Nucleic Acids Research, 1995, Vol. 23, No. 19 **3865–3871**

TFE3 contains two activation domains, one acidic and the other proline-rich, that synergistically activate transcription

Steven E. Artandi[1,+], Kevin Merrell[3], Nicole Avitahl[1,§], Kwok-Kin Wong[3] and Kathryn Calame[1,2,3,*]

Departments of [1]Microbiology, [2]Biochemistry and Molecular Biophysics and [3]Integrated Program in Cellular, Molecular and Biophysical Studies, Columbia University College of Physicians and Surgeons, New York, NY 10032, USA

Received July 7, 1995; Revised and Accepted August 25, 1995

GenBank accession no. U36393

ABSTRACT

TFE3 is a basic-helix–loop–helix-zipper (bHLHZIP) domain-containing protein that binds μE3 sites in regulatory elements in the immunoglobulin heavy chain gene. The protein is a transcriptional activator that is expressed *in vivo* as two alternately spliced isoforms with different activating properties: TFE3L contains an N-terminal acidic activation domain; TFE3S lacks this activation domain and is a dominant negative inhibitor of TFE3L. We show that TFE3L and TFE3S contain a second, C-terminal activation domain rich in proline residues. This pro-rich activation domain has activity in a Gal4 fusion assay comparable to the N-terminal acidic activation domain present in TFE3L. The TFE3 pro-rich activation domain contains regions of strong homology with the related proteins microphthalmia and TFEB, suggesting that these regions are important for function. Using two different assays, we show that the N- and C-terminal activation domains of TFE3 act synergistically. This synergism explains in part the ability of TFE3S to act as a dominant negative. Our domain analysis of TFE3 is incorporated into a general structural model for the TFE3 protein that predicts that the activation domains of TFE3 will be widely separated in space.

INTRODUCTION

Gene transcription is a key regulatory point in diverse developmental processes such as cell growth, differentiation, apoptosis and transformation. The transcription factors that control the rates of transcription initiation are often tightly regulated and the activities of these proteins can be modified at a variety of levels. Eukaryotic transcription factors are generally bipartite; DNA-binding domains contact the DNA binding site and activation domains transduce regulatory information via protein–protein interactions with components of the basal transcription apparatus (1).

TFE3 is a member of the bHLHZIP family of transcription proteins. It binds the μE3 elements in the immunoglobulin heavy chain (IgH) intronic enhancer, in Ig kappa enhancers and in some IgH variable region promoters (2,3). TFE3 exists *in vivo* as two alternately spliced isoforms with different activation potentials (4). The alternately spliced exon encodes an N-terminal acidic activation domain (AAD) (2). The activation potential of TFE3S, the isoform lacking this activation domain, is ~4-fold lower than that of TFE3L, the isoform containing the N-terminal domain (4). The TFE3S mRNA *in vivo* is expressed at low levels compared to TFE3L, ranging from 2 to 18% of total TFE3 mRNA in a variety of tissues and cell lines.

Cotransfection assays with TFE3L and TFE3S showed that TFE3S can act as a dominant negative. Two aspects of these data were striking. First, the dominant negative effect of TFE3S on TFE3L activity was achieved with substoichiometric amounts of TFE3S expression plasmid. The correlation between the plasmid ratios at which TFE3S inhibited TFE3L activity and the ratio of these mRNA isoforms *in vivo* suggested that the dominant negative activity of TFE3S is biologically important. Second, even when TFE3S represented only 20% of the total amount of input TFE3 expression plasmid, the activity of the mixture was equivalent to that of 100% TFE3S rather than an average between the activities of TFE3S and TFE3L.

The mechanism by which TFE3S exerts this dominant negative effect on TFE3L is unclear. It was originally proposed that TFE3S might poison the activity of a tetramer of TFE3 proteins (4). Tetramerization of TFE3 (5) and other bHLHZIP proteins including c-Myc (6), TFEB (7) and USF (8) has been demonstrated, however, these proteins appear to bind DNA as dimers in electromobility shift assays. Tetramerization may be important in mediating interactions between remotely and proximally bound bHLHZIP proteins (5), however either dimers or tetramers of TFE3 could mediate simple activation from a proximal element such as the one used in the mixing experiments with TFE3S and TFE3L (4). We previously observed that TFE3 exhibits synergistic activation of transcription as the number of μE3 binding sites are increased in a promoter (5). We wondered if synergistic effects present in the TFE3L homodimer, but absent in the TFE3S

* To whom correspondence should be addressed

Present addresses: Departments of [+]Medicine and [§]Dermatology, Massachusetts General Hospital, Harvard Medical School, Boston, MA, USA

homodimer and the TFE3L–TFE3S heterodimer could contribute to the dominant negative activity of TFE3S.

To understand more completely the mechanism of dominant negative inhibition by TFE3S, we characterized the activation sequences of TFE3 and developed an assay to test how these domains might work together. In this report, we show that a C-terminal proline rich domain of TFE3 is an activation domain. Furthermore, the proline-rich domain activates synergistically with the acidic domain of TFE3, both when these domains are in their normal context in the TFE3 molecule and when they are tethered to different DNA binding domains bound at adjacent DNA elements. These data are incorporated into a model for TFE3 activation.

MATERIALS AND METHODS

Expression plasmids

GAL4-TFE3PRO was constructed by cloning a *Stu*I fragment from the TFE3 cDNA into the *Ecl*136II site of pSV-GAL(1–147) (9), resulting in fusion of TFE3 residues 345–446 [numbering scheme as in (3)] to GAL4(1–147). GAL4-TFE3AAD was previously described as GAL4λ3-Δ2 and fuses 126 N-terminal residues from TFE3 to GAL4(1–147) (2). GAL4-VP16 has been previously described (9).

pSV2-TFE3L and pSV2-TFE3S were described previously (4). TFE3L-ΔCterm and TFE3S-ΔCterm represent C-terminal truncations of TFE3L and TFE3S, respectively, terminating at Val-314, immediately after the leucine zipper [numbering as in (3)]. They were constructed by PCR using pBS-TFE3L and pBS-TFE3S as templates. PCR fragments were digested with *Hin*dIII and *Bam*HI and cloned into the pSV2A expression plasmid (10).

Reporter plasmids

pGAL4-TATALuciferase (gift of J. Kaplan) was generated by isolating a GAL4 site pentamer from G_5E4T (gift of M. Carey) by digestion with *Xba*I and *Hin*dIII. This fragment was cloned by blunt ended ligation into the *Bgl*II site of pTATA-Luciferase (5), a reporter plasmid containing a minimal TATA box upstream of the firefly luciferase gene.

pGAL4/μE3-TATALuciferase is a reporter plasmid containing a pentamer of GAL4 sites and a tetramer of μE3 sites upstream of the minimal promoter of pTATA-Luciferase. The GAL4 site pentamer was cloned into the *Hin*dIII site of a pKS Bluescript polylinker that contained a μE3 tetramer in the *Sma*I site. A cassette containing both multimers was excised by *Xba*I and *Bam*HI digestion and cloned into the *Bgl*II site of pTATA-Luciferase by blunt ended ligation.

Transfection

NIH 3T3 cells were transfected by the calcium phosphate technique essentially as described (5). Reporter plasmid (1 μg) and 5–10 μg expression plasmid were used per plate of cells. Cells were harvested and assayed for luciferase activity as previously described (5).

Western blot

Whole cell extracts from transfected fibroblasts were prepared. Cells were pelleted by centrifugation at 600 *g*, washed once in PBS, and lysed in 50 mM Tris pH 8.0, 0.5% NP-40, 10% glycerol, 0.1 mM EDTA and 400 mM NaCl. Fibroblasts from one 10 cm plate (4×10^6 cells) resuspended in 100 μl of lysis buffer and allowed to remain on ice for 30 min then frozen in liquid nitrogen. Nuclear extract (10 μl; 4×10^5 cell equivalents) was loaded in each well of a 10% SDS–polyacrylamide gel. The nuclear proteins were transfered to nitrocellulose and Western blotted by standard techniques using a polyclonal anti-TFE3 antiserum.

RESULTS

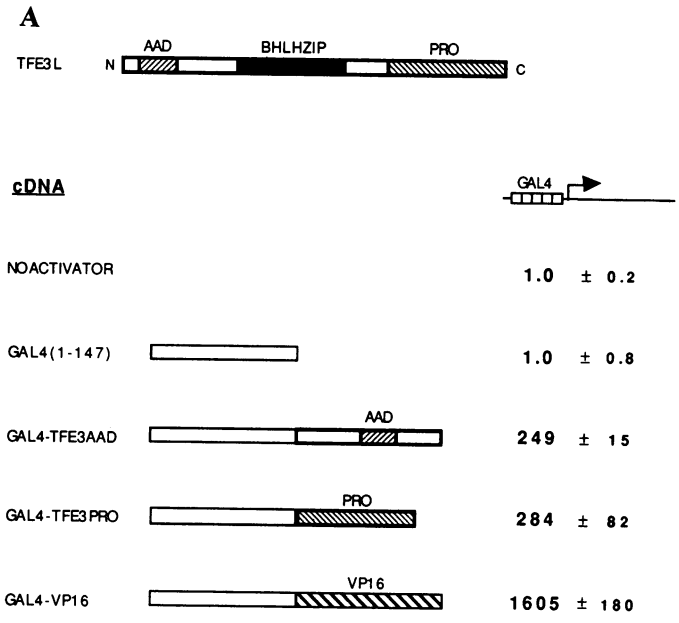
TFE3L contains a second transcriptional activation domain C-terminal to the bHLHZip region

The exon that is alternately spliced between isoforms TFE3L and TFE3S encodes an N-terminal activation domain that is rich in acidic amino acids. TFE3S, which lacks this domain, is a weaker transactivator but it does retain some transcriptional activating ability (4). The original GAL4 fusion analyses of TFE3L showed that the N-terminal acidic domain is a strong activation domain (2). However, in that study the remainder of the protein containing the bHLHZIP and C-terminal domains stimulated transcription weakly when fused to the GAL4 DNA binding domain (GAL4 DBD). These data led us to speculate that TFE3L might contain a second activation domain.

To test this hypothesis, we constructed a fusion protein in which the GAL4 DBD was fused to the C-terminus of TFE3. We tested this fusion construct and others in cotransfection experiments with a GAL4 site-dependent reporter plasmid driving the luciferase gene.

Figure 1A shows that the GAL4 DBD alone, GAL4(1–147), did not stimulate transcription. GAL4-TFE3AAD, containing an N-terminal fragment of TFE3 that includes the AAD fused to the GAL4 DBD (2), stimulated transcription 249-fold. The fusion protein containing 102 amino acids from the C-terminus of TFE3 fused to the GAL4 DBD activated transcription to a similar extent, 284-fold. For comparison, activation is shown for GAL4-VP16, a fusion of the GAL4 DBD and the potent VP16 activation domain (9). These data identify a second, C-terminal transcriptional activation domain in TFE3L, separated in primary amino acid sequence from the N-terminal activation domain by the large bHLHZIP domain. The N-terminal activation domain and the C-terminal activation domain have similar transcriptional activating abilities in the GAL4 fusion assay.

Analysis of the C-terminal 134 amino acids of TFE3 shows the region is rich in Pro (13%), Ser (16%), Leu (13%), Gly (9%) and acidic residues (15%, Asp plus Glu). This profile is similar to that of other activation domains that have been termed Pro-rich. The Pro-rich domain was originally described for CTFI; the C-terminal 99 amino acids of CTFI contain 19% Pro, 13% Ser, 10% Leu, and 10% Gly, but only 5% acidic residues. Oct-2, a B-cell enriched member of the POU domain family, contains a C-terminal domain (residues 314–460) that is 13% Pro, 14% Ser, 8% Leu, 10% Gly, and 4% acidic residues. TFEB, a bHLHZIP protein with homology to TFE3, has a 150 amino acid C-terminal domain that is 14% Pro, 13% Ser, 13% Leu, 10% Gly, and 15% acidic residues. AP2 has a 70 amino acid activation domain (residues 51–120) that is Pro and Gln rich; it is comprised of 21% Pro, 9% Ser, 9% Leu, 7% Gly, and 16% Gln. Hox4.2 contains an 86 amino acid domain (residues 24–109) containing 22% Pro, 6% Ser, 6% Leu, 23% Gly, and 11% Ala.



B

```
          464
TFE3   GLPVPPNPGL LSLTTSSVSD SLKPEQLDIE EEGRPSTTFH VSGGPAQNAP
TFEB   GLPTTSPSG. ..VNMAELAQ QVVKQELPSE DGPGEALMLG PEVPEPEQMP
  Mi   GLSLIPSTG. ..LCSPDLVN RIIKQEPVLE NCSQE..... ..........

TFE3   PQQPPAP... ...PSDALLD LHFPSD.HLG DLGDPFHLG. ..........
TFEB   ALPPQAPLPS AAQPQSPFHH LDFSHGLSFG GGGDEGPTGY PDTLGTEHGS
  Mi   LVQHQADLTC TT.......T LDLT...... ....DGTITF TNNLGTMPES

TFE3   .......... ...LEDILME EEGMVGGLSG GALSPLRAAS DPLLSSVSPA
TFEB   P....FPNLS KKDLDLMLLD D......... .SLLPL.A.S DPLFSTMSPE
  Mi   SPAYSIPRKM GSNLEDILMD D......... .ALSPV.GVT DPLLSSVSPG

TFE3   VSNASSRRSS FSIEEES*..
TFEB   ASKASSRRSS FSMEEGDVL*
  Mi   ASKTSSRRSS MSAEETEHA*
```

Figure 1. (**A**) Analysis of activation domains of TFE3. Different protein regions were fused to the GAL4 DBD and tested for activation by cotransfection with a GAL4 site-dependent reporter plasmid driving expression of the luciferase gene. GAL4-TFE3AAD was previously described as GAL4λ3-Δ2 and fuses 126 N-terminal residues from TFE3 to GAL4(1–147) (2). TFE3PRO comprised the C-terminal 102 residues from TFE3 fused to the GAL4 DBD and GAL4-VP16 contains the activation sequences of the VP16 protein. Luciferase values from at least three independent transfections were normalized to that of the reporter alone and are shown at the right for each activator plasmid. (**B**) Sequence homology among murine TFE3, TFEB and mi. The Pileup algorithm (GCG software) was used to compare the C-termini of these three proteins. Three regions of high homology are boxed and proline residues are shown in bold. The horizontal arrow indicates the beginning of the 102 amino acid region of TFE3 shown to be sufficient for activation when fused to the GAL4 DBD.

Despite similar amino acid contents, sequence homology has not previously been demonstrated among proline-rich domains. However, we compared the sequence of the proline-rich C-terminus of TFE3 with the proline-rich C-termini of TFEB and micophthalmia (mi) (11), two proteins in the μE3 family of bHLHZIP domain proteins. We have cloned a murine TFEB cDNA which extents further 3′ than that originally reported for human TFEB (12). This cDNA was sequenced to provide the amino acid data shown in Figure 1B. Substantial homology was detected in the C-termini of TFE3, mi and TFEB. This homology extends over 52 amino acids in the extreme C-terminus of TFE3 and is distributed among three homology boxes containing 38 amino acids (Fig. 1B). Within the homologous 38 amino acids, 50% of the residues are identical among all three proteins and 45% are identical between two of the three. The conserved region is contained within the 102 amino acid C-terminal portion of TFE3 shown to be sufficient for activation (Fig. 1A). Interestingly, although the proline content in the highly conserved region is low (3/38), all three prolines are perfectly conserved.

The two TFE3 activation domains activate transcription synergistically

We showed previously that TFE3L is capable of synergistic activation of transcription; as the number of cognate binding sites upstream of a TATA element is increased from four to eight, transcription driven by TFE3L is increased by substantially more than a factor of two (5). Although proteins with a single activation domain have been shown to exhibit this phenomenon of synergy, we wondered if part of the activation properties of TFE3L were dependent upon synergistic stimulation of transcription by the separate N-terminal and C-terminal activation domains.

We first addressed this question in the context of the native TFE3 protein. Transcriptional activation was compared among TFE3L (containing both activation domains), TFE3S (containing only the C-terminal activation domain) and TFE3L-ΔCterm (a truncated protein containing only the N-terminal activation domain). Plasmids encoding these proteins were cotransfected with a reporter dependent upon eight tandem μE3 sites. Figure 2A shows that TFE3L stimulated the promoter 105-fold. This high level of stimulation is dependent upon synergistic activation by multiple DNA-bound TFE3 molecules. TFE3S activated 19-fold and TFE3L-ΔCterm activated 9-fold. If activation by TFE3L on the multiple-site reporter were simply dependent upon the sum of activation by the AAD and the Pro-AD, one would predict that the activities of TFE3S and TFE3L-ΔCterm would add up to the activity of TFE3L. Since this is not the case, the data suggest that it is the combination of acidic and Pro-rich activation domains that enables TFE3L to activate at such high levels and that the AAD and Pro activation domain stimulate transcription synergistically with one another.

Figure 2A also shows that TFE3S-ΔCterm cannot activate the μE3 site reporter plasmid. The results with this C-terminal truncation of TFE3S confirm that the residual activity of TFE3S is attributable to the Pro-rich domain. Furthermore, without the exon encoding the 35 amino acid AAD, the remaining N-terminal and bHLHZIP domains are devoid of transactivating potential.

To rule out the possibility that differential protein expression confounds the interpretation of our results, immunoblots were performed on whole cell extracts of NIH3T3 fibroblasts transfected with equal amounts of expression plasmid. TFE3L-ΔCterm was expressed at slightly lower levels than TFE3L. This small difference in expression may be accounted for by a loss of epitopes in the truncated protein. The polyclonal antiserum was generated using TFE3S as an antigen and therefore is likely to have significant antibody activity directed against the proline-rich C-terminus. Expression of TFE3L and TFE3L-ΔCterm are therefore comparable.

Surprisingly, TFE3S and TFE3S-ΔCterm were expressed at significantly higher levels than TFE3L. Densitometric scanning of the autoradiogram showed that the TFE3S based isoforms were expressed at ~2.5-fold higher levels than their TFE3L based counterparts. These data show that the various forms of TFE3 protein used in Figure 2A are expressed at least as well as TFE3L.

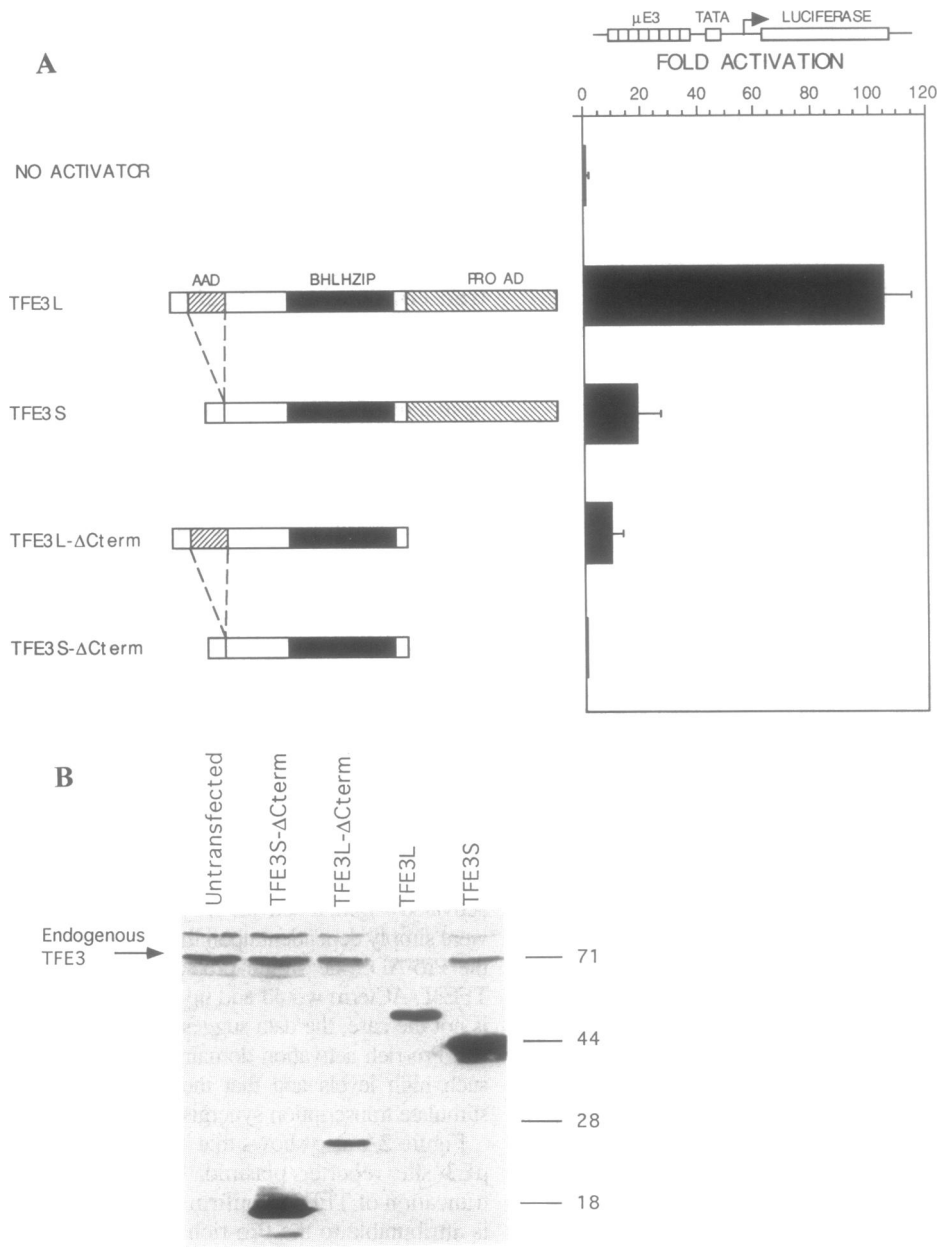


Figure 2. (**A**) The activation domains of TFE3 act synergistically. Cotransfections were performed with plasmids encoding the indicated TFE3 proteins and a reporter plasmid containing eight tandem μE3 binding sites. TFE3L contains both activation domains and TFE3S is the isoform that lacks the N-terminal activation domain. TFE3L-ΔCterm and TFE3S-ΔCterm were engineered to include a stop codon after the leucine zipper of TFE3L and TFE3S, respectively. TFE3L-ΔCterm contains only the N-terminal activation domain, and TFE3S-ΔCterm contains neither activation domain. (**B**) Western blot of extracts from transfected NIH3T3 cells. Cells were transfected with 20 μg of the expression plasmid indicated. Western blots using a polyclonal anti-TFE3S antiserum were performed on whole cell extracts. Lane 1, untransfected cells; lane 2, TFE3S-ΔCterm; lane 3, TFE3L-ΔCterm; lane 4, TFE3L; lane 5, TFE3S. The positions of endogenous TFE3 related proteins and the migration of molecular weight standards are shown.

Therefore, the lower transactivation potential of the shorter proteins is an innate property of these proteins and is not due to decreased expression. This strengthens our conclusion that the N- and C-terminal activation domains of TFE3L act synergistically.

The activation domains of TFE3 activate synergistically when tethered to different DNA-binding domains

To examine further the capacity of TFE3's activation domains to act synergistically, we asked if synergistic activation could be reproduced when the activation domains were joined to separate DNA-binding domains and bound to the same DNA template. Because this approach changes the spacing and, possibly, orientation of the activation domains relative to their positions in the native protein, it might reveal different functional consequences of domain interactions. A reporter plasmid containing five multimerized GAL4 binding sites and four multimerized μE3 binding sites upstream of a TATA box and the luciferase gene was constructed. Cotransfections were performed with TFE3 proteins containing either the AAD or the Pro-rich domain and with hybrid

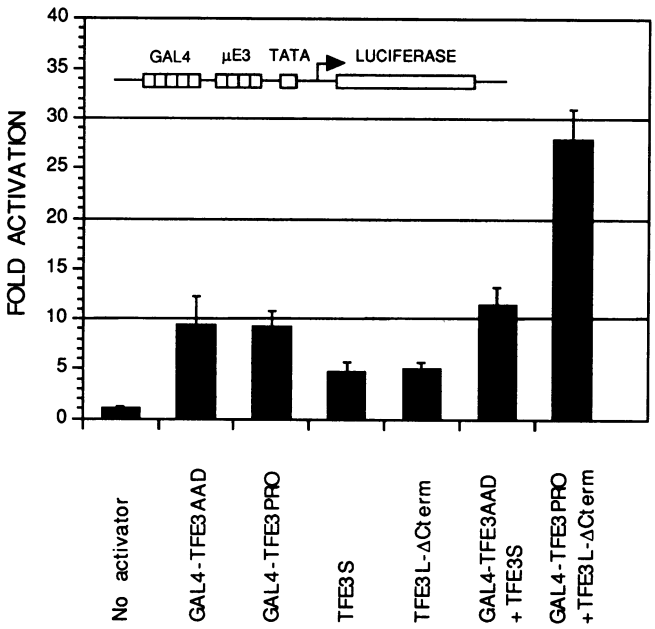


Figure 3. The activation domains of TFE3 act synergistically when tethered to different DNA binding domains on the same DNA template. TFE3 proteins containing a single activation domain were transfected either alone or in combination with GAL-4-fusion constructs expressing a single TFE3 activation domain. Activity was measured from a reporter plasmid with five GAL4 sites and four μE3 sites.

proteins comprised of the GAL4 DNA-binding domain fused to either the TFE3 AAD or the Pro-rich domain. GAL4-TFE3AAD and GAL4-TFE3PRO each activated this reporter plasmid ~9-fold (Fig. 3). Note that the fold activation is lower than in Figure 1 because the basal activity of the GAL4/μE3 reporter is higher than the GAL4 reporter due to activation by endogenous μE3 binding proteins. TFE3S and TFE3L-ΔCterm yield similar levels of activation of this reporter plasmid. Cotransfection of GAL4-TFE3Pro and TFE3L-ΔCterm results in 28-fold activation, ~2-fold higher than the sums of activation by either protein alone. Reversing the positions of the activation domains by transfecting GAL4-TFE3AAD and TFE3S did not result in synergistic activation. Thus, synergistic activation by AAD and Pro-rich activation domain can be recapitulated by removing these domains from their normal protein context and tethering them to different DNA-binding domains; however, this synergism is not reproduced in every geometric arrangement.

DISCUSSION

A new activation domain for TFE3

We have shown that the C-terminal 102 residues of TFE3 are sufficient to confer strong activation on a hybrid protein when fused to the DBD of GAL4. Although this region is proline-rich and has a similar amino acid content to other proline-rich domains, sequence comparisons among TFE3 family members revealed highly conserved sub-domains, containing only three prolines, located C-terminal to a highly proline-rich portion of the region. It will be interesting to determine if these conserved subdomains are either necessary or sufficient for transcriptional activation. It may be that both prolines and the conserved regions are necessary. For example, proline residues might provide unstructured 'punctuation' between small structured regions which contact the basal transcription machinery. Alternatively, conserved regions and not the prolines may be functionally important. Recent dissection of both glutamine and acidic domains has shown that hydrophobic residues, not the prevalent glutamine and acidic residues, are required for interaction of these domains with their protein targets (13,14). The prevalent amino acids upon which classification of activation domains has been based may serve a more general structural role or an as yet unidentified function.

Importance of synergistic activation by TFE3 activation domains

We have shown that multiple DNA-bound molecules of TFE3 synergistically activate transcription and that this synergism is dependent upon the presence of both activation domains. This was demonstrated within the context of the native TFE3 protein and when the AAD and the Pro-rich activation domain were fused to different DNA binding domains. In what circumstances is synergy between TFE3's activation domains likely to be important biologically? TFE3 has been shown to mediate communication between promoter and enhancer elements. The HLHZIP domain controls interaction between widely separated elements and the activation domains determine synergistic activation (5). The immunoglobulin heavy chain gene contains a μE3 site in its V_H promoters and in the intronic enhancer. TFE3 may facilitate association of these elements. Once these elements are in close proximity, synergistic activation by TFE3 molecules bound in the promoter and in the enhancer is likely to contribute to activation of the IgH gene.

Demonstration of synergistic activation by TFE3L's activation domains also provides insight into the mechanism of inhibition of TFE3L activity by the alternately spliced isoform TFE3S. TFE3S mRNA is expressed at lower concentrations than TFE3L mRNA, ranging from 2 to 18 % of TFE3L in all tissues and cell lines tested. The activity of TFE3L is approximately three to four times greater than TFE3S. When cotransfected at substoichiometric amounts similar to those that exist *in vivo*, TFE3S inhibits the activity of TFE3L (4). In fact, in experiments in which TFE3S comprised only 20% of the input mixture of TFE3 isoforms, activity was indistinguishable from that of TFE3S alone.

How is inhibition at substoichiometric ratios possible? We suggest that two mechanisms play a role. First, the presence of one short monomer in a mixed dimer of TFE3L and TFE3S (L–S) appears to destroy synergy between activation domains. If there were no synergy between the activation domains, a mixed dimer (L–S) would have an activation potential exactly intermediate to that of TFE3L and TFE3S dimers, assuming equal dissociation constants for dimer formation and for DNA binding. This is not the case because activity is low with subtoichiometric amounts of TFE3S—a situation where primarily L–S and L–L dimers are expected to be present (4). In view of the synergy observed in this work, we suggest that lack of inter-monomer synergy accounts for the low activity of L–S dimers. Thus, the model is that activity of L–L dimers is dependent upon the synergistic combination of AAD and Pro-rich domains and that removing one AAD from the dimer reduces the activity of the mixed dimer, rendering it much closer to the activity of the S–S dimer. Second, our data suggest that TFE3S protein is more stable that TFE3L protein which would lead to a higher molar ratio of TFE3S/TFE3L protein

relative to the RNA ratio. Nevertheless, the differences in protein expression are not large enough to produce a molar excess of TFE3S when TFE3L and TFE3S and transfected in a 4:1 plasmid ratio. Therefore, disruption of synergism among the activation domains of TFE3 appears to be an important component of the dominant negative action of TFE3S.

The generation of alternate forms of transcription factors from the same gene which is seen in other proteins as well, including FosB, LAP and CREM, is an economical means of regulating the activity of key proteins (15,16). In the absence of synergistic effects such as the ones demonstrated here, dominant negative forms of these proteins would have to be generated at significantly higher levels to inhibit the activities of the more strongly activating forms. It is likely that nature exploits the effects of activation domain synergism in such a way that dramatic changes in gene expression can be achieved by more subtle alterations in the concentrations of isoforms generated from the same transcription factor gene.

A model for TFE3 activation

Stimulation of transcription by TFE3 is mediated by two separate activation domains, one rich in acidic residues and located in the amino terminus, and one rich in Pro residues and located in the carboxy terminus. Analysis of many other transcription factors has shown they are often comprised of multiple activation domains. Typically, these domains are in contiguous segments of the protein (17–19) although sometimes activation domains are separated in the primary amino acid sequence (20). Frequently, activation domains are comprised of multiple subdomains, which, when deleted, reduce the activity of the protein in a gradual fashion, without evidence of synergistic activation. This has been shown for GAL4 (20), C/EBP (19) and USF (18), among others. The two activation domains of TFE3L, however, are not subdomains of a larger activation domain and are likely to be separated in three dimensional space by the relatively large alpha helices which confer DNA binding and subunit dimerization. We propose a model for the general structure of the TFE3 protein that takes into account the data presented above, the location of the domains in the primary amino acid sequence, and specific structural details of the bHLHZIP–DNA complex of Max determined by X-ray crystallography (21).

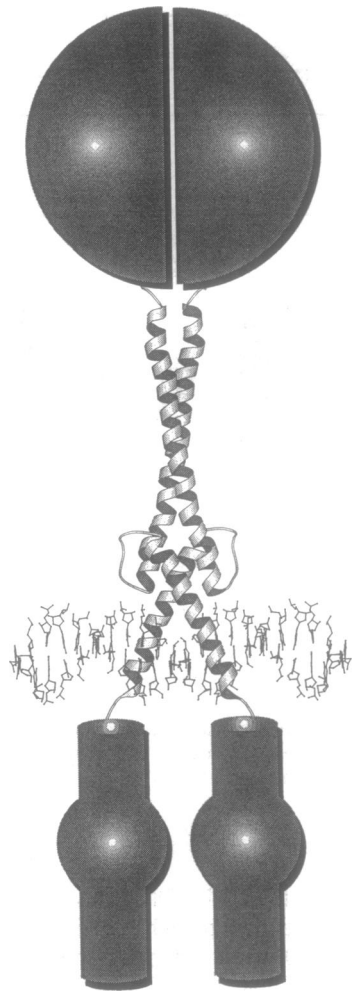


Figure 4. Model for general structure of TFE3. The TFE3L protein is proposed to form a dumbbell shaped protein with N-terminal (bottom) and C-terminal (top) activation domains widely separated in space. Activation domains are depicted as spheres. The bHLHZIP domain and DNA were generated as a ribbon structure from the max-DNA coordinates (21) using Molscript.

We propose that the TFE3 protein dimer, when bound to DNA, assumes a 'dumbbell shape'; the shaft of the dumbbell represents the bHLHZIP domain and the ends of the dumbbell represent the AADs and the Pro-rich activation domains (Fig. 4). The bHLHZIP domain that separates the activation domains is an unusual motif in a non-structural protein. It is an extended domain comprised of two parallel α-helices connected by a short loop. One helix contains the basic region that binds DNA and the helix1 portion of the HLH domain. The other helix contains the helix2 portion of the HLH domain as well as the leucine zipper. The axis of the HLHZIP domain is perpendicular to the DNA axis. The basic region α-helix passes through the major groove of the DNA binding site and terminates on the other side. Thus, sequences N-terminal to the basic region must at least begin on the other side of a plane that is perpendicular to the HLHZIP domain and that contains the DNA molecule. The dumbbell model predicts that the AAD and the Pro-rich domains will be widely separated in space and will lie on opposite sides of the DNA molecule. The alternative, which we cannot rule out, is that each activation domain turns back toward the other, ending up in close proximity. The extended nature of the bHLHZIP structure makes this less likely.

Targets for the activation domains of TFE3 have not yet been identified, but candidates exist. TBP (22), TFIIB (23), TAFII40 (24) and TAFII60 (14) have been shown to interact with other AADs. The Pro-rich domain of CTFI has been shown to require TFIIB (25) as well as TAFs (26) for transactivation. However, it is also possible that other targets will actually be demonstrated to mediate TFE3 activation. The minimal sequence homology among members of a given family of activation domains make a single protein target unlikely. In fact, TAFII110 interacts with the Gln-rich sequences of Sp1 and CREB, but not those of Antennapedia or bicoid (13,27).

The proposed dumbbell shape of TFE3 and our data that show synergistic interaction of both activation domains suggest a three dimensional view of how TFE3 might direct formation of a stereospecific preinitiation complex. Transactivation of a promoter by TFE3 would therefore involve molecular connections

between TBP and TFE3 both above and below the plane of the DNA helix. Thus, TAFII40 or TFIIB might bridge between the TFIID complex and the AAD of TFE3 on one side of this plane, while TFIIB or another TAF might bridge between the Pro-rich domain and the TFIID complex on the other side of this plane.

Of the many TFE3, USF and Myc/max related proteins that comprise the bHLHZIP family of proteins, only TFE3, TFEB and mi contain a proline-rich region C-terminal to the ZIP domain. USF, c-Myc and max terminate shortly after the ZIP domain. In addition to having different activation sequences, the overall geometric shapes and structures of these proteins are likely to be different. The criteria that dictate whether a gene is regulated by TFE3, USF or c-Myc/max are not known. It is tempting to speculate that the stereospecific complex formed by promoter-bound factors and components of the basal transcription apparatus will encourage or disfavor incorporation of a sequence specific factor based upon the sequence of its activation domains and their geometric organization.

Similarity between TFE3 and Oct-2

To our knowledge, Oct-2 is the only protein previously shown to contain multiple activation domains that activate synergistically (28). The overall organization of the Oct-2 protein is similar to that of TFE3. Each protein is comprised of a DNA binding domain sandwiched between an N-terminal activation domain (Gln-rich in Oct-2) and a C-terminal Pro-rich activation domain. The recent X-ray crystal structure of the Oct-1 POU domain (29) shows that the POU specific domain contacts the 5′ half of the binding site. This domain is connected by a flexible linker to the POU homeodomain which contacts the 3′ half of the OCT site on the opposite face of the DNA double helix. Therefore, bHLHZIP domains and POU domains achieve placement of their N- and C-termini on opposite sides of the DNA molecule by using completely different binding modules. If the N- and C-terminal activation domains of Oct-2 and TFE3 are indeed located on opposite sides of the DNA molecule, it suggests formation of a preinitiation complex circumferentially around the double helix in elements such as the V_H promoters that contain adjacent binding sites for both factors.

ACKNOWLEDGEMENTS

We thank members of the Calame lab for helpful discussions. We are grateful to Dr V. Racaniello for assistance with the Molscript software, to Dr M. Carey for providing constructs and to Dr A. Aggarwaal for critically reading the manuscript. S.A., K.W. and K.M. were supported by an NIGMS M.S.T.P. training grant and this work was supported by USPH grant GM29361 to K.C.

REFERENCES

1 Tjian, R. and Maniatis, T. (1994) *Cell*, **77**, 5–8.
2 Beckmann, H., Su, L.-K. and Kadesch, T. (1990) *Genes Dev.*, **4**, 167–179.
3 Roman, C., Matera, G., Cooper, C., Artandi, S., Blain, S., Ward, D. and Calame, K. (1992) *Mol. Cell. Biol.*, **12**, 817–827.
4 Roman, C., Cohn, L. and Calame, K. (1991) *Science*, **254**, 94–97.
5 Artandi, S. E., Cooper, C., Shrivastava, A. and Calame, K. (1994) *Mol. Cell. Biology*, **14**, 704–7716.
6 Dang, C. V., McGuire, M., Buckmire, M. and Lee, W. M. F. (1989) *Nature*, **337**, 664–666.
7 Fisher, D. E., Carr, C. S., Parent, L. A. and Sharp, P. A. (1991) *Genes Dev.*, **5**, 2342–2352.
8 Ferre-D'Amare, A., Pognonec, P., Roeder, R. G. and Burley, S. K. (1994) *EMBO J.*, **13**, 180–189.
9 Sadowski, I. and Ptashne, M. (1989) *Nucleic Acids Res.*, **17**, 7539.
10 deWet, J. R., Wood, K. V., DeLuca, M., Helsinki, D. R. and Subramani, S. (1987) *Mol. Cell. Biol.*, **7**, 725.
11 Hodgkinson, C. A., Moore, K. J., Nakayama, A., Steingrimsson, E., Copeland, N. G., Jenkins, N. A. and Arnheiter, H. (1993) *Cell*, **74**, 395–404.
12 Carr, C. and Sharp, P. (1990) *Mol. Cell. Biol.*, **10**, 4384–88.
13 Gill, G., Pascal, E., Tseng, Z. H. and Tjian, R. (1994) *Proc. Natl. Acad. Sci. USA*, **91**, 192–196.
14 Thut, C. J., Chen, J.-L., Klemm, R. and Tjian, R. (1995) *Science*, **267**, 100104.
15 Foulkes, N. and Sassone-Corsi, P. (1992) *Cell*, **68**, 411–414.
16 Descombes, P. and Schibler, U. (1991) *Cell*, **67**, 569–579.
17 Courey, A. J. and Tjian, R. (1988) *Cell*, **55**, 887–898.
18 Kirschbaum, B. J., Pognonec, P. and Roeder, R. G. (1992) *Mol. Cell. Biol.*, **12**(11), 5094–5101.
19 Friedman, A. D. and McKnight, S. L. (1990) *Genes Dev.*, **4**, 1416–1426.
20 Ma, J. and Ptashne, M. (1987) *Cell*, **48**, 847–853.
21 Ferre-D'Amare, A. R., Prendergast, G. C., Ziff, E. B. and Burley, S. K. (1993) *Nature*, **363**, 38–45.
22 Stringer, K. F., Infles, C. J. and Greenblatt, J. (1990) *Nature*, **345**, 783–786.
23 Lin, Y.-S., Ha, I., Maldonado, E., Reinberg, D. and Green, M. R. (1991) *Nature*, **353**, 569–571.
24 Goodrich, J. A., Hoey, T., Thut, C. J., Admon, A. and Tjian, R. (1993) *Cell*, **75**, 519–530.
25 Kim, T. K. and Roeder, R. G. (1994) *Proc. Natl. Acad. Sci. USA*, **91**, 4170–4174.
26 Tanese, N., Pugh, B. F. and Tjian, R. (1991) *Genes Dev.*, **5**, 2212–2224.
27 Hoey, T., Weinzierl, R. O. J., Gill, G., Chen, J.-L., Dynlacht, B. D. and Tjian, R. (1993) *Cell*, **72**, 247–260.
28 Tanaka, M., Clouston, W. M. and Herr, W. (1994) *Mol. Cell. Biol.*, **14**, 6046–6055.
29 Klemm, J. D., Rould, M. A., Aurora, R., Herr, W. and Pabo, C. O. (1994) *Cell*, **77**, 21–32.