Transcription factor AP-4 contains multiple dimerization domains that regulate dimer specificity

Yan-Fen Hu, Bernhard Lüscher,[1] Arie Admon, Nicolas Mermod,[2] and Robert Tjian

Howard Hughes Medical Institute, Department of Molecular and Cell Biology, University of California at Berkeley, Berkeley, California 94720 USA

Enhancer binding protein AP-4 is a transcription factor that activates both viral and cellular genes by binding to the symmetrical DNA sequence, CAGCTG. Here, we report the molecular cloning and characterization of human AP-4 cDNAs. The deduced amino acid sequence reveals that AP-4 is a helix–loop–helix (HLH) protein. Like other members of this family, the AP-4 HLH motif and the adjacent basic domain are necessary and sufficient to confer site-specific DNA binding. However, unlike other HLH proteins, AP-4 also contains two additional protein dimerization motifs consisting of leucine repeat elements LR1 and LR2. The analysis of various deletion and point mutants for their ability to dimerize in the presence or absence of DNA reveals several unusual features. Although the HLH basic region is sufficient for DNA recognition and binding, dimer formation between different truncated versions of AP-4 in solution requires an intact LR1 or LR2 domain. AP-4 is unable to form heterodimers with other helix–loop–helix family members such as the immunoglobulin enhancer binding factor, E12. In contrast, an AP-4 derivative, ΔC222, which lacks LR1 and LR2 but retains an intact HLH, can form heterodimers with E12. Moreover, AP-4 molecules containing LR2 or LR1 are unable to form mixed dimers with carboxy-terminally truncated AP-4 molecules such as ΔC222, but retain the ability to form complexes with longer versions of AP-4 that contain LR1 and/or LR2. Our findings strongly suggest that AP-4 contains multiple protein–protein interfaces that function to promote homodimer formation and restrict heterocomplexes. These findings provide a mechanism by which different members of the helix–loop–helix family of transcription factors can form functional dimers in a specific fashion with their appropriate partners to control transcriptional networks during cellular differentiation.

[*Key Words*: AP-4; helix–loop–helix; leucine repeat; dimerization specificity]

Received June 22, 1990; revised version accepted August 8, 1990.

The coordinated control of eukaryotic gene expression is, in large measure, directed by the action of sequence-specific transcription factors. These DNA-binding proteins modulate the frequency of transcriptional initiation by interacting with specific promoter and enhancer elements. Structure–function studies of various cloned eukaryotic transcription factors have revealed a modular construction consisting of DNA binding and transcriptional activation domains combined, in some cases, with specific homo- or heterodimer interfaces (Mitchell and Tjian 1989). A comparison of amino acid sequences deduced from several cloned genes has led to the identification of distinct structural motifs responsible for transcriptional activation, DNA binding, and dimerization. Studies of transcriptional activation domains reveal various, as yet poorly defined, structural features such as acidic helices, glutamine-rich, proline-rich, and serine/threonine-rich domains (Ptashne 1988; Courey et al. 1989; Mermod et al. 1989; Tanaka and Herr 1990). DNA-binding domains with more clearly defined structures have been characterized, such as zinc fingers, helix–turn–helix/homeo domains, basic/leucine-zipper (B-zip) proteins, and a variety of other "basic" motifs (Miller et al. 1985; Kadonaga et al. 1987; Clerc et al. 1988; Hoey and Levine 1988; Landschulz et al. 1988b; Müller et al. 1988; Sturm et al. 1988).

In addition to structural domains that function by contacting DNA or interacting with the basal transcription components, some regulatory factors also contain specialized protein–protein oligomerization domains. In particular, certain classes of DNA-binding structures rely on the formation of specific homo- or heterodimers to create a functional DNA recognition site. One of the dimerization motifs that provides an integral framework for a DNA recognition domain is the so-called "leucine repeat" of the B-zip class of DNA-binding proteins

Present addresses: [1]Institute of Pharmacology, University of Zürich, Switzerland; [2]Institut de Biologie Animale, Université de Lausanne, Lausanne, Switzerland.

(Landschulz et al. 1988b). This amphipathic α-helix motif is characterized by the presence of leucine residues at every seventh position, interdigitated with other hydrophobic amino acids at the 3-4 position. Two leucine repeats aligned in a parallel fashion are thought to form a coiled-coil structure resulting in either homodimers or heterodimers, as demonstrated by studies of CCAAT/enhancer binding protein (C/EBP), Jun–Fos, and GCN4 (Kouzarides and Ziff 1988; Landschulz et al. 1988b, 1989; Gentz et al. 1989; O'Shea et al. 1989; Turner and Tjian 1989).

Recently, a different class of DNA-binding and dimerization domain referred to as the helix–loop–helix (HLH) motif has been described (Murre et al. 1989a). This motif is found in a variety of enhancer binding proteins and putative transcription factors involved in the control of cellular proliferation and differentiation (Edmonson and Olson 1989; Lassar et al. 1989; Ludwig et al. 1989; Mellentin et al. 1989; Murre et al. 1989a,b; Beckmann et al. 1990; Henthron et al. 1990). A distinctive feature of HLH proteins is the capacity to form heteromeric complexes with many different members of this large and growing family of proteins. For instance, the immunoglobulin gene enhancer binding proteins, E12 and E47, the muscle-specific factors MyoD and myogenin, the *Drosophila* regulatory proteins *daughterless* and *achaete-scute* are all able to form heteromeric complexes with each other as well as homodimers (Murre et al. 1989b). Moreover, these different HLH proteins share in common the ability to recognize a conserved symmetrical core DNA sequence, CAXXTG. Thus, the combined action of different HLH proteins may provide an intricate network for directing gene transcription in eukaryotic cells.

Given their propensity to form heterodimers with each other, how do these different regulators recognize the correct partners to form the desired homo- or heterodimers to carry out their functions and, at the same time, prevent the incorrect combinations from occurring? One possibility is that different HLH proteins are expressed in a temporally or spatially restricted manner in the organism. This is apparently the case with some tissue-specific factors such as MyoD, and the *emc* and *achaete-scute* gene products (Davis et al. 1987; Edmonson and Olson 1989; Wright et al. 1989). On the other hand, many HLH proteins such as the immunoglobulin enhancer-binding proteins E12/E47 and TFE3, are ubiquitously expressed (Murre et al. 1989a; Beckmann et al. 1990). It would therefore be of interest to decipher the mechanisms that may regulate specific homodimer or heterodimer formation within this important family of regulatory proteins.

Here, we report that transcription factor AP-4 is a new member of the HLH family. AP-4 was identified first as a cellular protein that binds to the SV40 enhancer and activates viral late gene transcription (Mermod et al. 1988). In several cellular promoters, as well as some viral enhancers, AP-4 binding sites overlap the recognition sequences of other transcription factors such as AP-1/Jun, suggesting possible factor–factor interactions. Mermod et al. (1988) have found that in vitro, AP-4 and AP-1 can act synergistically to activate both SV40 late and the hMTIIA gene transcription. Recently, several studies have also suggested AP-4 or related factors in the regulation of human proenkephalin expression (Comb et al. 1988). Thus, AP-4 will offer an opportunity to examine the interplay of different transcription factors with each other and to extend our understanding of how homo- and heterocomplexes contribute to regulation.

Here, we describe the isolation of a partial cDNA clone encoding human AP-4. Analysis of the AP-4 cDNA product suggests that it encodes an HLH protein that recognizes the symmetrical core DNA sequence, CAGCTG. Although AP-4 is a member of the HLH family, the protein does not form heterodimers with other HLH proteins. Instead, AP-4 contains multiple distinct protein–protein interfaces, and displays novel dimerization properties. Studies of AP-4 reveal new mechanisms for directing dimerization specificities between different related transcription factors. We also used AP-4 as a model system to dissect further the role of the HLH motif in its capacity as a DNA binding and dimerization element. Finally, we speculate on the structural basis and biological significance of the elaborate AP-4 dimerization machinery.

Results

Isolation of AP-4 cDNAs

Transcription factor AP-4 was purified to near homogeneity by a combination of classical and sequence-specific DNA affinity chromatography (Kadonaga and Tjian 1987) as described by Mermod et al. (1988). The partial amino acid sequences of several tryptic peptides were determined. Two peptide sequences were used to derive best-guess oligonucleotide probes for screening a HeLa cell cDNA library. Four independent cDNA clones were isolated from 2×10^6 λgt10 phage plaques. The longest contiguous AP-4 sequence obtained consisted of 1.9 kb and contains an open reading frame encoding a protein of 321 amino acids (Fig. 1A). Northern blot and primer extension analyses indicate that there are ~240 nucleotides at the 5′ end of the mRNA that are not present in the cDNA (data not shown). It is therefore likely that our longest cDNA sequence lacks some coding regions as well as noncoding leader sequences. Further attempts to obtain a full-length cDNA clone for AP-4 from a variety of other cDNA libraries were unsuccessful.

Two lines of evidence establish that the clone we isolated encodes the AP-4 protein from HeLa cells. First, five independent peptide sequences obtained from purified AP-4 are present in the amino acid sequence of AP-4 deduced from the DNA sequence (underlined in Fig. 1A). Second, the AP-4 cDNA was subcloned into a vaccinia virus expression vector, and DNase I footprinting analysis confirmed that the recombinant AP-4 protein has DNA binding specificity and affinity indistinguishable from purified HeLa cell AP-4 protein, when assayed on both the proenkephalin (Fig. 2B) and α-amylase promoter

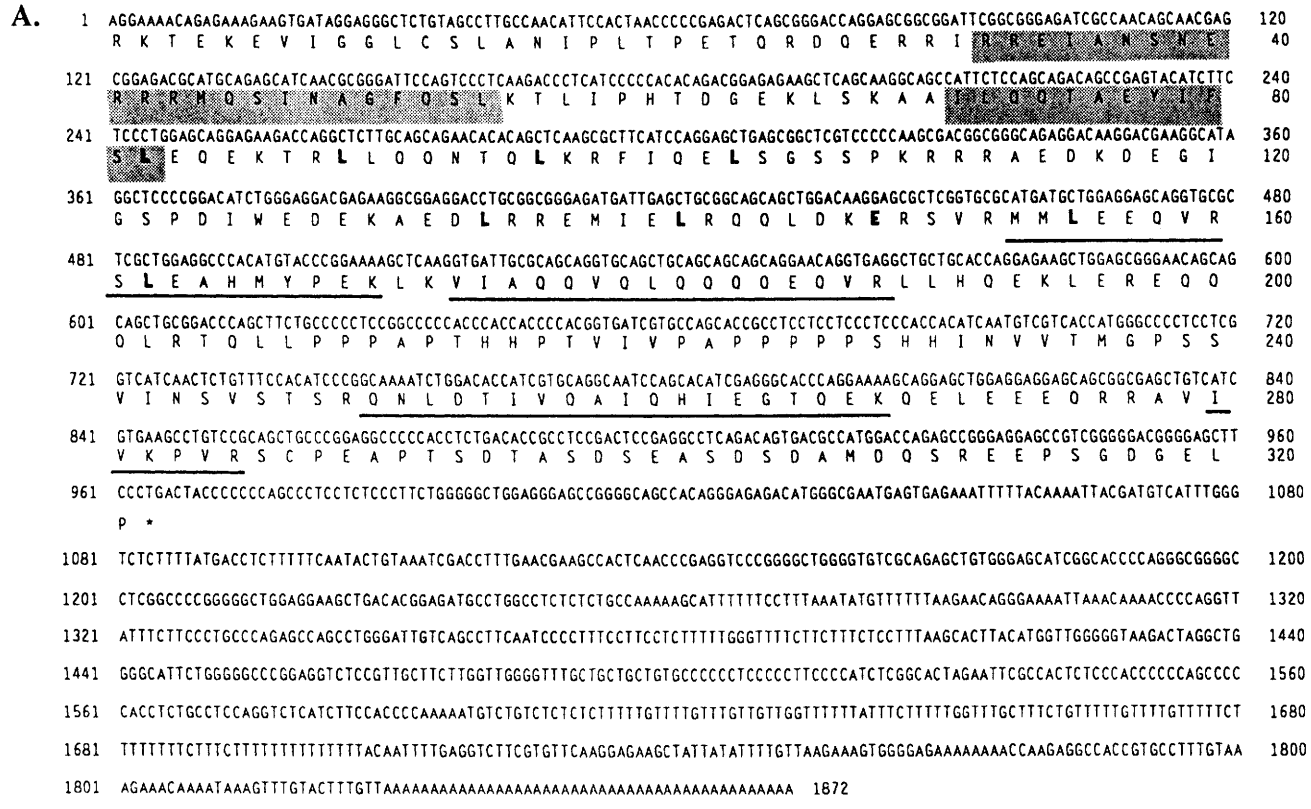

Figure 1. DNA and amino acid sequence of AP-4. (*A*) The nucleotide sequence of the cDNA and the deduced amino acid sequence of AP-4. The shaded sequences correspond to the conserved basic/HLH region; leucine residues of two putative leucine repeats are highlighted. The underlined sequences indicate the tryptic peptide sequences obtained from the purified HeLa cell AP-4 protein. (*B*) Schematic representation of the AP-4 protein. (B) Conserved basic stretch; (HLH) HLH motif; (LR1) first leucine repeat; (LR2) second leucine repeat. The amino acid numbers are marked on the *top*. (*C*) Comparison of the B/HLH motif, leucine repeats of AP-4 and other HLH, leucine repeat proteins. Identical or related amino acids are boxed and noted as a consensus at the bottom. ψ indicates hydrophobic residues. The relevant regions of the following HLH proteins are shown: TFE3 (amino acids 134–199; Beckmann et al. 1990), human c-myc (amino acids 346–401; Battey et al. 1983), human N-myc (amino acids 393–437; DePinho et al. 1987), MyoD (amino acids 108–164; Davis et al. 1987), E12/E47 (amino acids 336–393; Murre et al. 1989), daughterless (amino acids 554–613; Caudy et al. 1988), AS-C sc (amino acids 101–167; Villares and Cabrera 1987), and AS-C ac (amino acids 26–95; Villares and Cabrera 1987). The relevant regions of the following leucine repeat proteins are shown: TFE3 (amino acids 202–223; Beckmann et al. 1998), N-myc (amino acids 434–455; DePinho 1987), c-myc (amino acids 413–434; Battey et al. 1983), c-Jun (amino acids 289–316; Bohmann et al. 1987), C/EBP (amino acids 315–342; Landschulz et al. 1988a), and GCN4 (amino acids 253–278; Hinnebusch 1984).

sequences (B. Lüscher, unpubl.). Although the intact HeLa cell AP-4 has the molecular weight of 48 kD on SDS-PAGE (Mermod et al. 1988), the recombinant protein from the partial clone migrates as a 42-kD product (Fig. 2A), even though the predicted molecular weight should be 37 kD.

The deduced amino acid sequence of AP-4 reveals several noteworthy features that are summarized in Figure 1B. First, between residues 31 and 82 there is a highly conserved HLH motif and an adjacent amino-terminally located stretch of basic amino acids, which, together, most likely comprise the DNA-binding domain. This is

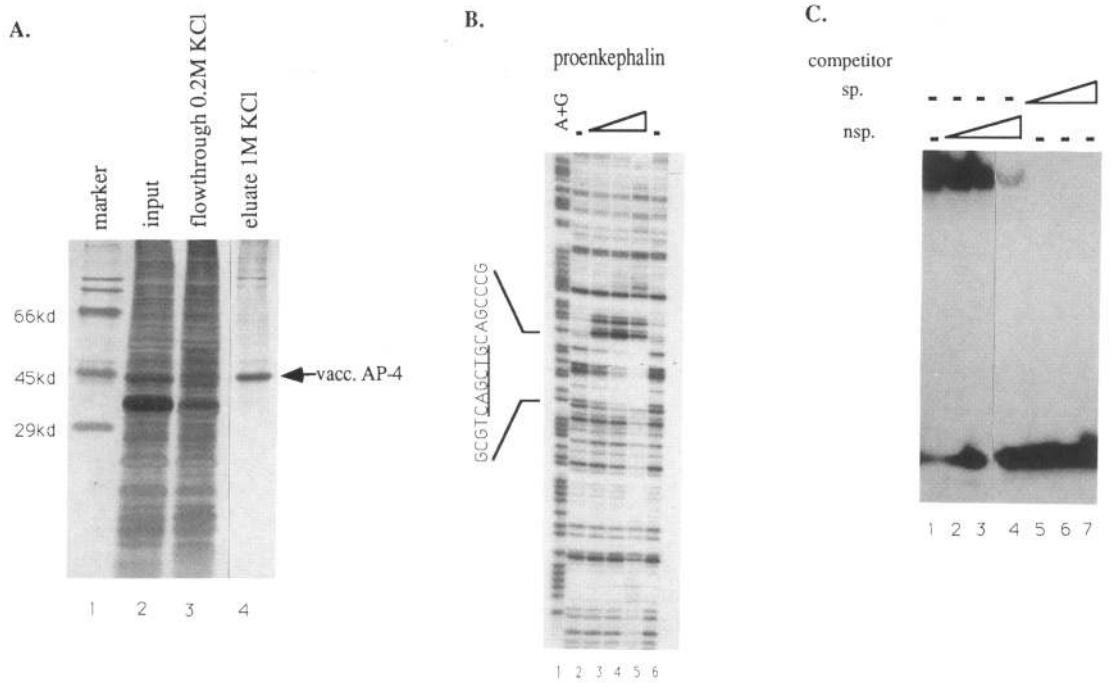

Figure 2. DNA binding activity of AP-4 expressed from a recombinant vaccinia virus or by in vitro translation. (*A*) Purification of vaccinia virus-expressed AP-4 by DNA affinity chromatography. Proteins were subjected to 10% SDS-PAGE and visualized by silver staining. (Lane *1*) Molecular weight standards; (lane *2*) nuclear extract from HeLa cells infected with the recombinant vaccinia virus (80 μg of protein); (lane *3*) flowthrough fraction from the AP-4 affinity column (80 μg of protein); (lane *4*) 1 M KCl eluate from the AP-4 affinity column (0.1 μg of protein). (*B*) DNase I footprinting analysis of recombinant AP-4 protein. Reactions were carried out as described previously (Mermod et al. 1988) using a probe derived from the human proenkephalin promoter that contains an AP-4 recognition site (underlined to left of the footprint). (Lane *1*) Partial purine-specific cleavage products of the probe (A + G). Footprint reactions were carried out with no AP-4 protein (lanes *2* and *6*) or with increasing amounts of partially purified AP-4 protein (0.5 ng, 2.5 ng, and 12.5 ng) (lanes *3–5*). (*C*) Electrophoretic mobility-shift assay (EMSA) of the in vitro-translated AP-4. The probe, designated NM3/4, was a ^{32}P-labeled oligonucleotide obtained from the SV40 enhancer sequence that contains an AP-4 binding site. The same NM3/4 probe was used in all subsequent EMSAs. Reaction shown in each lane was carried out using the same amount of the in vitro-translated AP-4 in the presence of (lane *1*) no competitor DNA; (lanes *2–4*) increasing amounts of poly[d(I-C)] nonspecific competitor DNA (50 ng, 500 ng, and 3000 ng, respectively); (lanes *5–7*) increasing amounts of specific competitor DNA (100 ng, 200 ng, and 500 ng, respectively).

consistent with the finding that the AP-4 binding site (CAGCTG; Mermod et al. 1988) is similar to that of other HLH proteins such as E12/E47 (CAGGTG; Murre et al. 1989a) or MyoD (CACCTG; Lassar et al. 1989). However, unlike MyoD or E12/E47, AP-4 contains two additional putative leucine repeat-type domains between residues 82 and 162. The first leucine repeat, LR1, with four leucines at seven amino acid intervals, is immediately carboxy-terminal to the HLH motif. This leucine repeat does not contain the consensus features of hydrophobic residues at the 3-4 position, and is therefore more similar to the leucine repeat found in the myc protein (Battey et al. 1983; DePinho et al. 1987). The second leucine repeat, LR2, is interrupted by the presence of a single glutamate residue in place of a leucine residue, but has hydrophobic amino acids at every fourth position, except one, and therefore fits more closely with the consensus leucine repeats of proteins such as C/EBP, Jun, and GCN4 (Bohmann et al. 1987; Landschulz et al. 1988b; O'Shea et al. 1989). Detailed comparisons of these motifs are shown in Figure 1C. Finally, between residues 175 and 227, there is a glutamine-rich stretch (13 of 30) followed by a proline-rich sequence (11 of 20). In addition, the carboxy-terminal amino acid tail is fairly acidic (17 of 59). Thus, the carboxy-terminal half of AP-4 contains several structural elements reminiscent of previously defined activation domains.

The basic/HLH motif is necessary and sufficient for DNA binding

The unusual presence of multiple potential dimerization motifs in AP-4 prompted us to address the function of these domains. To characterize AP-4 DNA binding and dimerization activities, we performed electrophoretic mobility-shift assays (EMSAs) with in vitro-translated AP-4 proteins (Hope and Struhl 1987). As shown in Figure 2C, AP-4 binds specifically to a labeled oligonucleotide containing the AP-4 recognition site, CAGCTG. To determine the minimum region of the AP-4 protein required for DNA binding, a series of carboxy-terminal deletion mutants were generated (Fig. 3, bottom). Deletion mutants ΔC87, ΔC141, and ΔC222 all bind the AP-4 recognition sequences (Fig. 3, top). The

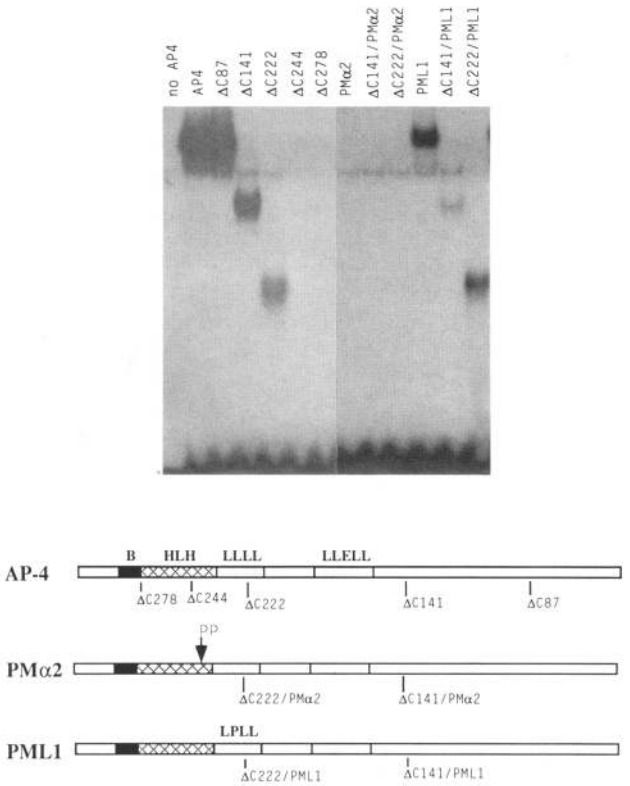

Figure 3. Localization of the minimum region of AP-4 required for DNA binding. The carboxy-terminal deletion mutants were generated by digestion of the AP-4 cDNA template at different restriction sites followed by in vitro transcription and translation of the cDNAs. The resulting products were analyzed by EMSA (*top*). ΔC87, ΔC141, and ΔC222 lack carboxy-terminal 87, 141, and 222 amino acids, respectively (*bottom*). PMα2 contains point mutations in the second amphipathic α-helix of the HLH with prolines substituting for Ile(71) and Leu(80). PML1 contains a point mutation replacing the second Leu of LR1 with Pro. The band seen across the entire gel near the *top* is due to a nonspecific DNA binding activity.

latter lacks the carboxy-terminal 222 amino acids containing LR1 and LR2. In contrast, ΔC244, which lacks part of the HLH region, and ΔC278, which retains the conserved basic region but lacks the entire HLH region, failed to bind DNA. Moreover, point mutant PMα2, in which two amino acid substitutions have been introduced in the second putative amphipathic α-helix of the HLH region, also abolished DNA binding. In contrast, PML1, which has a point mutation within LR1 (Leu2 → Pro), has no detectable effect on DNA binding. These findings strongly support the notion that the basic/HLH motif is necessary and sufficient to confer sequence-specific DNA binding activity and that the leucine repeats of AP-4 are not essential for DNA recognition.

The HLH motif is not sufficient for homodimerization of AP-4

The HLH motif has been demonstrated in HLH proteins such as E12 and E47, MyoD, etc., to be necessary and sufficient for dimerization, as determined by EMSA (Murre et al. 1989b). It has also been reported that the HLH motif is required for oligomerization of MyoD and E12 in solution as judged by coimmunoprecipitation (Davis et al. 1990). We therefore tested the potential involvement of the HLH motif in dimerization of AP-4 in the presence and absence of DNA binding sequences. Different truncated forms of AP-4 were translated, covalently cross-linked by treatment with glutaraldehyde, and subsequently analyzed by SDS-PAGE. As shown in Figure 4A, AP-4, ΔC87, and ΔC141 form homodimers, whereas ΔC222 is unable to form stable dimers in the absence of DNA template, even though this carboxy-terminally truncated mutant contains a complete HLH motif and binds sequence specifically to DNA, as determined by the EMSA. We also performed a cross-linking experiment on the HLH double mutant, ΔC141/PMα2, and found that it readily forms dimers in solution (Fig. 4B). These experiments indicate that, unlike other previously studied HLH proteins, the HLH domain of AP-4 is not sufficient and may not even be necessary for dimer formation in the absence of DNA, presumably because there are other dimerization regions in AP-4. This is the first hint that one or more of the leucine repeats present in AP-4 may play an important function in directing dimer formation.

We also analyzed the ability of AP-4 to form protein–protein complexes in the presence of DNA by EMSA. As expected, when various AP-4 deletion mutants are individually assayed for DNA binding, only one shifted band of protein–DNA complex is observed (Fig. 5A, lanes 2–4). However, when either ΔC87 and ΔC141 or the longest AP-4 and ΔC141 are cotranslated, an additional shifted band with intermediate mobility is observed, indicating the presence of heteroligomers, most likely heterodimers (lanes 7 and 9). In contrast, no intermediate bands representing heterodimer complexes are observed by mixing different truncated AP-4 proteins after separate in vitro translation (data not shown), suggesting that specific protein–protein complexes form in solution before binding to the target DNA sequence. Moreover, this finding suggests that once homo- or heterodimer complexes are formed, they remain stable within the time course of the experiments and do not readily exchange subunits. Interestingly, when AP-4, ΔC87, or ΔC141 were cotranslated with ΔC222 (which lacks LR1 and LR2), no intermediate shifted bands were observed (lanes 8, 10, and 11). These findings, combined with the cross-linking results, suggest that ΔC222, which retains sequence-specific DNA binding activity, is a monomer in solution, whereas the longer proteins exist as dimers and no evidence for tetramer or higher order structures was obtained. Moreover, ΔC222 is unable to form complexes with the longer species, even in the presence of DNA, presumably because this carboxy-terminally truncated protein lacks both LR1 and LR2 and is therefore precluded from forming dimers with the longer versions of AP-4.

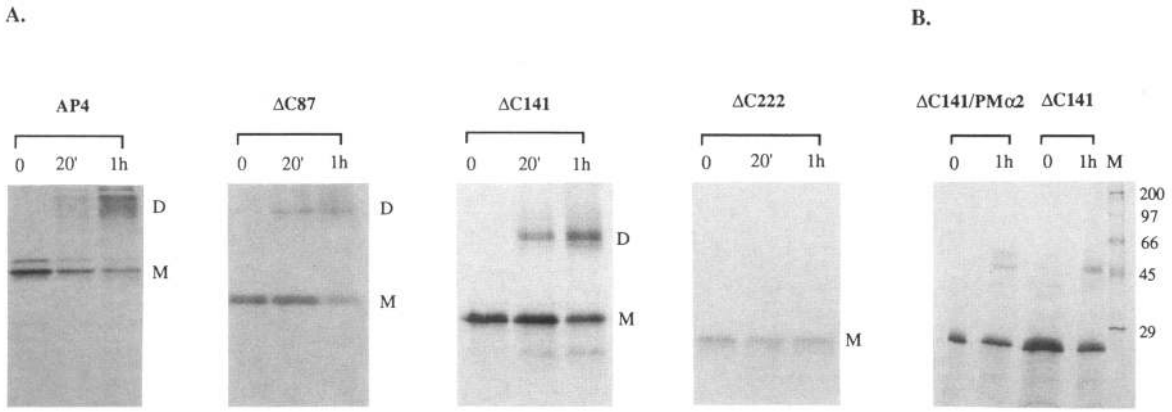

Figure 4. Glutaraldehyde cross-linking of in vitro-translated AP-4 and its derivatives. (*A*) In vitro-translated proteins were purified by DNA affinity chromatography and equal portions of the eluates were subsequently treated with 0% (0) or with 0.001% glutaraldehyde for 20 min (20') or 1 hr (1h). Samples were then resolved by either 12% (AP-4 and ΔC87) or 18% SDS-PAGE (ΔC141 and ΔC222). (M) Monomer; (D) dimer. (*B*) In vitro-translated proteins in crude reticulocyte extract were untreated (0) or treated with 0.001% glutaraldehyde for 1 hr (1h). The products were analyzed on 12% SDS-PAGE. The molecular weight standards are shown at *right*.

The two leucine repeats of AP-4 direct a hierarchy of dimerization interactions

We have shown that AP-4 forms a dimer but that the HLH motif is not required for this activity. Because ΔC141 dimerizes in solution but ΔC222 does not, we surmise that a specific protein–protein interface must be located in the region between residues 82 and 180. The amino acid sequence in this region reveals two putative leucine repeats, LR1 and LR2. To map the active dimerization domains more precisely, a number of additional deletions were generated and the resulting trun-

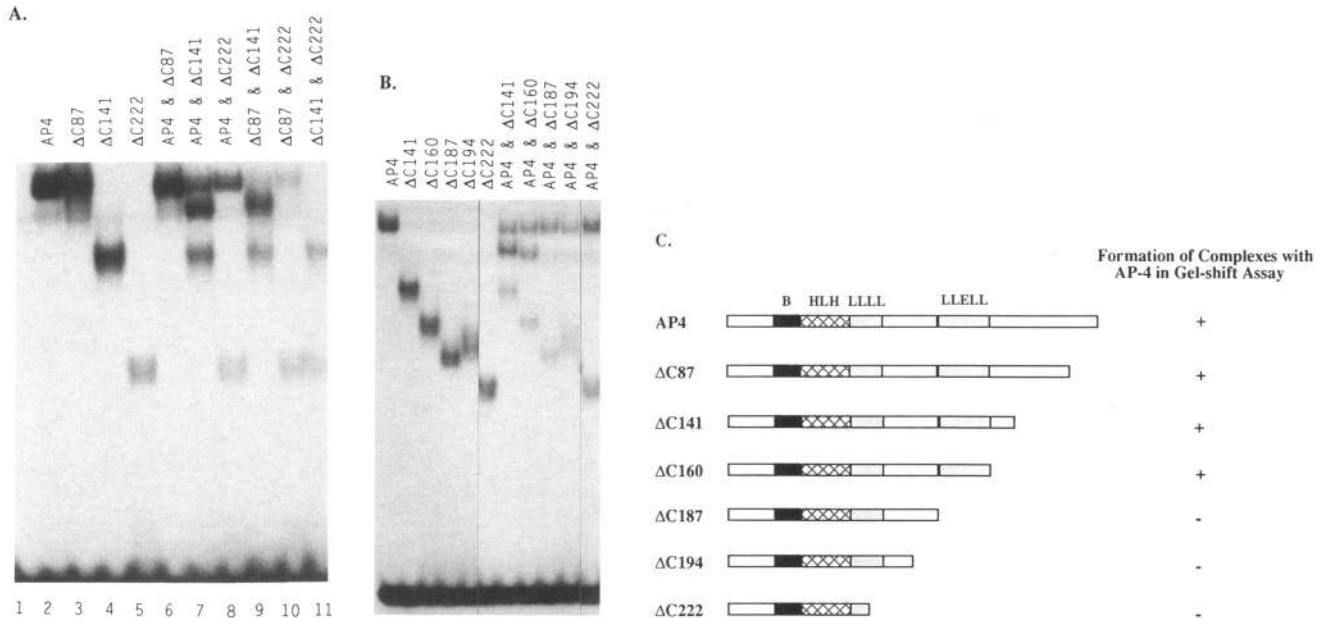

Figure 5. Dimerization properties of in vitro-translated AP-4 and its derivatives analyzed by EMSA. (*A*) Gel-shift reactions were carried out with reticulocyte lysates containing no RNA (lane *1*); in vitro-translated AP-4, ΔC87, ΔC141, and ΔC222 proteins, respectively (lanes *2–5*); AP-4 and ΔC87 cotranslation products (lane *6*); AP-4 and ΔC141 cotranslation products (lane *7*); AP-4 and ΔC222 cotranslation products (lane *8*); ΔC87 and ΔC141 cotranslation products (lane *9*); ΔC87 and ΔC222 cotranslation products (lane *10*); ΔC141 and ΔC222 cotranslation products (lane *11*). The expected difference in the mobility of AP-4 and ΔC87 complexes is not observed in this gel system. The acidic carboxy-terminal tail of AP-4 may be responsible for its faster mobility in this gel system. (*B*) Role of LR1 and LR2 in dimer formation. Additional carboxy-terminal deletions were generated by digestion of AP-4 template DNA with BAL-31 nuclease. The dimerization between the longest AP-4 and its truncated derivatives was analyzed by EMSA. In this assay, ΔC187 formed a faster migrating complex than ΔC194. This may be due to the absence of acidic residues at the carboxy-terminal end of the ΔC194 protein. (*C*) Representation of the AP-4 derivatives. The ability of each form to participate in complex formation with AP-4 is summarized at *right*.

cated products were assayed for dimer formation by EMSA (Fig. 5B). The results are summarized in Figure 5C. Mutant ΔC160, like ΔC141, is able to form a heterodimer with full-length AP-4. In contrast, ΔC187 and ΔC194 behave like ΔC222 and bind DNA specifically but do not form an intermediate band in the EMSA in the presence of the longest AP-4. Moreover, a point mutation in LR1, in which the second leucine has been replaced with proline, had no measurable effect on dimerization (data not shown). We conclude from these studies that the carboxy-terminally located leucine repeat, LR2, is essential for dimerization of AP-4 in solution and that LR1 is not required for dimerization when LR2 is present.

To address the question of whether LR1 could also participate in dimerization if we removed LR2, we performed EMSA and cross-linking experiments using ΔC187, which lacks LR2. Figure 6A shows that cotranslation of ΔC187 and ΔC222 does not result in the formation of heteromeric complexes, as determined by the EMSA. However, ΔC187 can form homodimers in the absence of DNA, as determined by chemical cross-linking (Fig. 6B). These results indicate that in the absence of LR2, LR1 assumes the responsibility for dimer formation in solution. Taken together with our other findings, we conclude that both LR1 and LR2 can contribute to the formation of stable AP-4 homodimers. However, LR2-mediated dimerization takes precedence over LR1-mediated dimerization.

AP-4 lacking the leucine repeats can heterodimerize with E12

Earlier, we performed EMSA with AP-4 and other HLH proteins such as E12 to test whether AP-4 also participates in the HLH regulatory network. We found that AP-4 is unable to form complexes with E12 (data not shown). This was puzzling because it has been well-documented that an intact HLH domain is usually sufficient for dimer formation. In light of the results concerning the dominant dimerization activity of the leucine repeats in AP-4, we were prompted to reevaluate the intrinsic capacity of the AP-4 HLH to form heterologous dimers with E12. In particular, we wished to test the possibility that in the absence of LR1 and LR2, the HLH motif of AP-4 would be sufficient to direct heterodimers with E12. Analysis of complexes by EMSA reveals that, unlike the behavior of other AP-4 derivatives, ΔC222 can indeed form a complex with the HLH protein E12. Moreover, ΔC222 and E12 can form heterocomplexes on DNA even after mixing of the individually translated products (Fig. 6C). These results suggest that the HLH motif of AP-4 is intrinsically capable of dimer formation with other HLH proteins, at least in the presence of DNA-binding sites. However, in the presence of the leucine repeats, AP-4 is unable to heterodimerize with E12. Apparently, the dominant effect of LR2 and LR1 in AP-4 is to preclude dimerization interactions with molecules that only contain HLH domains and, instead, favors homodimer formation of full-length molecules containing multiple interfaces. When LR2 is detected, the truncated protein can form homodimers using the HLH motif and LR1. When both LR1 and LR2 are deleted, leaving only the HLH intact, then the truncated AP-4 protein can complex with heterologous HLH proteins on DNA. Because a mutant LR1 (PML1) has the same dimerization potential as wild-type AP-4, we conclude that LR1 is not essential and that LR2 is the dominant dimerization motif. Thus, there may be a hierarchy of protein–protein interactions directed by multiple dimerization motifs in AP-4.

Discussion

Here, we report the molecular cloning of transcription factor AP-4 and the characterization of its DNA binding and dimerization properties. A novel feature of AP-4 is the presence of three separate and functional dimerization domains, which include an HLH motif and two distinct leucine repeat elements, LR1 and LR2. Analyses of mutant AP-4 molecules in vitro indicate that the HLH and adjacent basic domain are necessary and sufficient to direct sequence-specific DNA binding to its recognition element, CAGCTG. However, AP-4 requires LR2 to form a stable homodimer in solution. Interestingly, LR1 provides dimerization function if LR2 is deleted. If both leucine repeats are deleted, the HLH can now form a heterodimer with other HLH proteins in the presence of target DNA, whereas the longer AP-4 can form only homodimers (Fig. 7). Therefore, AP-4 contains three separate motifs that can participate in dimerization activity. Why does AP-4 have multiple functional dimerization domains? A priori, it seemed reasonable to suggest that one or more of these potential protein–protein interfaces may play a role in directing the formation of tetramers, or higher order structures. Thus far, a variety of experimental techniques, including chemical cross-linking and EMSAs, have failed to provide any evidence for higher-order structures other than dimers. We have not, at present, excluded the possibility that dimers of AP-4 dimers can form under some conditions, but have no compelling reason to postulate their existence.

An alternative hypothesis for the presence of three distinct dimerization elements in AP-4 is that this transcription factor has evolved to interact selectively with a variety of different partners and therefore would provide a large number of possible hetero-multimer configurations, in which each combination may give rise to unique DNA binding or transcriptional properties, perhaps generating tissue specificity in the resulting dimers. For example, it is conceivable that AP-4 could form stable oligomers with members of the leucine repeat family of transcription factors as well as the growing family of the cell-type-specific and ubiquitous HLH proteins. Systematic mutagenesis of dimerization domains indeed reveals that heterodimers between a truncated AP-4 molecule and another HLH protein, E12, can form, when the leucine repeat elements, LR1 and LR2, are deleted. In contrast, full-length AP-4 is incapable of forming heterodimers either with E12 and other HLH proteins, or with B-zip proteins such as Jun, Fos,

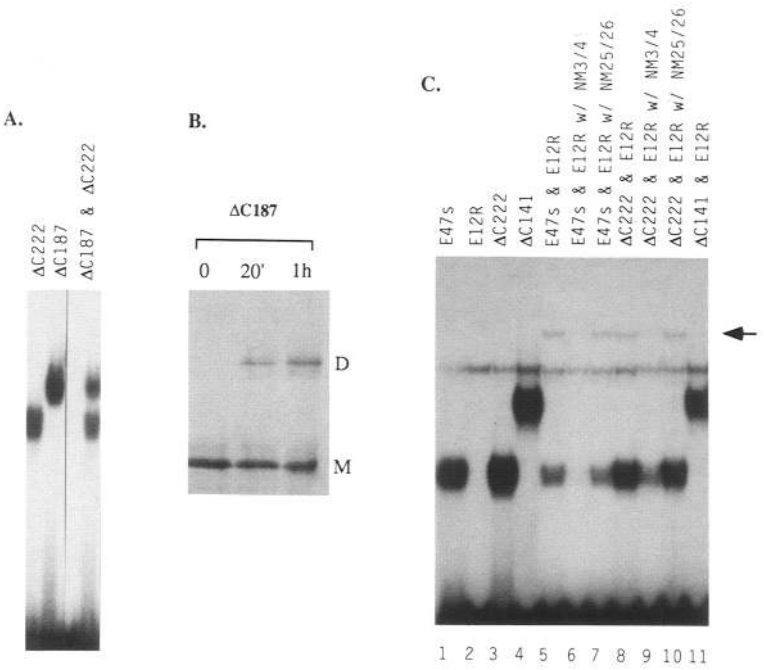

Figure 6. Characterization of the role of LR1 and HLH in dimerization. (*A*) Cotranslated AP-4 derivatives ΔC187 and ΔC222 do not dimerize with each other as analyzed by EMSA. ΔC187 and ΔC222 were translated either separately (ΔC187, ΔC222) or together in the same reaction mixture (ΔC187 & ΔC222) and the translation reactions were tested in the mobility-shift assay. (*B*) Glutaraldehyde cross-linking of ΔC187. In vitro-translated ΔC187 was purified by DNA affinity chromatography and treated with no glutaraldehyde (0), 0.001% glutaraldehyde for 20 min (20'), or 1 hr (1h) and the products were resoved on a 18% SDS-PAGE. (M) Monomer; (D) dimer. (*C*) Complex formation between AP-4 derivatives and E12/E47. E47S (lane *1*), E12R (lane *2*), ΔC222 (lane *3*), and ΔC141 (lane *4*) were translated independently and analyzed by EMSA. (Lanes *5–7*) E12R and E47S were cotranslated in the same reaction mixture. The EMSA was carried out with no specific competitor, 100 ng of unlabeled NM3/4, and 100 ng of NM25/26, respectively. NM25/26 differs from NM3/4 by 2 bp in the binding site (CAGCTG → CAGCGT). In lanes *8–11*, the two proteins indicated above each lane were generated separately but mixed with the DNA probe in the mobility-shift assay. No specific competitor (lane *8*), 100 ng of NM3/4 (lane *9*), or 100 ng of NM25/26 (lane *10*) were included in EMSA. Same results were obtained when the two proteins were cotranslated in the same reaction mixture (data not shown). The arrow at *right* corresponds to the heterodimer positions of E12R/E47S and E12R/ΔC222. E12R is the full-length E12 and has no detectable binding activity (Murre et al. 1989b); E47S is a derivative of E47 containing B/HLH region (Murre et al. 1989a).

GCN4, etc. (data not shown). Instead, our findings, taken in toto, suggest that the presence of multiple protein interfaces serves to restrict the formation of heterodimer complexes. Apparently, homodimers held together by three separate protein interfaces generate an extremely stable complex into which truncated AP-4 molecules or related factors containing only one or two dimerization motifs are unable to penetrate. Consequently, we are prompted to conclude that the major purpose of tandem dimerization domains within a single polypeptide serves to direct selective complex formation and prevent the incorrect combinations from occurring. This feature may be especially important for transcription factors such as AP-4, where its presence in several different cell types may severely interfere with the activity of other HLH or leucine repeat factors.

Although other DNA-binding proteins with a similar tripartite dimerization configuration have not been identified yet, it is conceivable that there exists in the cell one or more proteins that would readily pair with AP-4 through interactions with the proper combination of HLH and leucine repeats in the appropriate spatial register, to form stable heterodimers. For example, LR1 may have a more profound role than merely contributing to the homodimer stability in the intact AP-4 molecule. One could imagine that in vivo, if LR2 is inactivated by post-translational modification, specific protease digestion or other mechanisms, AP-4 may display a different dimerization specificity through the HLH and LR1. In other words, the multiple dimerization interfaces in AP-4 may not only offer a secure mechanism for homodimer formation but also provide a potential for more elaborate regulation and specificity. When antibodies directed against AP-4 become available, it should be possible to detect these putative heteromeric complexes by coimmunoprecipitation because these heterodimers are expected to be stable in solution.

The presence of multiple dimer interfaces brings to mind another noteworthy aspect of AP-4 structure. In most other cases of bona fide transcription factors that have been studied, HLH or leucine repeat domains have been associated with a basic DNA-binding domain. In fact, almost without exception, the dimerization domain of factors such as Jun, Fos, C/EBP, MyoD, TFE3, etc., comprise the structural framework for orienting the DNA-binding domain. Thus, disruption of the dimer interface simultaneously obliterates a functional DNA-binding domain. In the case of intact AP-4, the HLH dimer domain indeed appears to provide just such a scaffolding structure for the adjacent DNA-binding region. However, unlike other well-characterized B-zip transcription factors, the leucine repeats of AP-4 are not physically linked to the DNA-binding domain and, in fact, are totally dispensable for DNA recognition. Thus, the leucine repeats of AP-4 define a new function, which is independent for DNA binding and instead serves

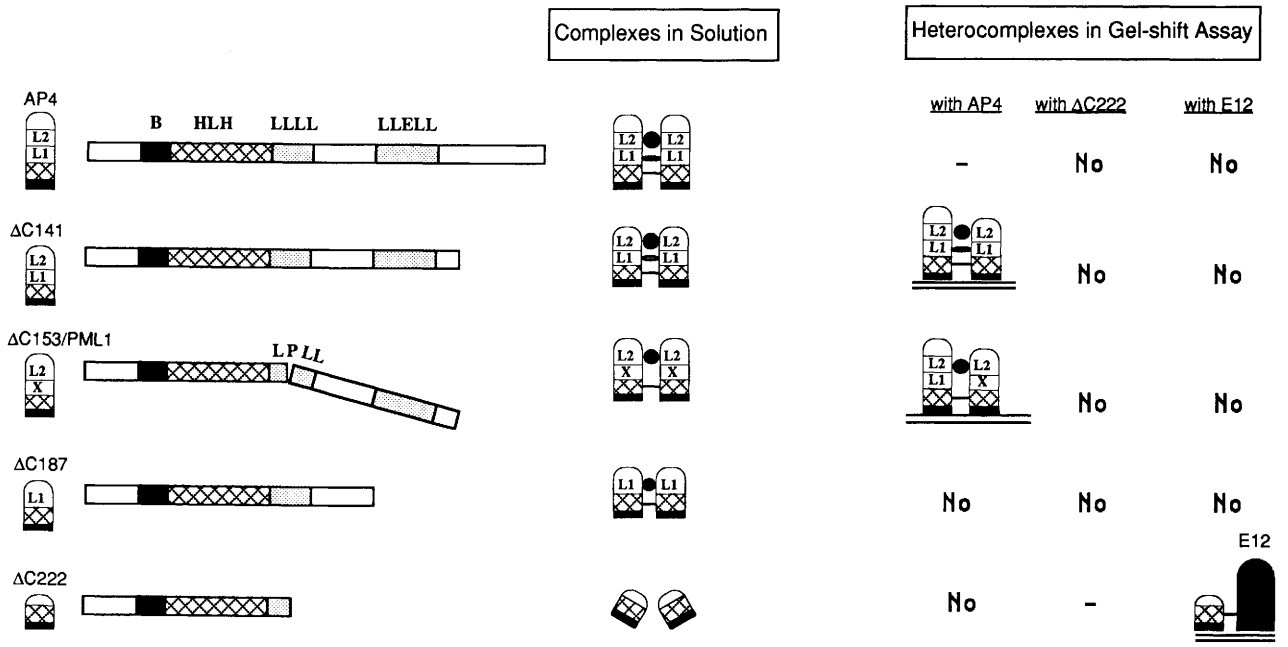

Figure 7. Summarization of the dimerization properties of AP-4 and its derivatives. Circular, oblong, and linear connections between two monomers represent the protein–protein interfaces.

strictly as a dimer interface. It is not difficult to imagine that some sequence-specific transcription factors may have weak B-zip DNA-binding domains and use strong HLH motifs as secondary dimerization interfaces. Recently, a number of putative regulatory transcription factors have been identified that contain a functional HLH dimerization domain but no associated basic DNA recognition sequences (Benezra et al. 1990; Ellis et al. 1990). Thus, the general principle that different classes of protein–protein interfaces can be used to specify dimerization for purposes other than generating DNA-binding sites has become more established.

It is intriguing that LR1 and LR2 represent two fairly divergent types of leucine repeats. LR1, which is directly adjacent to the HLH motif, is reminiscent of the structures found in Myc and TFE3 (Battey et al. 1983; DePinho et al. 1987; Beckmann et al. 1990). Despite the superficial similarity between Myc and AP-4, we have not detected any heterodimer formation between these two molecules. In contrast, LR2, which is not directly connected to the HLH region, resembles a more conventional Leu repeat, although one of the five Leu residues has been replaced by a glutamic acid residue. It has been shown that a c-Fos variant, which contains an Arg in place of one of the internal Leu residues, still dimerizes with Jun, suggesting that hydrophilic amino acids can be tolerated in the coiled-coil interface (Kouzarides and Ziff 1988). Furthermore, Lys residues are often found at the 3-4 position of myosin coiled coils, presumably because the methylene groups of the Lys side chain can make hydrophobic contacts while the charge is exposed to solvent (O'Shea et al. 1989). A similar replacement, Asn for Leu, was observed for one of the yeast AP-1 proteins (Moye-Rowley et al. 1989). Interestingly, when the Asn residue of yAP-1 was mutated to a Leu, DNA binding was abrogated rather than enhanced (C. Parker, pers. comm.). This finding suggests that the placement of a residue other than Leu (i.e., Asn or Glu) at this position may actually play some role in the structural integrity and function of the protein. It is therefore possible that the glutamic acid residue in LR2 is also important for dimerization specificity and/or affinity. In this respect, it is worth pointing out that the only nonhydrophobic residue at the fourth position in LR2, an Arg residue, occurs next to the Glu residue on the putative helix. The positioning of these opposite charged residues, when paired in a coiled-coil fashion, may help decrease charge repulsion between Leu repeats. Additional experiments will be required to assess the significance of these different Leu repeat structures.

In this study, we have not addressed the important question of the transcriptional properties of AP-4. However, our curiosity about the role of AP-4 in animal cells has become heightened by the unusual organization of multiple dimerization motifs and the presence of multiple potential activation domains, which include Gln-rich, Pro-rich, and acidic domains located toward the carboxyl terminus. We are also intrigued by the recent finding that AP-4 mRNA expression correlates with the exocrine function of pancreas cell lines (B. Lüscher, unpubl.). This latter finding is also consistent with a recent report suggesting that an AP-4-like molecule, in combination with another cell-type-specific regulatory factor, could direct the tissue-specific transcription of α-amylase in pancreatic cells (Roux et al. 1989; Fodor et al., unpubl.). The previously documented ability of AP-4 to synergistically activate transcription in vitro with other factors such as AP-1 (Mermod et al. 1988) suggests that

specific protein–protein interactions may be crucial to its role as a cell-type-specific regulator. These important issues will be the subject of subsequent studies.

Materials and methods

Purification of AP-4 and peptide sequencing

Human AP-4 was purified by combination of conventional chromatography and sequence-specific DNA affinity chromatography as described in Mermod et al. (1988) with substitution of lauryldimethylamine oxide (LDAO, Calbiochem) for NP-40. The purified proteins consist of two major polypeptides of 116 and 48 kD as judged by SDS-PAGE and silver staining. To remove the 116-kD contaminant, the proteins were precipitated with trichloroacetic acid (TCA) and washed twice with acetone, followed by resuspension in 8% triethylmamine (TEA) and incubation for 30 min at 65°C. Tris-HCl (pH 7.4) was added to 5 mM final concentration, and the pH was adjusted to 8.5 with 10% acetic acid and incubated on ice for 30 min. The protein mix was spun in a Beckman TL100 ultracentrifuge for 30 min at 30,000 rpm, 4°C, and the supernatant, which was free of the 116-kD protein, was digested overnight with trypsin at 37°C. The peptides were resolved by reverse-phase HPLC on a C_{18} column as described in Bohmann et al. (1987). The peptides were collected directly onto the glass fiber filters that had been premodified with polybrene and precycled on the Peptide Sequencer. Peptide sequencing was performed on a 477A Protein Sequencer (Applied Biosystems).

Isolation of a recombinant clone encoding AP-4

Two oligonucleotides, 5′-GCCCTCIATGTGCTGGATGGCCTGCACIATGGTGTCIAGGTTCTG-3′ and 5′-CAGGTGCAGCTGCAGCAGCAGCAGGAGCAG-3′, were designed on the basis of the sequences of two tryptic peptides, QNLDTIVQAIQHIEG and QVQLQQQQEQ of AP-4, according to Lathe (1985). The 5′-end-labeled oligonucleotides were used to screen a HeLa cell cDNA library in λgt10 as described by Kadonaga et al. (1987), except that the final washes of the nitrocellulose filters after the hybridization were done at 52°C. Two positives were obtained from the first screening of one million phage plaques. An internal *Eco*RI site was present in the insert of the larger recombinant clone, and the larger *Eco*RI fragment of this insert was used to screen another 10^6 phage plaques in the HeLa cell cDNA library. The inserts of four independent clones were subcloned into the *Eco*RI site of pBluescriptSK$^+$ (Stratagene). Nested deletions in the cDNAs were then prepared according to the manufacturer's instructions (Stratagene). Deletion mutants containing overlapping segments of the cDNAs were sequenced on both strands according to Sanger (1977). The four independent cDNA clones showed extensive sequence overlaps. The two longest cDNAs were fused at a common *Pst*I restriction site to give the AP-4 cDNA (pBSAP4) shown in Figure 1A.

Expression of AP-4 in HeLa cells by vaccinia virus expression vector

To construct a vaccinia virus expression vector carrying AP-4 sequences, the pBSAP4 DNA was cleaved with *Hin*dIII and *Sma*I, the ends were filled-in with Klenow fragment, and the insert was purified and subcloned in-frame into a *Sma*I-cut vaccinia expression vector (pAbT4537, kindly provided by Applied BioTechnology, Inc.). Recombinant virus was generated as described (Chakrabarti et al. 1985). HeLa cells were infected with the recombinant virus, and nuclear extracts were prepared as described (Dynlacht et al. 1989). As a control, HeLa cells were infected in parallel with a New York City Board of Health strain of vaccinia (kindly provided by Applied BioTechnology, Inc.). The crude nuclear extract was first partially purified by heparin-agarose chromatography essentially as described by Perkins et al. (1988), and the 0.8 M KCl eluate was purified further by DNA affinity chromatography as described previously (Mermod et al. 1988).

In vitro transcription and translation

The in vitro transcription/translation vector pT7βAP4 was constructed by ligation of the *Eco*RI fragment of the AP-4 cDNA into the *Nco*I–*Hin*cII sites of T7βSalI [a pGEM4 derivative containing the β-globin leader sequence downstream of the T7 polymerase promoter, kindly provided by Norman and Treisman (1988)] in the proper reading frame. The deletion mutants were generated by cutting of the construct at the various restriction sites or by digestion of the DNA with BAL-31 nuclease from the only *Pst*I site (at ΔC141 position) and subsequent ligation of *Pst*I linker. About 2 μg of linearized DNA template was transcribed in vitro in a 50-μl reaction for 30 min at 37°C, using T7 RNA polymerase in the presence of RNasin. The resulting RNAs were purified by phenol extraction and ethanol precipitation. In vitro translations were carried out in rabbit reticulocyte lysates (Promega) according to the manufacturer's instructions in the presence of [^{35}S]methionine (see also Turner and Tjian 1989).

DNA-binding assay

DNase I footprinting assay was carried out as described by Mermod (1988) except that the probe used contained proenkephalin promoter sequences.

EMSA was performed as follows: A double-stranded oligonucleotide containing an AP-4 site, designed NM3/4 (top strand 5′-GATCACCAGCTGTGGAATGTGTGT-3′; bottom strand 5′-GATCACACACATTCCACAGCTGGT-3′) was labeled at both ends by filling in with Klenow fragment in the presence of [α-^{32}P]dGTP. Labeled DNA probe (0.4 ng; 10^4 cps) was incubated with 4 μl of the translated or untranslated reticulocyte lysate. The binding reactions were carried out at room temperature for 15 min in the presence of 20 mM HEPES-KOH (pH 7.9), 50 mM KCl, 4 mM $MgCl_2$, 4 mM spermidine, 0.2 mM EDTA, 0.5 mM DTT, 0.05% (vol/vol) NP-40, and 20% (vol/vol) glycerol. Poly[d(I-C)] (500 ng) was added as nonspecific competitor DNA, unless otherwise indicated in the figure legends.

Glutaraldehyde cross-linking analysis

In vitro-translated proteins were purified by DNA affinity chromatography and eluted with buffer Z [25 mM HEPES-KOH (pH 7.5), 12.5 mM $MgCl_2$, 20% (vol/vol) glycerol, 0.1% NP-40, 1 mM DTT, and final pH adjusted to 7.7] to which 1 M KCl was added as described by Turner and Tjian (1989). When the cross-linking reactions were performed in the crude reticulocyte extract, the samples were diluted 15-fold prior to the addition of glutaraldehyde. Equal portions of the eluates or crude extract were treated with 0.001% glutaraldehyde for the amount of time indicated in the figures. Samples were then precipitated with TCA and subjected to electrophoresis on 12 or 18% (wt/vol) polyacrylamide–SDS gels, and the dried gels were fluorographed.

Acknowledgments

We thank Richard Turner for invaluable advice. We thank

Craig Hauser for providing the human cDNA (λgt10) library, Cornelis Murre for providing E12R/E47S plasmids, and Bobbi Johnson for help in generating recombinant vaccinia viruses. We thank Bruce Kimmel, Al Courey, Greg Peterson, and Trevor Williams for assistance with computer programs. We are also grateful to Karen Ronan for typing this manuscript and preparing the figures. We appreciate the critical reading of this manuscript by Vijay Baichwal, Grace Gill, Richard Turner, Frank Pugh, Michael Jantzen, Naoko Tanese, Matija Peterlin, and other members of the Tjian lab. B.L. and N.M. were supported by fellowships from the Swiss National Science Foundation. This work was supported in part by a grant from the National Institutes of Health to R.T.

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

Note added in proof

The sequence data of AP-4 have been submitted to the EMBO/GenBank data libraries.

References

Battey, J., C. Moulding, R. Taub, W. Murphy, T. Stewart, H. Potter, G. Lanoir, and P. Leder. 1983. The human c-myc oncogene structural consequences of translocation into the IgH locus in Burkitt lymphoma. *Cell* **34:** 779–787.

Beckmann, H., L. Su., and T. Kadesch. 1990. TFE3: A helix–loop–helix protein that activates transcription through the immunoglobulin enhancer μE3 motif. *Genes Dev.* **4:** 167–179.

Benezra, R., R.L. Davis, D. Lockshon, D.L. Turner, and H. Weintraub. 1990. The protein Id: A negative regulator of helix–loop–helix DNA binding proteins. *Cell* **61:** 49–59.

Bohmann, D., T.J. Bos, A. Admon, T. Nishimura, P.K. Vogt, and R. Tjian. 1987. Human proto-oncogene c-jun encodes a DNA binding protein with structural and functional properties of transcription factor AP-1. *Science* **238:** 1386–1392.

Caudy, M., H. Vassin, M. Brand, R. Tuman, L.Y. Jan, and Y.N. Jan. 1988. *daughterless*, a *Drosophila* gene essential for both neurogenesis and sex determination, has sequence similarities to *myc* and the *achaete-scute* complex. *Cell* **55:** 1061–1067.

Chakrabarti, S., K. Brechling, and B. Moss. 1985. Vaccinia virus expression vector: Coexpression of β-galactosidase provides visual screening of recombinant virus plaques. *Mol. Cell. Biol.* **5:** 3403–3409.

Clerc, R.G., L.M. Corcoran, J.H. LeBowitz, D. Baltimore, and P.A. Sharp. 1988. The B-cell-specific Oct-2 protein contains POU box- and homeo box-type domains. *Genes Dev.* **2:** 1570–1581.

Comb, M., N. Mermod, S.E. Hymann, J. Pearlberg, M.E. Ross, and H.M. Goodman. 1988. Proteins bound at adjacent DNA elements act synergistically to regulate human proenkephalin cAMP inducible transcription. *EMBO J.* **7:** 3793–3805.

Courey, A.J., D.A. Holtzman, S.P. Jackson, and R. Tjian. 1989. Synergistic activation by the glutamine-rich domains of human transcription factor Sp1. *Cell* **59:** 827–836.

Davis, R.L., H. Weintraub, and A.B. Lassar. 1987. Expression of a single transfected cDNA converts fibroblasts to myoblasts. *Cell* **51:** 987–1000.

Davis, R.L., P.-F. Cheng, A. Lassar, and H. Weintraub. 1990. The MyoD DNA-binding domain contains a recognition code for muscle-specific gene activation. *Cell* **60:** 733–746.

DePinho, R., K.S. Hatton, A. Tesfaye, G.D. Yancopoulos, and F.W. Alt. 1987. The human *myc* gene family: Structure and activity of L-myc and an L-myc pseudogene. *Genes Dev.* **1:** 1311–1326.

Dynlacht, B.D., L.D. Attari, A. Admon, M. Freeman, and R. Tjian. 1989. Functional analysis of NTF-1, a developmentally regulated *Drosophila* transcription factor that binds neuronal *cis* elements. *Genes Dev.* **3:** 1677–1688.

Edmonson, D.G. and E.N. Olson. 1989. A gene with homology to the *myc* similarity region of MyoD1 is expressed during myogenesis and is sufficient to activate the muscle differentiation program. *Genes Dev.* **3:** 628–640.

Ellis, H.M., D.R. Spann, and J.W. Posakony. 1990. Extramacrochaete, a negative regulator of sensory organ development in Drosophila, defines a new class of helix–loop–helix proteins. *Cell* **61:** 27–38.

Gentz, R., F.J. Rauscher, C. Abate, and T. Curran. 1989. Parallel association of Fos and Jun leucine zippers juxtaposes DNA-binding domains. *Science* **243:** 1695–1699.

Henthorn, P., M. Kiledjian, and T. Kadesch. 1990. Two distinct transcription factors that bind the immunoglobulin enhancer μE5/κE2 motif. *Science* **247:** 467–470.

Hinnebusch, A.G. 1984. Evidence for translational regulation of the activator of general amino acid control in yeast. *Proc. Natl. Acad. Sci.* **81:** 6442–6446.

Hoey, T. and M. Levine. 1988. Homeobox proteins as sequence-specific transcription factors. *Cell* **55:** 537–540.

Hope, I.A. and K. Struhl. 1987. GCN4, a eukaryotic transcriptional activator protein, binds as a dimer to target DNA. *EMBO J.* **6:** 2781–2784.

Kadonaga, J.T. and R. Tjian. 1986. Affinity purification of sequence-specific DNA binding proteins. *Proc. Natl. Acad. Sci.* **83:** 5889–5893.

Kadonaga, J.T., K.R. Carner, F.R. Masiarz, and R. Tjian. 1987. Isolation of cDNA encoding transcription factor Sp1 and functional analysis of the DNA-binding domain. *Cell* **51:** 1079–1090.

Kouzarides, T. and E. Ziff. 1988. The role of the leucine zipper in the fos–jun interaction. *Nature* **336:** 646–651.

Landschulz, W.H., P.F. Johnson, E.Y. Adashi, B.J. Graves, and S.L. McKnight. 1988a. Isolation of a recombinant copy of the gene encoding C/EBP. *Genes Dev.* **2:** 786–800.

Landschulz, W.H., P.F. Johnson, and S.L. McKnight. 1988b. The leucine zipper: A hypothetical structure common to a new class of DNA binding proteins. *Science* **240:** 1759–1764.

———. 1989. The DNA-binding domain of the rat liver nuclear protein C/EBP is bipartite. *Science* **243:** 1681–1688.

Lassar, A.B. et al. 1989. MyoD is a sequence-specific DNA binding protein requiring a region of *myc* homology to bind to the muscle creatine kinase enhancer. *Cell* **58:** 823–831.

Lathe, R.J. 1985. Synthetic oligonucleotide probes deduced from amino acid sequence data—Theoretical and practical considerations. *J. Mol. Biol.* **183:** 1–12.

Ludwig, et al. 1989. Lc, a member of the maize R gene family responsible for tissue-specific anthocyanin production, encodes a protein similar to transcriptional activators and contains the myc-homology region. *Proc. Natl. Acad. Sci.* **86:** 7092–7096.

Mellentin, J.D., S.D. Smith, and M. Cleary. 1989. lyl-1, a novel gene altered by chromosomal translocation in T cell leukemia, codes for a protein with a helix-loop-helix DNA binding motif. *Cell* **58:** 77–83.

Mermod, N., T.J. Williams, and R. Tjian. 1988. Enhancer binding factors AP-4 and AP-1 act in concert to activate

SV40 late transcription in vitro. *Nature* **332:** 557–561.
Mermod, N., E.A. O'Neill, T.J. Kelly, and R. Tjian. 1989. The proline-rich transcriptional activator of CTF/NF-1 is distinct from the replication and DNA-binding domain. *Cell* **58:** 741–753.
Miller, J., A.D. McLachlan, and A. Klug. 1985. Repetitive zinc-binding domains in the protein transcription factor IIIA from *Xenopus* oocytes. *EMBO J.* **4:** 1609–1614.
Mitchell, P.J. and R. Tjian. 1989. Transcriptional regulation in mammalian cells by sequence-specific DNA binding proteins. *Science* **245:** 371–378.
Moye-Rowley, W.S., K.D. Harshman, and C.S. Parker. 1989. Yeast YAP1 encodes a novel form of the *jun* family of transcriptional activator proteins. *Genes Dev.* **3:** 283–292.
Müller, M.M. et al. 1988. A cloned octomer transcription factor stimulates transcription from lymphoid-specific promoter in non-B cells. *Nature* **336:** 544–551.
Murre, C., P.S. McCaw, and D. Baltimore. 1989a. A new DNA binding and dimerization motif in immunoglobulin enhancer binding, *daughterless*, MyoD, and myc proteins. *Cell* **56:** 777–783.
Murre, C., S. McCaw, H. Vassin, M. Caudy, L.Y. Jan, Y.N. Jan, C.V. Cabrera, J.N. Buskin, S. Hauschka, A.B. Lassar, H. Weintraub, and D. Baltimore. 1989b. Interactions between heterologous helix–loop–helix proteins generate complexes that bind specifically to a common DNA sequence. *Cell* **58:** 537–544.
Norman, C., M. Runswick, R. Pollock, and R. Treisman. 1988. Isolation and properties of cDNA clones encoding SRF, a transcription factor that binds to the c-fos serum response element. *Cell* **55:** 989–1003.
O'Shea, E.K., R. Rutkowski, W.F. Stafford, and P. Kim. 1989. Preferential heterodimer formation by isolated leucine zippers from Fos and Jun. *Science* **245:** 646–648.
Perkins, K.K., G.M. Dailey, and R. Tjian. 1988. In vitro analysis of the *Antennapedia* P2 promoter: Identification of a new *Drosophila* transcription factor. *Genes Dev.* **2:** 1615–1626.
Ptashne, M. 1988. How eukaryotic transcriptional activators work. *Nature* **335:** 683–689.
Rogers, S., R. Wells, and M. Rechsteiner. 1986. Amino acid sequences common to rapidly degraded proteins: The PEST hypothesis. *Science* **234:** 364–368.
Roux, E., M. Strubin, O. Hagenbüchle, and P.K. Wellaher. 1989. The cell-specific transcription factor PTF1 contains two different subunits that interact with the DNA. *Genes Dev.* **3:** 1613–1624.
Sanger, F., S. Nicklen, and A.R. Coulson. 1977. DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci.* **81:** 7742–7746.
Sturm, R.A., G. Das, and W. Herr. 1988. The ubiquitous octamer-binding protein Oct-1 contains a POU domain with a homeo box subdomain. *Genes Dev.* **2:** 1582–1599.
Tanaka, M. and W. Herr. 1990. Differential transcriptional activation by Oct-1 and Oct-2: Interdependent activation domains induce Oct-2 phosphorylation. *Cell* **60:** 375–386.
Turner, R. and R. Tjian. 1989. Leucine repeats and an adjacent DNA-binding domain mediate the formation of functional cFos–cJun heterodimers. *Science* **243:** 1689–1694.
Villares, R. and C.V. Cabrera. 1987. The achaete-scute gene complex of D. melanogaster: Conserved domains in a subset of genes required for neurogenesis and their homology to myc. *Cell* **50:** 415–424.
Vogt, P.K., T.J. Bos, and R.F. Doolittle. 1987. Homology between the DNA-binding domain of the GCN4 regulatory protein of yeast and the carboxy-terminal region of a protein coded for by the oncogene jun. *Proc. Natl. Acad. Sci.* **84:** 3316–3319.
Wright, W.E., D.A. Sassoon, and V.K. Lin. 1989. Myogenin, a factor regulating myogenesis, has a domain homologous to MyoD. *Cell* **56:** 607–617.

Transcription factor AP-4 contains multiple dimerization domains that regulate dimer specificity.

Y F Hu, B Lüscher, A Admon, et al.

Genes Dev. 1990, **4:**
Access the most recent version at doi:10.1101/gad.4.10.1741