The Proto-oncogene HLF and the Related Basic Leucine Zipper Protein TEF Display Highly Similar DNA-Binding and Transcriptional Regulatory Properties

By Stephen P. Hunger, Shaoxing Li, Majilinde Z. Fall, Louie Naumovski, and Michael L. Cleary

Genes encoding transcription factors are frequently altered by chromosomal translocations in acute lymphoblastic leukemia (ALL), suggesting that aberrant transcriptional regulation plays a prominent role in leukemogenesis. E2A-hepatic leukemia factor (HLF), a chimeric transcription factor created by the t(17;19), consists of the amino terminal portion of E2A proteins, including two experimentally defined transcriptional activation domains (TADs), fused to the HLF DNA binding and protein dimerization basic leucine zipper (bZIP) domain. To understand the mechanisms by which E2A-HLF induces leukemia and the crucial functions contributed by each constituent of the chimera, it is essential to define the normal transcriptional regulatory properties of HLF and related bZIP proteins.

ABERRANT TRANSCRIPTIONAL regulation is important in the pathogenesis of human cancer. In the leukemias, genes encoding transcription factors are frequently affected by chromosomal translocations. One category of translocations leads to creation of fusion genes that encode chimeric proteins with novel structural and functional features not possessed by the constituent wild type proteins. In B-precursor acute lymphoblastic leukemia (ALL), the E2A gene, which encodes the basic helix loop helix (bHLH) transcription factors E12 and E47, is interrupted by two different translocations that create distinct chimeric transcription factors (Fig 1).2-6 As a result of the t(1;19)(q23;p13), the E2A carboxy terminus, including the bHLH DNA binding and protein dimerization domains, is replaced with a region of PBX1 containing a heterologous DNA binding domain.3,4 Following the t(17;19)(q21-22;p13), the same general portion of E2A proteins is fused to the basic leucine zipper (bZIP) DNA binding and protein dimerization domains of hepatic leukemia factor (HLF).5,6 The portion of E2A proteins included in these chimeras includes two separate experimentally defined transcriptional activation domains (TADs).7,9 Based on these structural features, it has been predicted that E2A-PBX1 and E2A-HLF contribute to leukemogenesis by aberrantly activating transcription of target genes that contain binding sites recognized by PBX1 or HLF (reviewed in Hunger10). Due to the substantial overlap in sites bound by related transcription factors, it is not certain whether the important E2A-PBX1 and E2A-HLF target genes are normally regulated by wild type PBX1 and HLF, or by other proteins with similar DNA binding properties. As E2A-PBX1 and E2A-HLF bind to completely different DNA sequences in vitro,11-13 it is assumed that the target genes affected by these two protein chimeras are distinct and nonoverlapping. However, because bona fide E2A-PBX1 and E2A-HLF target genes have not yet been identified, this assumption has not been confirmed experimentally.

Interestingly, wild type PBX1 and HLF are not present in normal lymphoid cells, as their genes are transcriptionally silent in this lineage.14-16 Thus, it is important to consider whether fusion to E2A is essential for leukemogenesis, or if ectopic expression of wild type PBX1 or HLF might also be transforming. This question has been addressed experimentally for PBX1 (which does not function as a transcriptional activator by itself11-13), and it is clear that fusion to E2A is absolutely required for transformation.16 Surprisingly, the PBX1 homeodomain, which mediates site-specific DNA binding, is not required for transformation in experimental systems.18 This does not necessarily mean that DNA-binding is a dispensable function, as E2A-PBX1 and PBX proteins appear to function in a heteromeric complex with other HOX proteins.17 Thus, the homeodomain-deleted E2A-PBX1 may be tethered to DNA targets via retained protein-protein interactions. E2A-HLF has also been shown to be capable of transforming NIH 3T3 fibroblasts, but the potential oncogenic properties of wild type HLF have not been tested.18

From the Section of Pediatric Hematology/Oncology, Department of Pediatrics, University of Colorado School of Medicine, Denver, CO; the Department of Pediatrics, Division of Hematology/Oncology, Stanford University School of Medicine, Stanford; and Laboratory of Experimental Oncology, Department of Pathology, Stanford University School of Medicine, Stanford, CA.

Submitted October 13, 1995; accepted January 26, 1996.

Supported in part by a grant from the National Institutes of Health (Grant No. CA42971). S.P.H. was supported by a National Research Service Award (F32 CA 09184) from the National Cancer Institute. L.N. is the recipient of a BLOOD/ASH Scholar Award, L.M.C. is a Scholar of the Leukemia Society of America.

Address reprint requests to Stephen P. Hunger, MD, UCHSC Campus Box C229, 4200 E Ninth Ave, Denver, CO 80262.

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

© 1996 by The American Society of Hematology.
Transformation requires the integrity of the HLF leucine zipper (that mediates protein dimerization, an essential prerequisite for DNA binding) and both E2A TADs, suggesting that the ability to activate transcription of target genes is necessary for E2A-HLF-induced transformation. We have shown previously that wild type HLF activates transcription of reporter genes containing concatamerized HLF binding sites in several different cell types, including human ALL cell lines phenotypically similar to those in which the t(17;19) occurs. If activating transcription of target genes that contain binding sites recognized by HLF is the sole requirement for transformation, wild type HLF, and other proteins with similar transcriptional regulatory properties, might also be able to function as oncogenic transcription factors.

HLF is a member of a small subfamily of bZIP proteins that also includes D binding protein (DBP), and thyrotroph embryonic factor (TEF) and its chicken homologue vitellogenin binding protein (VBP). These three proteins share substantial homology within the bZIP domains and within an adjacent domain rich in proline and acidic amino acids (the PAR). The three PAR proteins bind to an overlapping set of DNA target sequences as homodimers, or when heterodimerized with other PAR, but not other non-PAR, bZIP proteins. We and others have identified an HLF consensus DNA binding site, 5'-GTTACGTAAT-3', which E2A-HLF induces leukemia and the crucial functions contributed by each constituent of the chimera, we investigated the normal transcriptional regulatory properties of wild type HLF and TEF. Our studies show that these two proteins have virtually indistinguishable in vitro DNA-binding and transcriptional regulatory properties. Furthermore, the transcriptional activation potential of both proteins is contained within a highly conserved domain that we term the THAD (TEF/HLF Activation Domain). One fundamental consequence of the t(17;19) is replacement of the HLF THAD with heterologous E2A TADs, raising the possibility that alterations in protein-protein interactions between transcriptional activators and components of the general transcription machinery may contribute to leukemogenesis.

**MATERIALS AND METHODS**

**Cloning of human TEF cDNAs.** A cDNA library derived from the human B-precursor ALL cell line HB1119 was screened at reduced stringency (final wash 2× SSC [1× SSC is 0.15 mol/L NaCl with 0.015 mol/L sodium citrate] and 0.1% sodium dodecyl sulfate at 52°C) with a polymerase chain reaction (PCR)-generated cDNA probe, which consisted of the entire open reading frame (ORF) of HLF. Four independent plaques were purified to homogeneity, and the phase inserts were characterized by restriction enzyme mapping, then subcloned into the Bluescript plasmid vector (Stratagene, La Jolla, CA) for nucleotide sequence analysis using commercially prepared reagents (USB, Cleveland, OH).

**Electrophoretic mobility shift assays (EMSA) and consensus binding site selection.** EMSA were performed as described previously with proteins translated in vitro (IVT) from expression plasmids using a coupled reticulocyte lysate system (Promega, Madison, WI). The oligonucleotides used in EMSA contained the binding site of interest in an identical backbone derived from the VBP site in the chicken VTG II promoter. Binding sites used included the HLF/TEF consensus 5'-GTTACGTAAT-3', and (this report), a TEF binding site (5'-GTTACCGAAG-3') present in the rat growth hormone (GH) promoter, a VBP binding site (5'-TTTATGAAAA-3') present in the chicken VTG II promoter, and a consensus binding site (5'-CTGACGTCA-3') for ATF proteins. Single-stranded oligonucleotides were annealed, labeled by Klenow fill-in, and unincorporated nucleotides were removed by passing the resultant double-stranded oligonucleotide through a NucTrap column (Stratagene).

To determine the consensus TEF DNA binding site, a modified version of the selective amplification and binding (SAAB) procedure was performed exactly as described previously. After 5 rounds of SAAB, the amplified product was cloned into a pSP65 vector (Promega) to make pSP65-TEF. For transient transfections, the same portion of the TEF cDNA was cloned into the pCMV expression vector; pCMV-HLF has been described previously. To define the regions of TEF and HLF capable of activating transcription, various portions of the TEF and HLF cDNAs were fused in-frame to a cDNA encoding amino acids 1-147 of GAL4 in vator in cells of hepatic origin, but not in other cell types. TEF activates transcription in monkey kidney-derived CV-1 cells, and VBP has been shown to be a transactivator in chicken fibroblast- and hepatoma-derived cell lines. A cDNA library derived from the human B-precursor ALL cell line HB1119 was screened at reduced stringency (final wash 2× SSC [1× SSC is 0.15 mol/L NaCl with 0.015 mol/L sodium citrate] and 0.1% sodium dodecyl sulfate at 52°C) with a polymerase chain reaction (PCR)-generated cDNA probe, which consisted of the entire open reading frame (ORF) of HLF. Four independent plaques were purified to homogeneity, and the phase inserts were characterized by restriction enzyme mapping, then subcloned into the Bluescript plasmid vector (Stratagene, La Jolla, CA) for nucleotide sequence analysis using commercially prepared reagents (USB, Cleveland, OH).

**Electrophoretic mobility shift assays (EMSA) and consensus binding site selection.** EMSA were performed as described previously with proteins translated in vitro (IVT) from expression plasmids using a coupled reticulocyte lysate system (Promega, Madison, WI). The oligonucleotides used in EMSA contained the binding site of interest in an identical backbone derived from the VBP site in the chicken VTG II promoter. Binding sites used included the HLF/TEF consensus 5'-GTTACGTAAT-3', and (this report), a TEF binding site (5'-GTTACCGAAG-3') present in the rat growth hormone (GH) promoter, a VBP binding site (5'-TTTATGAAAA-3') present in the chicken VTG II promoter, and a consensus binding site (5'-CTGACGTCA-3') for ATF proteins. Single-stranded oligonucleotides were annealed, labeled by Klenow fill-in, and unincorporated nucleotides were removed by passing the resultant double-stranded oligonucleotide through a NucTrap column (Stratagene).

To determine the consensus TEF DNA binding site, a modified version of the selective amplification and binding (SAAB) procedure was performed exactly as described previously. After 5 rounds of SAAB, the amplified product was cloned into a pSP65 vector (Promega) to make pSP65-TEF. For transient transfections, the same portion of the TEF cDNA was cloned into the pCMV expression vector; pCMV-HLF has been described previously. To define the regions of TEF and HLF capable of activating transcription, various portions of the TEF and HLF cDNAs were fused in-frame to a cDNA encoding amino acids 1-147 of GAL4.
or by PCR amplification. The resultant plasmids express fusion proteins that consist of amino acids 1-147 of GAL4 (which allow binding to cognate GAL4 DNA binding sites and also contain a nuclear localization signal), from 3-15 amino acids encoded by the plasmid polylinker and varying portions of TEF or HLF. The nomenclature of constructs indicates the portion of TEF or HLF fused to GAL4 (1-147); for example, GAL4-TEF(-12---303) contains the entire TEF open reading frame including 12 residues amino terminal to the first methionine codon. Based on the results of transcriptional assays with the GAL4 chimeras, TEF and HLF expression vectors that lacked the putative TAD were constructed by using PCR to add restriction sites for deletion and cloning. To detect both full-length and deleted TEF and HLF proteins in transfected cells, pCMV-TEFΔ4-AP-CAT, pCMV-TEFΔ74-114ΔGAL4, pCMV-HLFΔ59-103ΔGAL4 were constructed; each contains the heterologous FLAG epitope (amino acid sequence DYKDDDDK) added to the carboxy terminus of the protein. For experiments performed in yeast, the plasmid pBTM116 (constructed originally by Paul Bartel and Stan Fields, Department of Microbiology, State University of New York, Stony Brook) was used.

The plasmid carries the TRP1 gene that allows selection in media lacking tryptophan and encodes full-length lexA followed by several unique restriction endonuclease recognition sites. Wild type and deleted versions of TEF and HLF were cloned in-frame to the lexA sequence of pBTM116 to generate lexA-TEF, lexA-TEFΔ74-114, lexA-HLF, and lexA-HLFΔ59-103. The integrity of all constructions was verified by restriction enzyme mapping and the nucleotide sequence of portions generated by PCR was determined to ensure that no mutations had been introduced.

Transient transfections and transcriptional activation assays. Various reporter constructs were used that contain concatenated DNA recognition elements adjacent to one of several promoters that drive expression of a chloramphenicol acetyl transferase (CAT) gene. The reporters referred to as (HLF/TEF)4-AP-CAT and (HLF/TEF)4-AP-TEFΔ1--154-AP-CAT have been described previously as HLF4-TK-CAT and HLF4-TEFΔ1--154-TEFΔ1--154-AP-CAT have been described previously as HLF4-TK-CAT and HLF4-TEFΔ1--154-TEFΔ1--154-AP-CAT. These contain four tandem repeats of the consensus binding site 5'-GTTACGTAAT-3' (contained within the VBP virus thymidine kinase (TK) promoter)

Diego, CA). β-galactosidase activity was determined by measuring the optical density at 420 nm as described. CAT activities were normalized based on the relative transfection efficiencies as determined by the luciferase or β-galactosidase assays. Each experiment was performed in duplicate on at least three separate occasions. Values are expressed as the arithmetic mean ± the standard deviation.

Immunodetection of transfected proteins. For Western (immunoblot) analyses, cells were transfected in 60 mm plates exactly as previously described and harvested 48 to 72 hours following transfection. The cell pellet from each plate was lysed in protein loading buffer, boiled for 5 minutes and separated by SDS-polyacrylamide gel electrophoresis. Following electrophoresis, proteins were transferred onto polyvinylidene difluoride (PVDF) membranes (Millipore, Bedford, MA) by semidry electrophoretic transfer. Proteins were detected using commercially obtained monoclonal antibodies as recommended by the supplier, and visualized by chemiluminescence (Amersham, Buckinghamshire, UK). The M2 anti-FLAG antibody (IBI, New Haven, CT) was used with the FLAG-tagged constructs and the GAL4 chimeras were detected with a mouse monoclonal antibody raised against GAL4 amino acids 1-147 (Santa Cruz Biotechnology, Santa Cruz, CA).

To determine the subcellular localization of transfected proteins, 4 × 10^6 NIH 3T3 cells were plated into each well of 2-well chamber slides and allowed to grow for approximately 18 hours. Cells were transfected by the calcium phosphate precipitation method with 5 μg of each plasmid construct. Approximately 20 hours following transfection, the cells were washed with PBS and fresh media added. The following day the cells were washed with PBS, fixed with 2% paraformaldehyde in PBS for 15 minutes, rinsed with PBS, permeabilized with 0.5% Triton X-100 in PBS for 30 minutes and blocked with a solution of 2% bovine serum albumin (BSA) in PBS (BSA/PBS) for 5 minutes. The anti-FLAG antibody was added at a dilution of 1:200 in PBS/BSA and allowed to incubate at room temperature for 2 hours. Following incubation, the slides were washed for 5 minutes in PBS and then incubated with a biotinylated goat antimouse antibody (1:500 in PBS/BSA for 1 hour). The washing step was repeated and the slides were incubated with streptavidin-horseradish peroxidase (1:500 in PBS/BSA for 1 hour). The slides were washed again, developed with 3-amino-9-ethylcarbazole, counterstained with aqueous hematoxylin and mounted with Crystal Mount (Biomeda Corp, Foster City, CA).

Yeast transformations and β-galactosidase filter assays. Yeast transformations were performed in the Saccharomyces cerevisiae strain L40, a generous gift from Stan Hollenberg (Vollum Institute, Oregon Health Sciences University, Portland), using lithium acetate and polyethylene glycol basically as described by Gietz et al. This strain contains two reporter genes, lacZ and HIS3, whose expression is driven by minimal GALI and HIS3 promoters, respectively, fused to multimerized lexA binding sites. Yeast transformed with the lexA-encoding plasmid pBTM116 cannot grow well on media lacking histidine or produce significant amounts of β-galactosidase. If a TAD capable of functioning in yeast is fused to lexA, the lexA chima can then grow on histidine-negative media and produce β-galactosidase.

β-galactosidase activity was assayed on nitrocellulose filters after transfer of yeast colonies from plates. The filters were quickly frozen...
by immersion in liquid nitrogen and, after thawing, placed in a petri dish containing a filter paper soaked in 1.5 mL of Z buffer (60 mmol/L Na2HPO4, 40 mmol/L NaH2PO4, 10 mmol/L KCl, 1 mmol/L MgSO4, pH 7.0) containing 15 µL of X-gal (5-bromo-4-chloro-3-indolyl-β-D-galactosidase; GIBCO-BRL, Gaithersburg, MD). The plates were covered, incubated in a humidified chamber at 30°C for several hours to overnight and inspected visually for evidence of blue color change.

RESULTS

Cloning and expression pattern of human TEF cDNAs.

To directly compare the DNA-binding and transcriptional regulatory properties of TEF to those of HLF, we set out to clone the cDNA encoding the human homologue of TEF/VPB. A cDNA library derived from the ALL cell line HB111928 was screened at reduced stringency with an HLF cDNA probe, and four independent clones were isolated that contained cDNA inserts ranging in size from approximately 3.3 to 4.5 kb. The longest cDNA contained an ORF of 945 nucleotides starting at its 5’ end and included approximately 3.3 kb of 3’ untranslated sequence followed by a poly-A tail (Fig 2A). The other three cDNAs contained shorter versions of the same sequence; no evidence of alternative splicing within the coding region was observed. An ATG codon beginning at nucleotide 37 is in a favorable context for initiation of translation.29 Data base searches indicated that the 303 amino acid protein encoded by these cDNAs (GenBank accession number U44059) was the human homologue of TEF/VPB and will be referred to herein as TEF.20,21

Chicken VPB (both the α and β isoforms23) diverges significantly from human TEF at the amino terminus (Fig 2B), but the remainder of the proteins share 88% amino acid identity. The rat TEF cDNA originally reported by Drolet et al28 encodes a protein that is 99% identical to human TEF, but includes 42 fewer amino-terminal residues. Similarly, Khatib et al38 recently reported isolation of a human TEF cDNA that begins at nucleotide 113 of that described herein, conceptual translation of this cDNA also predicted a protein lacking the initial 42 amino acids shown in Fig 1B. These two human TEF cDNAs are otherwise identical except for two minor differences: the cDNA reported by Khatib et al contains a G, rather than an A, at nucleotide 196 (predicting Glu41 rather than Lys41), and contains a T, rather than a C, at nucleotide 498. The difference in the predicted amino terminal sequence of TEF results from conceptual translation beginning at an ATG codon corresponding to the second present in the human TEF cDNA we isolated. Several observations suggest that the rat TEF cDNA and the human TEF cDNA described by Khatib39 are not full length. First, 31 nucleotides are present in the rat cDNA 5’ of the first ATG codon; 28 of these are identical to those in the corresponding position of human TEF, and they specify the same 10 amino acids. In addition, when these cDNAs are in vitro translated side-by-side, the human TEF cDNA we isolated encodes a protein 5 to 7 kD larger than the rat cDNA, consistent with translation beginning at the first ATG codon (data not shown). We cannot rule out the possibility that human TEF includes additional amino terminal residues, as the longest clone isolated does not contain a stop codon 5’ of the first ATG.

By Northern analysis the TEF transcript is a single band measuring approximately 4.5 kb (data not shown). Consistent with previous observations,38 TEF mRNA was found to be expressed at modest levels in each of several different human ALL cell lines tested and in the HepG2 hepatocellular carcinoma cell line (data not shown). This pattern of expression is distinct from that of HLF, which is also expressed in HepG2, but is transcriptionally silent in all hematolymphoid cell lines tested.4,6

TEF and HLF bind preferentially to the same core consensus DNA sequence in vitro. To identify the TEF consensus DNA binding site, a modified version of the SAAB procedure28 was performed using IVT TEF and a radiolabeled double-stranded oligonucleotide that contained 20 central

---

Fig 2. Sequence of human TEF. (A) Nucleotide sequence of the open reading frame of the longest TEF cDNA clone. The first in-frame ATG and stop codons are underlined. (B) The predicted amino acid sequence (single letter abbreviation) of human TEF (hTEF) starting from the first methionine codon is compared with that of the homologues rat TEP (rTEF) and chicken VBP (cVBP).
HLF AND TEF CONTAIN HOMOLOGOUS TADs

A

<table>
<thead>
<tr>
<th>position</th>
<th>-7</th>
<th>-6</th>
<th>-5</th>
<th>-4</th>
<th>-3</th>
<th>-2</th>
<th>-1</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td>% G</td>
<td>25</td>
<td>42</td>
<td>74</td>
<td>0</td>
<td>11</td>
<td>21</td>
<td>74</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>17</td>
<td>2</td>
<td>0</td>
<td>11</td>
<td>26</td>
</tr>
<tr>
<td>% A</td>
<td>19</td>
<td>32</td>
<td>26</td>
<td>0</td>
<td>0</td>
<td>75</td>
<td>0</td>
<td>26</td>
<td>0</td>
<td>95</td>
<td>106</td>
<td>0</td>
<td>72</td>
<td>22</td>
<td>22</td>
</tr>
<tr>
<td>% T</td>
<td>25</td>
<td>10</td>
<td>100</td>
<td>89</td>
<td>30</td>
<td>0</td>
<td>74</td>
<td>0</td>
<td>0</td>
<td>5</td>
<td>74</td>
<td>11</td>
<td>17</td>
<td>17</td>
<td>17</td>
</tr>
<tr>
<td>% C</td>
<td>31</td>
<td>16</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>18</td>
<td>0</td>
<td>26</td>
<td>5</td>
<td>5</td>
<td>16</td>
<td>55</td>
<td>33</td>
<td>33</td>
</tr>
</tbody>
</table>

**Fig 3. Selection of a high affinity TEF consensus DNA binding site.**

(A) A modified SAAB procedure was used to identify a consensus TEF binding site. The sequences of individual clones (n = 19) bound by TEF following five rounds of selection from a pool of random oligomers were visually aligned to derive the consensus. The percentage of clones that contain each nucleotide at a particular position is indicated in tabular form, and the consensus recognition site is listed below. (B) EMSA was performed using radiolabeled double-stranded oligonucleotides containing the consensus sequence and IVT protein. In the absence of competitor, a shifted complex is seen with the TEF IVT, but not with the negative control IVT programmed with vector alone. The specificity of binding was assessed by preincubation with 100-fold molar excess of unlabeled oligonucleotides containing the consensus or related sites. Binding is eliminated by competition with either the consensus site or a TEF site (5'-GTTACGTAAT-3') present in the rat GH gene promoter. A cVBP binding site from the VTG II gene promoter (5'-TTATACTGAA-3') competes partially, and little or no competition is observed with a canonical ATF site (5'-CGACAGTCAG-3').

For personal use only.on October 31, 2017. By guest. From www.bloodjournal.org by guest on October 31, 2017. For personal use only.
Cat reporters that contained either an adenovirus Elb or TEF, respectively, when this domain was discreetly deleted (constructs 3, 4, 6, and 7, yet the transcriptional potency of the first two differed markedly from the latter two (Fig 6B and C). We conclude from these experiments that, in the context of fusion to the heterologous GAL4 DNA binding domain, TEF amino acids 72-114 are absolutely required for transcriptional activation and contain most, if not all, of the transcriptional activation potential present in the full-length protein.

To determine whether the homologous region of HLF also functioned as a TAD, a more limited series of GAL4-HLF constructs was generated, and the ability of GAL4-HLF fusion proteins to activate transcription of the GAL4-E1b-CAT reporter was assessed following transient transfection into CV-1 cells (Fig 6D). GAL4-HLF activated transcription approximately 25-fold over background levels observed with GAL4 1-147 alone. Deletion analysis showed that HLF amino acids 60-102, the region of TEF homology, functioned as a discreet TAD when fused to GAL4 1-147 and, similar to TEF amino acids 72-116, contained most, if not all, of the transcriptional activation potential present in the full-length protein. Experiments performed in Cos-7 cells confirmed these observations and shown that similar amounts of each of the GAL4-HLF chimeric proteins were produced (data not shown).

Deletion of the conserved domain completely abrogates transcriptional activation by native TEF and HLF. To assess the function of this domain in the context of the native proteins, TEF and HLF expression vectors were constructed

<table>
<thead>
<tr>
<th>Fold Activation</th>
</tr>
</thead>
<tbody>
<tr>
<td>pCMV</td>
</tr>
<tr>
<td>pCMV-TEF</td>
</tr>
<tr>
<td>pCMV-HLF</td>
</tr>
</tbody>
</table>

Fig 4. TEF and HLF display similar transcriptional potencies. Histogram showing relative transcriptional potencies of TEF and HLF as assessed by transient transfections in REH cells with an (HLF/TEF)TK-CAT reporter. Values are expressed as the fold activation of CAT activity as compared with cotransfection of the reporter with the empty pCMV vector. The results shown are the arithmetic mean of duplicate determinations performed in three independent experiments with the standard deviation displayed as error bars.

To ensure that the observed differences in transcriptional properties were due to differences in the level of protein expressed, identical experiments were performed in Cos-7 cells, and the CAT activity of cell lysates was compared with the level of protein expressed as determined by immunoblot analysis with a monoclonal antibody raised against GAL4 amino acids 1-147 alone (Fig 6A). Similar results were observed using GAL4-TEF reporters that contained either an adenovirus E1b or HSV TK promoter. All constructs that contained TEF amino acids 72-116 (the region of HLF homology) activated transcription as much, or more, than the full length GAL4-TEF chimera. Constructs that lacked this portion of TEF were at or near background levels of CAT activity. Most instructively, when this domain was discreetly deleted (constructs 9 and 10), CAT activity was reduced to background levels. Some differences were observed in the magnitude of activation of the two reporter constructs; these are primarily a consequence of higher baseline CAT activity observed with GAL4-TK-CAT.

To asess the function of this domain in the context of the native proteins, TEF and HLF expression vectors were constructed

### Fold Activation

<table>
<thead>
<tr>
<th>Construct</th>
<th>Fold Activation</th>
</tr>
</thead>
<tbody>
<tr>
<td>pCMV</td>
<td>1</td>
</tr>
<tr>
<td>pCMV-TEF</td>
<td>2</td>
</tr>
<tr>
<td>pCMV-HLF</td>
<td>3</td>
</tr>
</tbody>
</table>

Fig 5. Homology between PAR proteins. (A) Domains of TEF are schematically demarcated. Amino acid residues are numbered above. In the lower portion, the percent amino acid identity between TEF and DBP is indicated for each domain. The cross-hatched region is the minimal TEF transcriptional activation domain identified in these studies. PAR, proline, and acidic amino acid rich region; BRE, basic region extension; bZIP, basic leucine zipper. (B) The amino acid sequence (single letter abbreviation) of the regions of TEF and HLF that comprise the THAD and are deleted in pCMV-TEF.A74-114 and pCMV-HLF.A59-103. The corresponding region of human DBP is also shown: gaps, indicated by dashes, are introduced to maintain alignment. Amino acid identity is denoted by dashes and similarity by pluses between the residues. Within this region TEF shares 72% amino acid identity and 85% similarity with HLF, 51% identity and 78% similarity with DBP, and HLF shares 56% identity and 77% similarity with DBP. A portion of the THAD predicted to assume an α-helical configuration is indicated.

<table>
<thead>
<tr>
<th>A</th>
<th>74</th>
<th>114</th>
<th>166</th>
<th>218</th>
<th>238</th>
<th>363</th>
</tr>
</thead>
<tbody>
<tr>
<td>TEF</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HLF</td>
<td>17%</td>
<td>72%</td>
<td>18%</td>
<td>67%</td>
<td>67%</td>
<td>78%</td>
</tr>
<tr>
<td>DBP</td>
<td>14%</td>
<td>51%</td>
<td>14%</td>
<td>67%</td>
<td>62%</td>
<td>62%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>B</th>
<th>94</th>
<th>132</th>
<th>74</th>
<th>114</th>
<th>14</th>
<th>59</th>
<th>103</th>
</tr>
</thead>
<tbody>
<tr>
<td>TEF</td>
<td>VPAQGILLWMLWKPFLPDQ---VETYQDAFLAKLHPLPPSF</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HLF</td>
<td>VSASLMPFVMDQITYGDEGFHYKGLSKFLKMDQIFAPSF</td>
<td>114</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

From www.bloodjournal.org by guest on October 31, 2017. For personal use only.
HLF AND TEF CONTAIN HOMOLOGOUS TADs

were fused in frame to the expression construct pSG424 that encodes GAL4 amino acids 1-147. The ability of GALCTEF chimeric proteins to activate transcription of the GAL4-Elb-CAT reporter following transient transfection into CV-1 cells. Results are expressed as the fold activation of CAT activity as compared with cotransfection of the reporter with pSG424 and are the arithmetic mean of duplicate determinations performed in three independent experiments. Standard deviations are displayed as error bars. No binding site-dependent baseline activation seen with the (HLF/TEF)_k-AP-CAT reporter. When the full-length and deleted constructs are directly compared, pCMV-TEFL^p^FLAG and pCMV-HLF\Delta79-103^p^FLAG were completely incapable of activating transcription of the reporter gene. In fact, these constructs acted as dominant negative inhibitors of transcriptional activation by endogenous proteins as they significantly reduced the binding of GAL4 chimeric proteins to activate transcription of the GAL4-Elb-CAT reporter. When the full-length and deleted constructs are compared, pCMV-TEF\Delta^p^FLAG is 27.6 ± 7.7-fold more active than pCMV-HLF\Delta59-103^p^FLAG and pCMV-TEF\Delta^p^FLAG is 13.0 ± 4.6-fold more active than pCMV-TEFL74-114^p^FLAG. Similar results were observed when identical experiments were performed with the (HLF/TEF)_k- TK-CAT reporter (data not shown). Western (immunoblot) analysis with an anti-FLAG monoclonal antibody showed that equivalent amounts of each of the protein constructs are present in transfected Cos-7 cells indicating that protein stability was not affected by the deletions (Fig 7B and C).

In summary these studies show that TEF and HLF each contain a highly homologous ~40 aa TAD that will be referred to subsequently as the THAD (Fig 5B). Data base searches indicate that the THAD is unique to the PAR subfamily of bZIP proteins. Structural analysis using the Chou-Fasman algorithm predicts that a central 16 amino acid portion of this domain assumes an a-helical configuration. There are no other obvious structural features present in the THAD; of 40 amino acids TEF has 5 prolines, 4 serines, and 1 threonine, while HLF possesses 5, 3 and 3 of the respective amino acids.

**THAD functions as a TAD in yeast.** It is currently believed that TADs effect their function by interacting, directly or indirectly, with protein components of the general transcription machinery. As many of these factors are highly conserved throughout evolution, we investigated whether HLF and TEF could activate transcription in yeast. Full-length and THAD-deleted HLF and TEF cDNAs were fused in frame to a cDNA encoding lexA in the plasmid pHBTM116. To ensure that the failure to activate transcription was not because of an alteration in protein subcellular localization, NIH-3T3 cells were transiently transfected and immunocytochemistry was performed with an anti-FLAG monoclonal antibody. Each of the full-length and deleted proteins was found to be located almost exclusively in the nucleus (Fig 8).

**Fig 6. Mapping of TEF and HLF transcriptional activation domains using GAL4 chimeric proteins.** (A) Various portions of the TEF cDNA were fused in frame to the expression construct pSG424 that encodes GAL4 amino acids 1-147. The ability of GAL4-TEF chimeric proteins to activate transcription of the CAT reporter genes driven by a minimal promoter derived from the adenovirus Elb or herpes simplex virus thymidine kinase gene with adjacent tandem GAL4 recognition elements was assessed following transient transfection into CV-1 cells. Results are expressed as the fold activation of CAT activity as compared with cotransfection of the reporter with pSG424 and are the arithmetic mean of duplicate determinations performed in three independent experiments. Standard deviations are displayed as error bars. No binding site-independent transactivation was observed when activators were cotransfected with CAT reporter constructs lacking GAL4 binding sites, only background levels of CAT activity (indistinguishable from those observed when pSG424 was cotransfected with GAL4-TK-CAT or GAL4-E1b-CAT) were observed. (B) Representative results of CAT assays assessing the ability of select GAL4-TEF chimeric proteins to activate transcription of the GAL4-E1b-CAT reporter following transient transfection into Cos-7 cells. Transfections were performed in duplicate, and one plate was harvested for CAT assays and the other for immunoblot analysis. In these experiments, CAT assays were allowed to go to completion to highlight the profound differences in transcriptional activation. The location of acetylated and nonacetylated products is indicated at the right. The constructs are numbered as in (A). (C) Immunoblot analysis of transfected Cos-7 lysates using a monoclonal antibody directed against GAL4 amino acids 1-147. The migration of molecular weight markers (in kilodaltons) is indicated at the left of the blot. The constructs are numbered as in (A). (D) The ability of GAL4-HLF chimeric proteins to activate transcription of GAL4-E1b-CAT was assessed following transient transfection into CV-1 cells. Results are expressed as the fold activation of CAT activity as compared with cotransfection of the reporter with pSG424 and are the arithmetic mean of duplicate determinations performed in three independent experiments. Standard deviations are displayed as error bars.

with and without this conserved domain (the amino acid sequence of portions of TEF and HLF deleted in these constructs is shown in Fig 5B). The heterologous FLAG epitope was added to the carboxy terminus of each construct so that protein expression could be directly compared. Transcriptional activation of an (HLF/TEF)_k-AP-CAT reporter was increased approximately fivefold over background levels seen with the pCMV vector alone by both pCMV-TEF^p^FLAG and pCMV-HLF\Delta79-103^p^FLAG (Fig 7A). The transcriptional potency of these constructs was indistinguishable from constructs lacking the FLAG epitope (data not shown). In contrast, pCMV-TEFL74-114^p^FLAG and pCMV-HLF\Delta59-103^p^FLAG were completely incapable of activating transcription of the reporter gene. In fact, these constructs acted as dominant negative inhibitors of transcriptional activation by endogenous proteins as they significantly reduced the binding of GAL4 chimeric proteins to activate transcription of the GAL4-Elb-CAT reporter. When the full-length and deleted constructs are directly compared, pCMV-TEF\Delta^p^FLAG is 27.6 ± 7.7-fold more active than pCMV-HLF\Delta59-103^p^FLAG and pCMV-TEF\Delta^p^FLAG is 13.0 ± 4.6-fold more active than pCMV-TEFL74-114^p^FLAG. Similar results were observed when identical experiments were performed with the (HLF/TEF)_k- TK-CAT reporter (data not shown). Western (immunoblot) analysis with an anti-FLAG monoclonal antibody showed that equivalent amounts of each of the protein constructs are present in transfected Cos-7 cells indicating that protein stability was not affected by the deletions (Fig 7B and C).

To ensure that the failure to activate transcription was not because of an alteration in protein subcellular localization, NIH-3T3 cells were transiently transfected and immunocytochemistry was performed with an anti-FLAG monoclonal antibody. Each of the full-length and deleted proteins was found to be located almost exclusively in the nucleus (Fig 8).

**Summary.** These studies show that TEF and HLF each contain a highly homologous ~40 aa TAD that will be referred to subsequently as the THAD (Fig 5B). Data base searches indicate that the THAD is unique to the PAR subfamily of bZIP proteins. Structural analysis using the Chou-Fasman algorithm predicts that a central 16 amino acid portion of this domain assumes an a-helical configuration. There are no other obvious structural features present in the THAD; of 40 amino acids TEF has 5 prolines, 4 serines, and 1 threonine, while HLF possesses 5, 3 and 3 of the respective amino acids.
HLFΔ59-103 and lexA-TEFΔ74-114 were incapable of growth (Table 1). Similarly, yeast transformed with lexA-HLF and lexA-TEF were strongly positive in the β-galactosidase filter assay, while those transformed with lexA-THADΔ59-103 and lexA-TEFΔ74-114 did not exhibit any color change, even after overnight incubation. These data indicate that the THAD functions as a TAD in yeast, as well as in mammalian cells, suggesting that it interacts with proteins that are conserved throughout millions of years of evolution.

**DISCUSSION**

In these studies, we have compared the DNA-binding and transcriptional regulatory properties of human TEF and HLF and found them to be strikingly similar. Both bound preferentially to the same core consensus DNA sequence in vitro and no differences were detected in binding to a panel of related sites. DBP, the third known PAR-bZIP protein, binds to similar target sites, but its optimal recognition element has not been defined. These findings are consistent with the fact that the three PAR proteins share extensive homology within the basic and adjoining regions known to mediate DNA binding. In transient transfection assays, TEF and HLF activated transcription of consensus site-containing CAT reporter genes in each of several cell types (REH, CV-1, Cos-7, NIH 3T3) tested and their potencies were quite similar to one another within each cell type. These observations suggested that TEF and HLF might contain structurally
similar TADs. Previous studies had shown that the transactivation properties of both TEF and HLF were not contained within the highly conserved PAR or bZIP domains, but rather mapped to the amino terminal halves of each protein.14,20

This line of reasoning focused our attention on a region of ~40 amino acids within which TEF and HLF share 72% amino acid identity and 85% similarity. Using both GAL4 chimeras and native TEF and HLF proteins, we found that this domain, termed the THAD, contains most, if not all, of the transcriptional potency and is absolutely required for transcriptional activation by TEF and HLF. Because the transcriptional competency of subfragments of the THAD was not determined, it is possible that the actual TAD might be somewhat smaller than that defined in this report.

Based on these observations, one would anticipate that TEF and HLF would be capable of activating transcription of an overlapping set of target genes in vivo. One manner by which the function of these proteins might vary is via differences in their expression profiles. Indeed, HLF displays a very restricted pattern of mRNA expression in adult tissues, while TEF/VBP mRNA is expressed relatively ubiquitously in rat and chicken.5,9,19,20 Our results, combined with those of Khatib et al.,18 indicate that TEF mRNA expression differs in at least one important manner from HLF, as it is consistently expressed in human ALL cell lines, whereas HLF is transcriptionally silent in this lineage. Efforts to detect TEF protein expression in ALL cell lines have been unsuccessful, but it is uncertain whether this was due to true lack of expression or inability of the reagents employed to detect low levels of endogenous TEF.15

Our conclusion that the highly conserved THAD is both absolutely necessary and sufficient for transcriptional activation by TEF and HLF substantially refines earlier studies that localized transactivation properties to the amino terminal halves of these proteins.14,20 Subsequent to completion of the current investigations, Burch et al.19 reported studies that used GAL4 chimeric proteins to map the VBP TAD. They found that residues 79-106 (corresponding to human TEF residues 72-99 that comprise the first two thirds of the TAD defined herein) were absolutely required for transcriptional activation to occur. These findings are concordant with our observations. However, in contrast to our findings, full transcriptional activity of GAL4-VBP chimeric proteins required VBP residues 79-172, an intermediate construct containing residues 79-137 was approximately fivefold less potent, and residues 79-106 were transcriptionally inactive. There are a number of technical factors that might explain these differences including the use of different reporter constructs and different cell types. It is also possible that the TADs of these homologues differ from one another in some respects, as the amino acid identity between human TEF and residues 118-167 of chicken VBP is only 57% (see Fig 2B).

It is also important to consider what role the THAD might play in transcriptional activation by DBP, which contains a similarly located region that shares significant homology with both TEF and HLF (see Fig 5B). DBP differs from the other two PAR proteins by virtue of the fact that it activates transcription only in cells of hepatic origin and not in other cell types.19,24 In support of this, we have found that rat DBP does not transactivate the (HLF/TEF),-TK-CAT or (HLF/TEF),-AP-CAT reporters in CV-1 or REH cells (SPH and MLC, unpublished observations, May 1994). In future studies it will be important to determine if the region of DBP homologous to the THAD functions as a TAD. If so, these proteins would provide ideal reagents with which to study the mechanisms by which cell type specificity is conferred on a TAD.

The primary amino acid sequence of the THAD is unique to the PAR subfamily of bZIP proteins, and analysis of its amino acid composition and predicted secondary structure do not yield any obvious insight into its mode of function. A number of TADs of several different classes have now been described. Recent evidence indicates that transcription factors effect their function by interacting, directly or indirectly, with various protein components of the basal transcription apparatus that is essential for transcription of mRNA by RNA polymerase II (reviewed in Tijan 40). The primary interaction appears to occur between discrete residues of TADs and individual protein components of the basal transcription apparatus including (but not necessarily limited to) the TATA box binding protein (TBP) and certain of its associated factors (TAFs) within the TFIID complex, TFIIb and TFIIH.40,44-48 The fact that the THAD also functions as a TAD in yeast suggests that its target has been highly conserved throughout evolution and will facilitate efforts to understand the biochemical mechanisms by which TEF and HLF activate transcription.

These studies lay important groundwork for future investigations aimed at defining the essential functional properties that E2A and HLF contribute to the t(17; 19)-ALL-associated chimeras E2A-TEL. Deletion analyses showing that the HLF leucine zipper and both E2A TADs are absolutely required for transformation of NIH-3T3 cells support the hypothesis that E2A-TEL contributes to leukemogenesis by activating transcription of crucial target genes.18 If this is true, then one should be able to reconstitute an oncogenic transcription
factor by taking advantage of the modular nature of functional domains. In the context of this model, our observations suggest that HLF and TEF might also be oncogenic if expressed in the correct cellular context. If they are not, there must either be some fundamental (qualitative or quantitative) differences between the E2A TADs and the THAD, or E2A must contribute other crucial nontranscriptional regulatory properties to the leukemia-associated chimera. Indeed, we have reported previously that E2A-RLF1 appears to be a modestly more potent transactivator than HLF in transient transfection assays. The magnitude of these differences is not striking and it is unclear whether or not it is functionally significant. With respect to the latter possibility, it is important to note that E2A proteins possess growth-inhibitory properties and, when over-expressed in fibroblasts, can block cell cycle progression near the G1-S border. The fact that growth inhibition is not completely dependent on the integrity of the bHLH domains supports the possibility that E2A proteins might have important nontranscriptional regulatory properties.

One major unresolved question concerning oncogenic transcriptional activators is whether the nature of the TAD and the manner in which it interacts with the basal transcription apparatus are important for oncogenic activity. Are transforming properties retained if an oncoprotein’s TAD is replaced with a heterologous TAD (of equivalent potency) that interacts with the same target molecule? With one that contacts a different target molecule? Lessnick et al. have begun to address this question using the Ewings Sarcoma-derived EWS/FLI1 chimera. Similar to HLF and E2A-RLF1, both FLI-1 and EWS-FLI1 can activate transcription of reporter genes, and the chimera is more potent than wild type FLI-1. When over-expressed via retroviral-mediated gene transfer, full-length EWS/FLI1 efficiently transformed NIH 3T3 cells, whereas wild type and truncated FLI-1 did not. Transformation competency was retained when the EWS portion of the chimera was replaced with some, but not other, heterologous TADs suggesting that these domains are not fully interchangeable. To understand how aberrant transcriptional regulation contributes to malignant transformation, it is essential to perform similar studies with other oncogenic transcription factors. Delineation of the THAD is an important prerequisite for performing such investigations with E2A-RLF1 and HLF.

ACKNOWLEDGMENT

We thank John Burch for his gift of the VBP-CAT reporter, and Ivan Sadowski and Mark Pusheine for the SG424 construct. Grateful appreciation is expressed to Stan Hollenberg and Jackie Lee for providing reagents and assisting with the yeast transformation experiments. We also thank Bob Garcea and Jackie Lee for helpful comments on the manuscript.

REFERENCES

11. Van Dijk MA, Voorhoeve M, Murre C: Pbx1 is converted into a transcriptional activator upon acquiring the N-terminal region of E2A in pre-B-cell acute lymphoblastic leukemia. Proc Natl Acad Sci USA 90:6061, 1993
13. Lu Q, Wright DD, Kamps MP: Fusion with E2A converts the Pbx1 homeodomain protein into a constitutive transcriptional activator in human leukemias carrying the t(1;19). Mol Cell Biol 14:3938, 1994


23. Haas NB, Cantwell CA, Johnson PF, Burch JBE: DNA-binding specificity of the PAR basic leucine zipper VBP partially overlaps those of the C/EBP and CREB/ATF families and is influenced by domains that flank the core basic region. Mol Cell Biol 15:1923, 1995


41. Lavery DJ, Schibler U: Circadian transcription of the cholesterol 7a hydroxylase gene may involve the liver-enriched bZIP protein DBP. Genes Dev 7:1871, 1993


43. Goodrich JA, Hoey T, Thut CJ, Admon A, Tijan R: Drosophila TAF40 interacts with both a VP16 activation domain and the basal transcription factor TFIIB. Cell 75:519, 1993


The proto-oncogene HLF and the related basic leucine zipper protein TEF display highly similar DNA-binding and transcriptional regulatory properties

SP Hunger, S Li, MZ Fall, L Naumovski and ML Cleary