins belonging to the ets oncogene family.

Introduction

The transcription of eukaryotic genes is regulated by specific DNA binding proteins that assemble on cis-acting DNA sequences in both promoters and enhancers (reviewed in Maniatis et al., 1987; Jones et al., 1988; Plashne, 1988; Strehl, 1989). Many of these DNA binding proteins are ubiquitous in their expression and probably serve a general role in gene transcription. Others are restricted in their expression to one or a few cell types. Several examples have been documented in which the expression of this latter type of DNA binding protein correlates with expression of genes to which they bind (Staudt et al., 1988; Müller et al., 1988; Scheidenreit et al., 1988; Bodner et al., 1988; Frain et al., 1988). The biochemical characterization of these cell type-specific activators of gene expression will provide a critical step in our understanding of gene regulation.

Recently, a novel genetic approach was used to isolate a variant of simian virus 40 (SV40) in which growth was restricted to CV-1 cells and lymphoid cells (Petterson and Schaffner, 1987). This variant was selected from an SV40 construction that lacked the 72 bp enhancer region. Characterization of this variant revealed that a duplicate region containing a purine-rich sequence called the PU box (5′-GAGGAA-3′) was able to substitute for the deleted wild-type enhancer. Nuclear factors present only in lymphocytes bound to the PU box, and the binding was correlated with the growth of the SV40 variant. We have used an in situ filter binding technique (Singh et al., 1988) to clone a mouse DNA binding protein, PU.1, that may be related to this lymphoid-specific enhancer activity. The PU.1 protein binds to the PU box, is a transcriptional activator, and is expressed in macrophages and B cells. The amino-terminal end of PU.1 is highly acidic, while the carboxy-terminal end is highly basic and encodes the DNA binding domain. The DNA binding domain of PU.1 exhibits considerable amino acid sequence identity with proteins encoded by the ets oncogene family.

Results

Isolation of the PU.1 Gene

In our studies on the promoter of the MHC class II I-Aβ gene, we have demonstrated the binding of a nuclear factor to a CCAAT sequence present at −60 to −65 from the transcription start site (Pelada et al., 1986). Attempts to isolate the gene for this CCAAT binding factor using an in situ filter binding technique (Singh et al., 1988) have not been successful. This may be due to the finding that this CCAAT binding factor requires two components for efficient binding of the protein to DNA (Pelada and Maki, 1989a). While screening a macrophage λgt11 cDNA library with a double-stranded multimerized 33-mer oligonucleotide that contained the sequence from −76 to −46 of the MHC class II gene I-Aβ (5′-gtacctgctcccaatgtgaggaaccaactcaggttga-3′), however, we isolated several clones of a DNA binding factor we have named PU.1. An initial screening of 500,000 phage plaques yielded two recombinant clones designated 125.1 and 124.1. As shown below, DNA sequence analysis revealed that both clones code for the same gene. Since the 125.1 insert was larger, it was used in many of the subsequent experiments.

To demonstrate that 125.1 contained a β-galactosidase fusion protein that had the capacity to bind DNA in a sequence-specific manner, a Southwestern blot was performed using a labeled tetramer of the 33-mer as a probe (Figure 1A). Bacterial extracts prepared from λgt11 and 125.1 lysogens were fractionated on a 75% SDS-polyacrylamide gel, transferred to nitrocellulose, and probed with labeled multimerized 33-mer. Only the lane with the 125.1 extract contained a protein able to bind the probe. When a duplicate filter was incubated with a different multimerized oligonucleotide, no signal was seen in either lane, indicating that binding to the multimerized 33-mer sequence was specific (data not shown). The positive filter was subsequently treated with rabbit antiserum to β-galactosidase and mouse anti-rabbit immunoglobulin horse-radish peroxidase conjugate (Figure 1B). In the lane containing the 125.1 extract, the slowest migrating species had an apparent molecular size of 150 kd. The appearance of several bands of lower molecular size is probably the result of degradation of the 150 kd species. The largest stained band in the λgt11 lane had a molecular size equivalent to β-galactosidase (MW = 116,000). Some breakdown of the β-galactosidase occurred in this sample resulting in a band at about 69 kd. These results indicated that the extracts prepared from the λgt11 lysogen con-
DNA Binding Specificity of the Fusion Protein

A series of methylation interference experiments using the 33-mer and the 25.1 fusion protein revealed that the fusion protein protected the three guanines within the sequence 5’-GAGGAA-3’ (data not shown). A subsequent search of previously described DNA binding activities revealed an activity present in lymphocytes that binds to a purine-rich sequence present within the SV40 enhancer (Petterson and Schaffner, 1987). The binding site for this activity contains the core sequence 5’-GAGGAA-3’ and was called the PU box.

To determine if the 25.1 fusion protein could bind to the PU box in the SV40 enhancer, a double-stranded oligonucleotide (PU-1) was synthesized corresponding to this region of the SV40 enhancer (Figure 2B). This oligonucleotide was cloned into the BamHI site of the plasmid pGEM1 and subsequently reisolated as a fragment for the gel electrophoresis DNA binding assay. In addition, for these and subsequent experiments, the 25.1 β-galactosidase fusion protein was partially purified using anti-β-galactosidase–Sepharose. This partially purified protein will be referred to as Ly25.1. When the PU-1 fragment was labeled and a gel electrophoresis DNA binding assay performed using Ly25.1, a retarded complex was observed (Figure 2A). This complex was effectively competed by a 50-fold molar excess of cold PU-1. Oligonucleotides that contained an altered PU core sequence (PU mut 5) or those in which the central two guanines were changed to cytosines within the PU box (PU mut 1) did not effectively compete for binding of Ly25.1 to labeled PU-1. These results suggested that Ly25.1 recognized the PU box in a sequence-specific manner. Since Ly25.1 had a higher affinity for the PU box of PU-1 than the PU box in the 33-mer, PU-1 was used in all subsequent experiments.

A search of other nuclear factor binding sites revealed a similarity in the PU box core sequence (5’-GAGGAA-3’) with the core sequence of the AP-3 binding site (5’-GTGGGAA-3’) (Chiu et al., 1987). When a double-stranded oligonucleotide containing the AP-3 binding site was used as a competitor for binding of the Ly25.1 protein to the oligonucleotide PU-1, no competition was observed, even at a 500-fold molar excess. Based on this result we conclude that Ly25.1 is not the transcription factor AP-3.

To delineate the Ly25.1 binding site on PU-1 further, we performed methylation interference experiments using the double-stranded PU-1 oligonucleotide (Figure 3). Binding of Ly25.1 to the PU-1 oligonucleotide resulted in the protection of three guanine residues on the upper strand within the core PU sequence (see Figure 2B for sequence). No guanine residues on the lower strand were protected. This footprint is identical to a footprint generated by a lymphocyte-specific PU box binding protein described earlier (Petterson and Schaffner, 1987). Thus, the data are consistent with PU1 being a mouse PU box DNA binding protein.

Comparison of Ly25.1 to a Binding Activity Present in Macrophage Nuclear Extracts

To determine if the Ly25.1 protein bound to DNA in a manner similar to that of a nuclear factor present in macrophages, the binding pattern of the Ly25.1 protein and the PU binding activity present in macrophages were compared by DNAase I footprinting using the cloned double-
stranded PU.1 oligonucleotide (Figure 4). A DNAase I-protected region of 16–17 bp centered over the PU box was obtained with both Ly25.1 and a nuclear extract prepared from P388 D1. These data demonstrate that the murine macrophage cell line P388 D1 expresses a nuclear protein that binds to the PU box in a manner similar to that seen for Ly25.1.

DNA Sequence of PU.1

Both the λ25.1 and λ24.1 cDNA inserts were subcloned into the Bluescript KS+ vector (Stratagene, Inc., La Jolla, CA) and sequenced (Figure 5). In addition, the insert of λ25.1 was used to screen a WEHI-3 macrophage cDNA library. The sequence of the longest clone isolated, W25.35.1, was determined and found to be identical to the sequence obtained from clones λ25.1 and λ24.1 where the sequence overlapped (Figure 5). Clone λ25.1 began with nucleotide 117, while clone λ24.1 started at nucleotide 517. Inspection of the PU.1 cDNA sequence revealed an open reading frame of 816 bases, which codes for a protein of 272 amino acids. This reading frame was also in frame with the fusion protein reading frames of λ25.1 and λ24.1. Analysis of multiple cDNAs isolated from the WEHI-3 cDNA library were used to characterize the 5' and 3' termini of PU.1. We have assigned the initial methionine residue based on this sequence information. A methionine codon located 7 amino acids downstream, however, has a nucleotide sequence flanking the ATG codon that is a better match to the consensus sequence for the initiation of eukaryotic translation (Kozak, 1986). Without additional information defining the sequence of the amino-terminal end of the protein, we will call the first methionine codon the start of the protein sequence. A stop codon 63 bases upstream in the same reading frame as the first methionine suggests that we have the entire PU.1 coding sequence.

A hydrophilicity plot of the predicted amino acid sequence revealed two major domains (Figure 5). The amino-terminal half of PU.1 is slightly hydrophobic, while the carboxy-terminal half is hydrophilic. These two domains are separated by a PEST sequence (a region rich in proline [P], glutamic acid [E], serine [S], and threonine [T]), which has been correlated with the increased degradation of intracellular proteins (Rodgers et al., 1986). An inspection of the amino acid sequence revealed that the amino-terminal 60% of PU.1 is highly acidic (16% aspartic plus glutamic acids versus 2% lysine plus arginine in the first 165 amino acids) and contains a glutamine-rich region (31%) between amino acids 75 and 93. Both of these components have been linked to the transcriptional activating qualities of several DNA binding proteins (Pashne, 1988; Kadonaga et al., 1988; Courre et al., 1988). No ser-
Figure 5. DNA Sequence of PU.1

(A) A diagram of the two original phage clones, 25.1 and 124.1, and the full-length cDNA clone, W25.35.1, isolated from a WEHI-3 cDNA library. The open box represents the coding sequence, while the black line represents the noncoding sequence. (B) DNA sequence of the full-length PU.1 gene and the predicted amino acid sequence. Start points for the original two clones are nucleotide 117 for 25.1 and nucleotide 517 for 124.1.

The image includes a DNA sequence diagram and text discussing the sequence and its implications for understanding the protein's structure and function.
Cloning of the PU.1 Gene

117

Figure 6. Hydrophilicity Plot of the Protein Sequence Predicted for PU.1.
The analysis was performed using the algorithm of Hopp and Woods (1981). Hydrophilic values are plotted above the center line. The core PEST sequence extends from amino acid 99 to 159.

carboxy-terminal region. As will be shown below, the carboxy half of the protein encodes the DNA binding domain.

Mapping of the DNA Binding Domain

A series of experiments were designed to map the DNA binding domain of the PU.1 protein, using the technique of in vitro transcription and translation of a series of PU.1 constructions. These products were tested in the gel electrophoresis DNA binding assay using the cloned SV40 PU.1 oligonucleotide as a probe (Hope and Struhl, 1985; Sturm et al., 1988). The λ25.1 cDNA was inserted into the Bluescript KS+ vector (p25.1) such that sense RNA could be produced in vitro using the T3 promoter and T7 RNA polymerase. RNAs generated from p25.1 and the deletion clones were translated in vitro using a rabbit reticulocyte lysate. The in vitro translated proteins were fractionated by SDS–polyacrylamide gel electrophoresis, and proteins of the appropriate size were detected for each deletion.

The results of these experiments demonstrated that the DNA binding activity of the PU.1 protein was localized within the carboxyl terminus (Figure 7A). When p25.1 was digested with XbaI, which cut in the polymerase just outside of the cDNA insert, a full-length protein (272 amino acids) was produced in the in vitro transcription/translation system. This protein gave a retarded band in the gel electrophoresis DNA binding assay using the cloned PU.1 double-stranded oligonucleotide as a probe (Figure 7B). The plasmid, p25.1, was also digested with NarI, which cut within the coding sequence, such that the in vitro synthesized protein was 5 amino acids shorter (267 amino acids) than the full-length protein. The resultant protein bound to the labeled PU.1 fragment as shown by a retarded band

Figure 7. Mapping of the Binding Domain of PU.1

(A) The indicated restriction enzymes were used to generate deletions of p25.1 for in vitro transcription and translation reactions, followed by gel electrophoresis DNA binding analyses. Binding activity is shown for each construction. The region required for binding is stippled in the top diagram. The size of the inserts encoded by the two original phage clones is shown for comparison.

(B) Gel electrophoresis DNA binding assay of proteins generated as described in (A) using the cloned PU.1 oligonucleotide as a probe. The dash indicates probe alone, and BMV (brome mosaic virus) was included as a control.
in the gel electrophoresis DNA binding assay. Digestion of p25.1 with BssHII resulted in an in vitro translated protein of 230 amino acids that no longer bound to the PU.1 fragment (Figure 7B). Therefore, we conclude that the 42 carboxy-terminal amino acids of the protein are necessary for DNA binding.

The NciI construction, which removed 117 amino-terminal amino acids, created a 155 amino acid in vitro translated product by using an internal methionine (Met-115) as a translation start site. The protein generated in this manner bound to the PU.1 probe and gave a retarded band. This result was similar to that achieved with the Ly24.1 fusion protein. Finally, an internal deletion of p25.1 (NcoI–KpnI) produced a protein of 172 amino acids that did not bind the cloned PU.1 oligonucleotide.

These results established 5' and 3' boundaries for the DNA binding domain of the PU.1 protein. The 5' deletions and the result from the Ly24.1 clone placed the amino-terminal end of the DNA binding domain between amino acids 129 and 200. The 3' deletions placed the carboxy-terminal end of the DNA binding domain between amino acids 230 and 267.

Expression of PU.1 mRNA Is Tissue Specific
To define the cell type specificity of PU.1 expression, we examined various cell lines and tissues by Northern blot analysis using the PU.1 cDNA as a probe. We found that the PU.1 gene was expressed in the macrophage-like cell lines WEHI-3, P388 D1, and WR19 M.1 as well as in bone marrow–derived macrophages (Figure 8). The size of this RNA was 1.4 kb, which agrees well with the length of the longest cDNA clone isolated. Since the activation of macrophages is primarily mediated by the lymphokine interferon γ (IFN-γ), we incubated WR19 M.1 cells and bone marrow derived macrophages in either the presence or absence of IFN-γ (300 IU/ml) for 24 hr before RNA was isolated. Macrophages treated with IFN-γ under these conditions have been shown to induce I-A antigen expression (Celada and Maki, 1999b). We observed no change in the level of PU.1 expression in these cells after IFN-γ treatment (Figure 8D). The PU.1 gene was also expressed in the B cell lymphoma A20-2J as well as the plasmacytomas 70A and MOPC315 and at a lower level in the plasmacytoma P3 (Figure 8C). There appeared to be no expression of PU.1 in the mouse fibroblast cell line L929 (Figure 8C), and, interestingly, there was no expression in the T cell line EL4 (Figure 8A). We extended this analysis to additional T cell lines and found no expression in the T cell lines OH2 and Ot6 (Kanagawa, 1988), BWS147 (Ralph, 1973), A32-26 (Koehler et al., 1977), or in L3 (Glasebrook and Fish, 1980) (data not shown). When RNA was analyzed from the thymus, however, we observed expression of the PU.1 gene. At present, we do not know whether the PU.1 expression seen in the thymus is due to a subset of T cells that express PU.1 or by contaminating macrophages or other cells that express PU.1. In other tissues examined (Figure 8B), PU.1 was expressed in the spleen as expected, but not in the brain, heart, kidney, or lung. Since PU.1 is not expressed in the liver, PU.1 is probably not the nuclear factor IEFga, which recognizes the sequence 5'-GGGAAA-3' (LaMarco and McKnight, 1989). The same blots used for PU.1 hybridization in Figures 8A and 8B were subsequently hybridized with a probe for the ribosomal gene L32 (Dudov and Perry, 1984) to ensure that the RNA was not degraded and to control for the amount of RNA added to each lane.

PU.1 Is a Transcriptional Activator
The PU.1 protein was tested for its ability to regulate tran-
Cotransfection of HeLa cells, which were previously shown to lack PU box DNA binding activity (Pettersson and Schaffner, 1987). A series of DNA constructions were made using the vector pBLCAT2, which contains the chloramphenicol acetyltransferase (CAT) gene linked to the thymidine kinase (TK) promoter (Lusky and Schultz, 1987). From one to eight copies of the PU-1 oligonucleotide were inserted both 5’ and 3’ of the CAT gene, as well as in both orientations. Expression of the PU-1 protein was generated by inserting the Xl25.1 gene into the vector pECE, which places the expression of a foreign gene under the control of the SV40 early promoter (Ellis et al., 1986). Each CAT construction was cotransfected with the β-galactosidase expression plasmid pCH110 and with or without the PU pECE expression plasmid. All CAT values were then normalized to the level of β-galactosidase expression to correct for any differences in transfection efficiency. When the pu pECE plasmid was cotransfected with CAT constructions containing one PU element, CAT activity increased 3- to 4-fold (Table 1). The orientation of the PU box did not drastically alter the level of CAT activity measured. When four copies of the PU box were inserted upstream of the TK promoter, and this construction was cotransfected with PU pECE, the CAT activity increased 4-fold. When eight pu boxes were inserted upstream of the TK promoter, CAT activity increased 8-fold in the presence of the PU element. No increase in CAT activity was seen when the pECE vector or a retinoic acid receptor pECE expression plasmid (Graupner et al., 1989) was cotransfected with the PU box containing CAT vectors, suggesting that expression of the PU-1 protein was necessary for an increase in CAT activity to be seen (data not shown).

When the vector pBLCAT2 was cotransfected with the PU pECE plasmid, no increase in CAT activity was observed. Thus, only those CAT plasmids containing a PU box could be activated by the PU-1 protein. Curiously, a consistent increase in CAT activity was observed with plasmids containing the PU-1 or PU mut 1 oligonucleotides as compared with pBLCAT2. At present, we do not have a satisfactory explanation for this result, except that we have noticed a similar increase in CAT activity when several other DNA fragments were inserted upstream of the TK promoter in this plasmid.

The PU-1 protein was shown not to bind to a PU box in which the two central guanines have been changed to cytosines (PU mut 1; Figure 2). This oligonucleotide was inserted into the plasmid pBLCAT2 and CAT activity measured in the cotransfection assay with the pu pECE plasmid. No trans-activation was observed in this experiment, suggesting that in order for PU-1 protein to increase the level of CAT activity in this assay, it must bind to the DNA.

The effect of placing the PU box downstream of the CAT gene was also determined. In these examples, when four or eight copies of the PU-1 box containing oligonucleotide were placed downstream of the CAT gene and cotransfected with the pu pECE plasmid, a modest increase of about 3-fold in CAT activity was observed. Therefore, we conclude that trans-activation by PU-1 can take place when the PU box is located either upstream or downstream of the CAT gene.

**Table 1. Trans-Activation by the PU-1 DNA Binding Protein**

<table>
<thead>
<tr>
<th>Reporter Constructions</th>
<th>% Acylation</th>
<th>Fold Increase</th>
</tr>
</thead>
<tbody>
<tr>
<td>No Expression Vector</td>
<td></td>
<td></td>
</tr>
<tr>
<td>pBLCAT2</td>
<td>2.72</td>
<td>2.00</td>
</tr>
<tr>
<td>PU (x X) B</td>
<td>5.60</td>
<td>19.32</td>
</tr>
<tr>
<td>PU (x X) 8X</td>
<td>5.54</td>
<td>19.97</td>
</tr>
<tr>
<td>PU (x X) X</td>
<td>9.04</td>
<td>34.80</td>
</tr>
<tr>
<td>PU (x X) X</td>
<td>11.30</td>
<td>43.50</td>
</tr>
<tr>
<td>PU (X X) X</td>
<td>9.14</td>
<td>167.54</td>
</tr>
<tr>
<td>PU mut (x X) B</td>
<td>5.62</td>
<td>2.15</td>
</tr>
<tr>
<td>PU (X X) X</td>
<td>3.36</td>
<td>9.77</td>
</tr>
<tr>
<td>PU, 3’ (x X) Z</td>
<td>19.66</td>
<td>24.12</td>
</tr>
<tr>
<td>PU, 3’ (x X) Z</td>
<td>10.15</td>
<td>26.86</td>
</tr>
<tr>
<td>PU, 3’ (x X) Z</td>
<td>10.89</td>
<td>29.06</td>
</tr>
</tbody>
</table>

Table 1 contains the data from a representative experiment. Each reporter construct was tested more than six times. The level of acetylation was corrected for differences in transfection efficiency using a β-galactosidase expression plasmid. The PU (X X) X sample was adjusted for each reporter activity. Relative activities in pBLCAT2 are as follows: B. BamHI; X. XbaI; S. SalI; Z. Smal.

Notation for reporter constructions: oligonucleotide inserted # of binding sites and orientation of the PU element, arrow to the right indicates 5’ GAAGCCAA-3’ on upper strand; polylinker insertion point; X indicates insertion of oligonucleotides 3’ of the CAT gene.

**Discussion**

The results presented here describe the cloning of the gene for a new transcription factor we have named PU-1. Originally, PU was defined as a lymphoid-specific transcriptional activity that allowed the growth of an SV40 variant in lymphoid cells (Pettersson and Schaffner, 1987). The activity was shown to bind to a PU box within a duplicated sequence that substitutes for the wild-type enhancer in this SV40 variant. Based on several criteria, it is likely that PU-1 codes for a protein with the same transcriptional activity described previously for PU. First, the DMS footprint generated by the PU-1 protein on the PU box is identical with the DMS footprint obtained using nuclear extract prepared from the B cell line BJAB (Pettersson and Schaffner, 1987). Second, the PU-1 protein must bind to DNA in order for it to act as a transcriptional activator, which is consistent with the PU activity described earlier in lymphoid cells. Third, the cell type-specific expression of PU-1 is consistent with the previous work, which found PU binding activity in B cells and T cells. The interesting feature of PU-1, however, is that it is only expressed in B cells and macrophages but not in T cell lines (Figure 8). This suggests that a different PU box binding protein may be present in T cells. Preliminary data from our laboratory using the technique of proteolytic clipping (Schreiber et al., 1986) to characterize the PU binding proteins in macrophages, B cells, and T cells suggest that T cells do express a different PU DNA binding protein than is expressed in macrophages and B cells (unpublished...
The proteins examined are indicated at the left and include mouse ets-2, human ets-1, human ets-2 (Watson et al., 1988), Drosophila ets-2 (Prybil et al., 1988), sea urchin ets-2 (Protein Identification Resource, release 22), human ets-1 (Reddy et al., 1987), and human ets-1 (Rao et al., 1988). The percent identity is indicated at the right. The dots indicate the amino acids that are identical with the amino acid sequence in the PU.1 protein. The stars indicate the basic amino acids that are conserved. Restriction sites are indicated as in Figure 7.

**Binding of PU.1 to DNA**

Experiments conducted to identify the DNA binding domain of PU.1 placed this domain at the carboxy-terminal end of the protein. The carboxy-terminal 144 amino acids appear to be sufficient for DNA binding (Figure 7). We found no evidence that this amino acid sequence resembled either a zinc finger (Miller et al., 1986), leucine zipper (Landschulz et al., 1988), helix-loop-helix (Murre et al., 1989), antiparallel β sheet (Tanaka et al., 1984), or the SPK histone motif (Suzuki, 1989). An analysis of the PU.1 protein, however, using either the algorithm of Garner et al. (1978) or Chou and Fasman (1978) revealed a helix-turn-helix motif in the carboxy-terminal half of the protein that resembled the helix-turn-helix motif found in a number of other DNA binding proteins (Pabo and Sauer, 1984). A tri-helical structure, each helix separated by a turn, was identified between the amino acids 207 and 249 of the PU.1 protein. This helix-turn-helix motif did not have significant amino acid sequence homology to homeodomain-containing proteins (Garcia-Blanco et al., 1989; Scott et al., 1989).

**The DNA Binding Domain of PU.1 is Related to the ets Gene Family**

A search of known protein sequences with regions of similarity to PU.1 revealed that the carboxy-terminal amino acid sequence of the PU.1 protein has significant identity with a region of amino acid sequence in the ets gene family (Figure 9). Within a stretch of 99 amino acids near the carboxy terminus of PU.1 there is 41% sequence identity with 95 amino acids located near the carboxy terminus of mouse ets-2, human ets-2, and human ets-1 proteins. In contrast, there is only a 6% sequence identity between the first 160 amino acids of PU.1 and mouse ets-2. For two other members of the ets gene family, human ets-1 and human ets-1, the percentage of amino acid identity to PU.1 within the 99 amino acid region is 37% and 38%, respectively.

The ets gene was originally identified in the avian leukemia virus E26, which causes erythroblastosis and myeloblastic leukemias in chickens (Moscovici et al., 1981; Radke et al., 1982). Although no function for the ets protein has been defined, it is found in the nucleus, is phosphorylated, and has affinity for DNA (Boulukos et al., 1988; Fujiiwara et al., 1988a; Fujiiwara et al., 1988b; Pognonec et al., 1989). Furthermore, it has been shown that mouse ets-1 and ets-2 are expressed in the thymus and may be involved in T cell activation and proliferation (Bhat et al., 1987; Pognonec et al., 1988; Bhat et al., 1989; Seth et al., 1989).

Recently, the region of c-ets-1 with a high degree of amino acid sequence identity with PU.1 was shown to be essential for nuclear localization and DNA binding (Boulukos et al., 1989). Thus, it is likely that common elements within this region are important for DNA binding. When several of the ets proteins were analyzed using the algorithms of Garner et al. (1978) or Chou and Fasman (1978), however, no helix-turn-helix motif was seen in this region. Thus, despite the sequence similarity, these proteins may employ a different DNA binding motif. Another possibility is that the amino acid identity between PU.1 and ets reflects a novel DNA binding motif. In this regard, there is a conserved repeat (amino acids 214–248) of basic amino acids (lysine or arginine) in a region known to be important for DNA binding. In the in vitro transcription experiments with PU.1, the Bsh111 construction generated a protein that lacked several of these basic amino acids and lost the ability to bind DNA. In the case of c-ets-1, loss of this basic region resulted in loss of DNA binding (Boulukos et al., 1989). Another area in
which PU.1 and the members of the ets family exhibit considerable amino acid sequence identity in amino acids 166–182 (PU.1 numbering). A PU.1 mutant protein lacking these amino acids also lacked DNA binding activity (Figure 7). It is not known whether these amino acids are directly involved in DNA binding or if the Ncol–KpnI deletion altered the conformation of the protein so as to prevent DNA binding. To answer this question, a more detailed analysis of the requirements for DNA binding by the PU.1 protein is required.

Transcriptional Activation by PU.1

The PU.1 protein was shown to be a transcriptional activator using a series of CAT constructions in which the PU box was inserted upstream of the TK promoter or downstream of the CAT gene. When any of these CAT constructions were cotransfected with the PU eCE expression plasmid, an increase in CAT activity was seen. Mutation of the PU box or the substitution of PU.1 protein with the retinoic acid receptor a did not increase in CAT activity. Based on these results, we conclude that PU.1 increases the level of CAT transcription by binding to DNA in a sequence-specific manner.

The PU.1 protein has a number of features that are characteristic of a transcriptional activator. As discussed above, the carboxyl-terminal 144 amino acids are highly basic and sufficient for DNA binding. The amino-terminal 165 amino acids contain a number of elements that have been associated with activation domains of transcription factors. This region of PU.1 is highly acidic with a net negative charge of −23 in the first 165 amino acids. The transcription factors GCN4 (Hope and Struhl, 1986) and GAL4 (Ma and Ptashne, 1987) have also been shown to contain net negative charges that are necessary for transcriptional activation. In addition, within the amino-terminal portion of PU.1, between amino acids 75 and 93, there is a high concentration of glutamines (31%). Glutamine-rich regions have been shown to be important for transcription activation mediated by the transcription factor SP1 (Courey and Tjian, 1988). Whether the glutamine-rich region or the net negative charge in the amino-terminal half of PU.1 are important for transcriptional activation is currently being assessed.

There is evidence that PU DNA binding proteins are involved in the regulation of gene expression. A PU binding activity is known to promote the growth of SV40 in B cells and T cells (Petterson and Schaffner, 1987) and has been implicated in the expression of the IL-2 gene (Ghia et al., 1988; Serfling et al., 1989). Putative PU binding sites have been found associated with a number of other genes, which includes GM-CSF (Miyatake et al., 1985; Nimer et al., 1988), G-CSF (Nagata et al., 1986), IL-1 (Clark et al., 1986; Furutani et al., 1986), IL-3 (Cohen et al., 1986), IL-5 (Tanabe et al. 1987; Campbell et al., 1988), TNF-α (Nedwin et al., 1985), TNF-β (Gardner et al., 1987), and lymphotoxin (Lucow and Schütz, 1989). Preliminary results suggest there may be a hierarchy of PU binding sites: PU.1 binds with higher affinity to the TNF-α PU box than the IL-2 PU box (unpublished data).

The isolation of the PU.1 gene will allow us to characterize further its functional domains. The fact that it is a cell type–specific transcriptional activator suggests that it may play an important role in the differentiation or activation of macrophages and B cells.

Experimental Procedures

Isolation of lgt11 Recombinant PU Binding Protein

Recombinant lgt11 clones were isolated from a normal murine macrophage cDNA library (Cinomet Laboratories, Inc., Palo Alto, CA) and confirmed using an in situ filter binding technique (Singh et al., 1988). Oligonucleotides were synthesized on a model 380A Applied Biosystems DNA Synthesizer (Foster City, CA) and purified using Applied Biosystems oligonucleotide purification cartridges. The probe used to screen the filters was a tetramer of the 33-mer oligonucleotide (5'-gaccctcccaaatdagaacaattcagcatc-3'), inverted in the BamHI site of the vector pGEM 1 (Promega Biotech, Inc., Madison, WI). The probe was isolated from the plasmid by digesting with HindIII and AvaI, labeled with [γ-32P]ATP, and purified by phenol extraction. The recombinant fragment of cell DNA was resolved and separated from the vector on an 8% polyacrylamide gel. Phage were plated at a density of 2 x 10^6 plaques per 150 mm plate for the first screening. Four rounds of plaque purification were used to purify the positive clones.

Sequencing of PU1 Clones

The inserts from the two lgt11 recombinant clones and the clone W2553,1 were subcloned into the Bluescript KS+ vector (Stratagene, Inc., La Jolla, CA), and DNA sequencing was performed on double-stranded plasmid templates using the dye-terminator method of Sanger (Sanger et al., 1977), according to the Sequenase protocol (U.S. Biochemical Corp., Cleveland, OH). Deletion clones were constructed by the cloning of smaller restriction fragments into the Bluescript KS+ vector. Both strands of DNA were sequenced with G and I reactions to resolve any compression discrepancies. The full-length cDNA clone W2553,1 was isolated from a W6-3G cDNA library constructed in ZAP II (Stratagene, catalog number 936304) by standard methods (Maniatis et al., 1982) using the lgt11 CDNA insert as a hybridization probe. Sequence analysis was performed using the Microgenie computer program (Beckman Instruments, Inc., Palo Alto, CA). Protein comparisons were conducted using the homology algorithm of GENEPRO (window = 29, matches = 11, id = 2, speed = 1, score = 100) (Riverside Scientific Enterprises, Seattle, WA). The analysis was performed on the sequences in the Protein Identification Resource (National Biomedical Research Foundation) release 22 (September, 1988).

Production of β-Galactosidase Fusion Protein and Western and Southwestern Analyses

Lymphocytes were prepared from the lgt11 and lgt24 clones as described (Sirimanna et al., 1988) and the fusion proteins partially purified using anti-β-galactosidase antibody bound to Sepharose (5 Prime-3 Prime, Inc., Baulk, CA), as described by the manufacturer. Western and Southwestern analyses were carried out as described (Saulit et al., 1988).

Northern Analysis

RNA was isolated from cell lines and mouse tissues as described (Chirgwin et al., 1979). Some samples of RNA were passed over oligo(dT). RNA was fractionated on a 1% agarose gel containing formaldehyde (Maniatis et al., 1982) and transferred to a nylon membrane as described by the manufacturer (Schleicher & Schuell, Inc., Keene, NH). The probes, 251 cDNA or L22-44, were labeled by the random oligo method (Pharmacia LKB Biotechnology, Inc., Piscataway, NJ). Hybridizations were performed at 42°C for 2 hr in buffer containing 50% formamide, 5 x SSC, 0.1% SDS, 10 μg/ml salmon sperm DNA, and 1 x 10^6 cpm/ml probe. The filters were washed three times with 0.1 x SSC, 0.1% SDS for 15 min at 37°C before exposing to Kodak XAR film. Hybridized probe was removed from the filter by incubating the filter in 10 ml TTS-HCl (pH 8.0), 0.2 mL EDTA for 2 hr at 65°C. Bone marrow macrophages were prepared as described (Celada et al., 1988a).

Gel Electrophoresis DNA Binding Assay

For binding reactions, the PU.1 double-stranded oligonucleotide was
cloned into the BamHI site of the vector PGEM 1. This plasmid was digested with HindIII, labeled by filling in the ends with α-32P-labeled dNTPs and Klenow polymerase, and digested with EcoRI, and a 20 bp fragment was isolated from a 6% polyacrylamide gel. DNA binding reactions were carried out for 15 min at room temperature in a final volume of 20 μl containing 10,000 cpm of end-labeled P1 probe (specific activity 2 x 107 cpm/μg) and 5% glycerol, using 2 μg of poly(dI-dC) (Pharmacia) and indicated amounts of competitor double-stranded oligonucleotides. Samples were separated on 6% polyacrylamide gels and fluorographed; the RNA negatives were isolated using a 6% polyacrylamide gel in 1× TBE buffer. Gel bands were excised from Whatman 3MM paper and exposed to Kodak XAR film at room temperature using an intensifying screen.

Methylation Interference Analysis

The methylation interference analysis was performed as described (Celada et al., 1988). The P1 oligonucleotide was end-labeled using α-32P-pATP and T4 polynucleotide kinase and partially methylated with DMS (final concentration 0.05 M) for 2 min at 20°C. Preparative gel retardation experiments were carried out using the L25S1 fusion protein, followed by autoradiography. The bound and free bands were isolated by electrophoresis into an SSC concentrator (Iatron, Tokyo, Japan), ethanol precipitated, and cleaved with pipernoline for 30 min at 37°C. The products were analyzed on a 20% acrylamide-8 M urea sequencing gel and exposed to XAR film at 70°C.

DNAase I Protection Analysis

DNAase I protection analysis was performed as previously described (Celada et al., 1988). For this analysis we cloned a P1 oligonucleotide that had been isolated from the vector PGEM1 (Promega Biotech), end-labeled with Klenow polymerase. The purified DNA fragment (20,000 cpm) was incubated for 60 min at room temperature in 40 μl of the same buffer as described above for gel retardation analysis and with the indicated amounts of protein. Nuclear extracts from F344 D1 were prepared as described (Celada et al., 1988). DNAase I (1 mg/ml; Worthington Grade) was diluted 20-fold in dH2O and incubated at 4°C for 45 to 60 min prior to use in digestion. Diluted DNAase I (4 μl) was added to binding reactions for 20 min, followed by addition of an equal volume of 2× stop buffer (1% SDS, 20 mM EDTA, 40 mM HEPEs, 100 mM NaCl, 0.5 mM EDTA) to the reaction mixture. The samples were phenol-chloroform extracted and ethanol precipitated, and the analysis of the resulting products was carried out on an 8% polyacrylamide-8 M urea sequencing gel.

In vitro Transcription/Translation of PL1 cDNA

To map the DNA binding region of the PL1 protein, in vitro transcribed RNA was made as described by the manufacturer (Promega) from deletion clones created by successive digests of the KS-plasmid containing the 25.1 insert. RNA was transcribed using T3 RNA polymerase for all constructs except for the NciI construct, which used T7 RNA polymerase. The RNA templates (~2 μg) were used as RNA in a rabbit reticulocyte lysate (Promega) to transcribe into vitro (SP6)polythio-nucleotide-labeled protein. Protein production was confirmed by SDS-polyacrylamide gel electrophoresis and autoradiography. Gel retardation analysis with the end-labeled P1 oligonucleotide was performed directly with 1 μl of the translation mix as described above. The gels were dried and exposed to two sheets of Kodak XAR film to distinguish the PL1 label from the 25.1 label.

The 3′ deletions were generated by the digestion with the indicated restriction enzymes (XbaI, NarI, BsHII, or NcoI). The 5′ deletion was generated by isolating the NciI fragment, filling in the ends with Klenow polymerase, and cloning the fragment into the EcoRI site of the KS-plasmid. The internal deletion clone NcoI–KpnI was created by digesting the plasmid p25.1 with NcoI and KpnI, followed by treatment with the DNA made with mung bean nucleases and religating.

PU1 Transfection Assays

The PU1 expression vector, PU1 pOE, was constructed by ligating the 25.1 CDNA clone into the EcoRI site of the expression vector, pOECE (Ellis et al., 1988). The CAT reporter constructions were made as follows: PU (t+), PU mut 1 (t+), and PU mut 2 (t+) were made by ligating the double-stranded oligonucleotides, PU1-U or PU1 mut 1, into the BamHI site of the CAT vector pBR322 (Lucas and Schultz, 1987). The PU (t+) was made by digesting the PU1 pOECE (see above) with NciI and XbaI and directionally ligating this fragment into SalI-digested and Klenow-treated XbaI-digested pBR322. XbaI-digested pBR322: PU (t-) and PU (t+L) were made by digesting dimers of the PU1 oligonucleotide in both orientations in the BamHI site of the vector pBR322 into the XbaI site of pBR322. The 3′ insertions were made by digesting PU (t-) with XbaI, isolating the tetramer of PU1, Hierow blunt ending, and ligation into the XbaI site of pBR322. HeLa cells were transfected by the method of lipofection, according to the recommendations of the manufacturer (Bethesda Research Laboratories, Gaithersburg, MD) with two modifications: the lipofection reagent was allowed to equilibrate in DMEM for 10 min prior to use, and the cells were incubated for 8–10 hr with the lipofection–DNA complex. These two modifications resulted in a higher level of transfection in HeLa cells. Cotransfection of the backbone 1-1.2 μg of reporter vector, 3 μg of PU pOECE expression vector, and 3 μg of pCH110, a β-galactosidase expression vector used to measure transfection efficiency (Pharmacia), and KS-Bluescript, to a total of 20 μg of DNA. In some experiments the PU pOECE construction was replaced by the recombination receptor expression vector RAR pOECE (Graupner et al., 1989). CAT assays were performed using a standard protocol (Gorman et al., 1985).

Briefly, cells were isolated after 40–48 hr and subjected to these freeze-thaw cycles in dry ice and ethanol at 3°C. The extract (1–60 μl) was incubated with [3H]Toluene and acetyl coenzyme A for 60–120 min at 37°C, followed by extraction with ethyl acetate. The ethyl acetate was dried down and the sample resuspended in 20 μl of ethyl acetate for thin layer chromatography.

Acknowledgments

We thank John Knight for the synthesis of the oligonucleotides and David Chiller and Robert Collins for technical assistance. We would also like to thank William Roeder and Russell Doolittle for helpful discussions. Elizabeth Mathes, Wanda Reynolds, Robert Oshima, and Magnus Plath for critically reading the manuscript, and Kathleen Sweeting and Janie Burchartz for preparation of the manuscript. We thank Mallinckrodt, Inc., St. Louis, Missouri, for supplying us with acrylamide, bisacrylamide, and calcium chloride.

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

Received January 18, 1990.

References

Cloning of the PIU1 Gene


GenBank Accession Number

The accession number for the sequence reported in this paper is M32270.