MOLECULAR AND CELLULAR BIOLOGY, Jan. 1996, p. 121–129 Vol. 16, No. 1
0270-7306/96/$04.00+0

The AD1 Transactivation Domain of E2A Contains a Highly Conserved Helix Which Is Required for Its Activity in both *Saccharomyces cerevisiae* and Mammalian Cells

MARK E. MASSARI,[1] PATRICIA A. JENNINGS,[2] AND CORNELIS MURRE[1]*

Departments of Biology[1] and Chemistry and Biochemistry,[2] University of California, San Diego, La Jolla, California 92093

Received 7 July 1995/Returned for modification 16 August 1995/Accepted 11 October 1995

A conserved region, designated the AD1 domain, is present in a class of helix-loop-helix (HLH) proteins, E proteins, that includes E12, E47, HEB, E2-2, and a *Xenopus laevis* HLH protein closely related to E12. We demonstrate that the AD1 domain in E2A and the conserved region of E2-2 activate transcription in both yeast and mammalian cells. The AD1 domain contains a highly conserved putative helix that is crucial for its transactivation properties. Circular dichroism spectroscopy data show that AD1 is structured and contains distinctive helical properties. In addition, we show that a synthetic peptide corresponding to the conserved region is unstructured in aqueous solution at neutral pH but can adopt an α-helical conformation in the presence of the hydrophobic solvent trifluoroethanol. Amino acid substitutions that destabilize the helix abolish the transactivation ability of the AD1 domain. Both structural and functional analyses of AD1 reveal striking similarities to the acidic class of activators. Remarkably, when wild-type and mutant proteins are expressed in mammalian cells and *Saccharomyces cerevisiae*, identical patterns of transactivation are observed, suggesting that the target molecule is conserved between *S. cerevisiae* and mammals. These data show that transactivation by E proteins is mediated, in part, by a strikingly conserved peptide that has the ability to form a helix in a hydrophobic solvent. We propose that the unstructured domain may become helical upon interaction with its cellular target molecule.

A new class of helix-loop-helix (HLH) proteins, the E proteins, that includes E12, E47, E2-2, HEB, and daughterless has recently emerged (for a review, see reference 29). Each of these proteins is expressed in many different lineages, and each of them has the ability to bind either as a homo- or as a heterodimer to DNA (29, 30). They are closely related in their DNA-binding and dimerization domain, the HLH motif (29, 30). Two E proteins, E12 and E47, are encoded by one gene, the *E2A* gene, and arise through differential splicing (31, 44). E12 and E47 are involved in many different developmental pathways, including B-cell differentiation, myogenesis, and pancreatic development (9, 25, 33, 42). In B cells, E2A products bind as homodimers to E box elements present in both the immunoglobulin heavy- and light-chain gene enhancers (33). Studies with E2A knockout mice have demonstrated that E2A proteins are required for B-cell development (4, 50). In muscle-derived cells, E2A molecules form heterodimers with myogenic regulators, including MyoD, myogenin, and myf-5 (25, 32). E2-2 is highly expressed in pre-B cells and binds as homodimers to E box sites present in the immunoglobulin heavy- and light-chain gene enhancers (17). HEB is particularly abundant in T cells, and it has been suggested to be involved in the control of CD4 gene expression (18, 41). Daughterless is involved in *Drosophila* sex determination and neurogenesis (5, 6, 8, 11).

The *E2A* gene is also involved in two acute lymphoblastoid leukemias (ALL). In pro-B ALL, the *E2A* gene is translocated to a gene, designated HLF, located on chromosome 17 (19, 20). This translocation event replaces the E2A HLH domain with that of a leucine zipper DNA-binding and dimerization domain (19, 20). In pre-B ALL, involving a t(1;19) translocation, the *E2A* HLH domain is replaced with a homeodomain derived from the *Pbx1* gene (24, 35). The E2A N-terminal domain has been shown previously to be involved in transcriptional activation (2, 17, 38). One transactivation domain, designated the loop-helix or activation domain 2 (AD2), has been well characterized (2, 38). Mutations in this region abolish transactivation. An additional region of E2A, designated AD1, has also been shown to function as an activation domain (2).

Here, we identify a strikingly conserved stretch of amino acids, containing helical structure, that is present in each of the E proteins. We show, using circular dichroism (CD) spectroscopy, that this conserved region of E2A is unstructured in aqueous solution but has distinctive helical properties in a hydrophobic solvent. We further demonstrate that mutations of conserved residues in the helix obliterate E2A-mediated transactivation in both yeast and mammalian cells and propose that the conserved helix plays a crucial role in mediating the transactivation abilities of E proteins.

MATERIALS AND METHODS

Plasmid construction. The sequence encoding amino acids 1 to 99 of E2A was amplified by PCR with a full-length E12 cDNA template and the following primers: 5′ GGG GAA TTC ATG AAC CAG CCG CAG AGG ATG 3′ and 5′ GGG GAA TTC CCG AGT CCC GGT CCC AGG AAT 3′. The resulting *Eco*RI fragment was ligated into plasmid pBXG1 digested with *Eco*RI to create an in-frame fusion with the GAL4 DNA-binding domain. The sequence encoding amino acids 1 to 99 of E2-2 was amplified by PCR with the following primer set: 5′ ATA GCC CGG GAA TGC ATC ACC AAC AGC GAA TG 3′ and 5′ ATA GGG ATC CTC ATC TGG AAT TGA CAA AAG GTG G 3′. The PCR fragment was digested with *Sma*I and *Bam*HI and ligated into pBXG1 which was cut with *Eco*RI (blunt-ended with Klenow fragment) and *Bam*HI. 3′ deletions were generated with a Double Stranded Nested Deletion Kit according to the manufacturer's protocol (Pharmacia Biotech, Inc.).

Site-directed mutagenesis of the AD1 domain was performed with a Double-Take Double-Stranded Mutagenesis Kit as per the manufacturer's instructions (Stratagene). The following oligonucleotides were used for mutagenesis: 5′ GAA GTC CAG GAG G**GG** ACT GAG C**GG** CTT GTC TGT GCC 3′ (mutant 1); 5′

* Corresponding author. Mailing address: Department of Biology, 0366, University of California, San Diego, La Jolla, CA 92093.

GAA CAT CAT GCT G**C**G GTC CAG G**C**G GTC ACT GAG CTC 3′ (mutant 2); 5′ GTC GCC GCT GCC **GT**A GG**C** GCC TGA G**C**T GGG 3′ (mutant 3); 5′ CAG CGG GAA CAT C**GG** GCT GAA G**GG** CAG GAG GTC ACT 3′ (mutant 4); 5′ GTC CTC AAG ACC TG**C** ACC TCC GAA CTG 3′ (mutant 5); and 5′ GCT GGG GTC AAA GG**C** GG**C** GCT GCT CTG GTC 3′ (mutant 6). Mismatched nucleotides are indicated in boldface. All constructs described above were verified by sequencing with a Sequenase version 2.0 DNA sequencing kit in accordance with the manufacturer's instructions (U.S. Biochemical).

Cell culture and transfections. HeLa and COS cells were grown at 37°C in a 5% CO_2-containing atmosphere in Dulbecco's modified Eagle's medium supplemented with 10% bovine calf serum. HeLa cells were transiently transfected by the DEAE-dextran method (40). Cells were harvested after 48 h, and equivalent amounts of protein were assayed for chloramphenicol acetyltransferase (CAT) activity as described before (40).

Western immunoblot analysis. COS cells were plated at a density of 10^6 cells per 10-cm dish and transfected 12 to 24 h later with 10 μg of expression plasmid by the DEAE-dextran method. At 48 h posttransfection, the cells were harvested, and 60 μg of protein extract was fractionated on a 12% polyacrylamide–sodium dodecyl sulfate (SDS) gel, transferred to an Immobilon-P membrane (Millipore), and probed with a monoclonal anti-GAL4 antibody (kindly provided by Rene Bernards). The blots were then probed with an alkaline phosphatase-conjugated secondary antibody. Bands were then visualized by an alkaline phosphatase reaction (Promega). Individual yeast transformants were grown in synthetic medium lacking histidine and harvested at an optical density at 600 nm (OD_{600}) of ~1. Yeast protein extracts were prepared as described before (48), and 28 μg of protein was fractionated on a 15% polyacrylamide–SDS gel. Western blotting was performed as above except that membranes were probed with a horseradish peroxidase (HRP)-conjugated secondary antibody, and proteins were detected by chemiluminescence (Amersham).

Yeast expression plasmids. E2A AD1 wild-type and mutant sequences were excised as *Eco*RI fragments from the pBXG1 parent plasmids and ligated into pMA424 (26) digested with *Eco*RI. A *Sma*I- and *Bam*HI-digested PCR product encoding E2-2 residues 1 to 99 was cloned into pAS1CYH2, a derivative of pAS1 (14), linearized with *Sma*I and *Bam*HI. All constructs were sequenced to confirm correct orientation and reading frame.

β-Galactosidase assay. Yeast strain yWAM2 (*MATα Δgal4 Δgal80 URA3*::*GAL1-lacZ lys2801^{amber} his3-Δ200 trp1-Δ63 leu2 ade2-101^{ochre} CYH2* [46]) was transformed with the various GAL4-E2A AD1 expression plasmids by the lithium acetate method (21). Individual transformants were grown to saturation in SC-His (3) and diluted 1:100 into synthetic medium lacking histidine and containing 2% ethanol, 3% glycerol, and 2% galactose. Cells were harvested at mid-log phase and assayed for β-galactosidase activity as described before (3).

Protein purification. To create a six-histidine tag (6×His) at the C terminus of AD1, the E2A PCR fragment encoding amino acids 1 to 99 was ligated into pET21a (Novagen) cut with *Eco*RI. The resulting bacterial expression vector was transformed into host strain BL21(DE3). Bacterial cultures were grown to an OD_{600} of 0.6 and induced with 1 mM IPTG (isopropylthiogalactopyranoside). Thirty minutes postinduction, rifampin (Sigma Chemical) was added to a final concentration of 25 μg/ml, and the cultures were incubated for an additional 2.5 h at 37°C. Harvesting, cell lysis, and purification by nickel chelate affinity chromatography were performed as described by the manufacturer (Novagen). The E2A AD1 6×His-tagged protein was >95% pure, as estimated by Coomassie blue staining of samples run on SDS-polyacrylamide gels. Protein concentration was determined by using an extinction coefficient (ε_{280}) of 5,690 M^{-1} cm^{-1} (16). Peptides were purchased from Chiron Mimotopes Peptide Systems and purified by reverse-phase high-pressure liquid chromatography (HPLC). The purity of the peptides was estimated by HPLC to be >95%. The concentrations of the peptides used for CD studies were confirmed by quantitative amino acid analysis.

CD analysis. The CD spectra of protein and peptide samples were taken in 10 mM phosphate buffer, pH 7, containing 137 mM NaF. The E2A AD1 6×His protein was unfolded in the same buffer including 6 M guanidinium chloride. 2,2,2-Trifluoroethanol (TFE) was added to some samples, as indicated in the legend to Fig. 7. CD spectra were recorded on an AVIV 61DS spectropolarimeter calibrated with a standard solution of 10-camphorsulfonic acid (22). Spectra were recorded with a 0.5-nm bandwidth, a 0.5-nm step size, and a 4.0-s time constant. Data were smoothed with a third-order polynomial function. CD spectra were measured at 22°C for E2A AD1 6×His or at 9°C for peptide samples. The mean residue ellipticity, [θ], is given in degrees times square centimeters per decimole.

RESULTS

The amino terminus of E2A contains a stretch of amino acids, including a putative helix, that is conserved among E proteins. Previous studies established that the N terminus of E2A contains two distinct transcriptional activation domains (2, 17, 38) (Fig. 1A). Using the BLAST sequence alignment algorithm (1), we identified a region of extensive homology between E2A and two other E proteins, HEB and E2-2 (Fig. 1B). Search results revealed that E2A is 45% identical to HEB and 44% identical to E2-2 over a 99-amino-acid stretch (Fig. 1B). This conserved stretch overlaps the previously identified E2A activation domain, AD1 (2). We next examined this region for potential secondary structure with both the Chou-Fasman and Robson-Garnier algorithms (7, 15). Chou-Fasman analysis predicts an α-helical structure spanning amino acids 11 to 28 of E2A (Fig. 1D). The Robson-Garnier method also predicts an α-helix within this same region (Fig. 1D). The putative helix is not amphipathic, as determined from helical wheel analysis (data not shown). This stretch of amino acids is extremely well conserved among the various E proteins, displaying an 83% identity to regions in both HEB and E2-2 (Fig. 1B). The evolutionary conservation of the putative α-helix is striking, with XE12, a *Xenopus laevis* homolog of E2A.E12, showing 100% identity, and Pan-2, the rat homolog of E2A.E12, showing 94% identity within this 18-amino-acid stretch (Fig. 1C). A recently identified zebra fish E12 homolog also displays a remarkable 78% identity within this region (not shown) (47).

The conserved region of E2A functions as a transcriptional activation domain. To determine whether the conserved region of E2A functions as a transcriptional activation domain, 99 amino acids surrounding the proposed helix were fused in frame to the DNA-binding domain of GAL4. The resulting expression plasmid was transfected into HeLa cells with a CAT reporter plasmid containing five GAL4 binding sites upstream of the E1b TATA sequence and subsequently assayed for CAT activity. Activation of the reporter gene to high levels was consistently observed, indicating that the region surrounding the putative helix can function as an independent transcriptional activation domain (Fig. 2A). Furthermore, activation was seen in all cell types tested, including LK, COS, Jurkat, NIH 3T3, and PD31 (data not shown).

To determine whether the proposed helix is sufficient for transactivation, we designed a series of E2A C-terminal deletions (Fig. 2). Deletion of 26 residues completely abolished transactivation potential in HeLa cells, reducing activity to approximately 3%, indicating that the conserved helix is not sufficient for transactivation (Fig. 2).

To determine whether the deleted products are actually synthesized in the cells, we transfected the deleted GAL4-E2A constructs, and protein from the transfected cells was isolated and analyzed by Western blotting with a GAL4 antibody. The GAL4-E2A deletion derivatives were expressed at levels comparable to that of the full-length fusion protein and were of the expected size (Fig. 3, lanes 1 and 9 to 12). Thus, we have identified a minimal domain in E2A that can function as a transcriptional activator. Since this 99-amino-acid region of E2A overlaps an activation domain identified by Aronheim et al. (2), we will refer to it as activation domain 1 (AD1). Note that the numbering designation for E2A is from that of the full-length E2A.E12 sequence (24) and is different from the numbering used by Quong et al. (38), which is in reference to E2-5 (17). E2-5 is a truncated, differentially spliced form of E47 which does not contain the AD1 activation domain.

Transactivation properties of E2-2(1–99) in mammalian cells and in yeast cells. Given that the potential helix of E2A AD1 is so well conserved, we wished to determine whether the amino terminus of E2-2 could also activate transcription. Both the E2A AD1 and E2-2(1–99) domains were able to activate the reporter construct when transfected into COS cells (Fig. 4A). Western blot analysis confirmed that the GAL4–E2-2(1–99) fusion protein was being made (Fig. 4B). Previously, we showed that a second transactivation domain, designated the loop-helix or AD2, in E2A was capable of activating transcription in yeast cells. To determine whether the E2A AD1 and

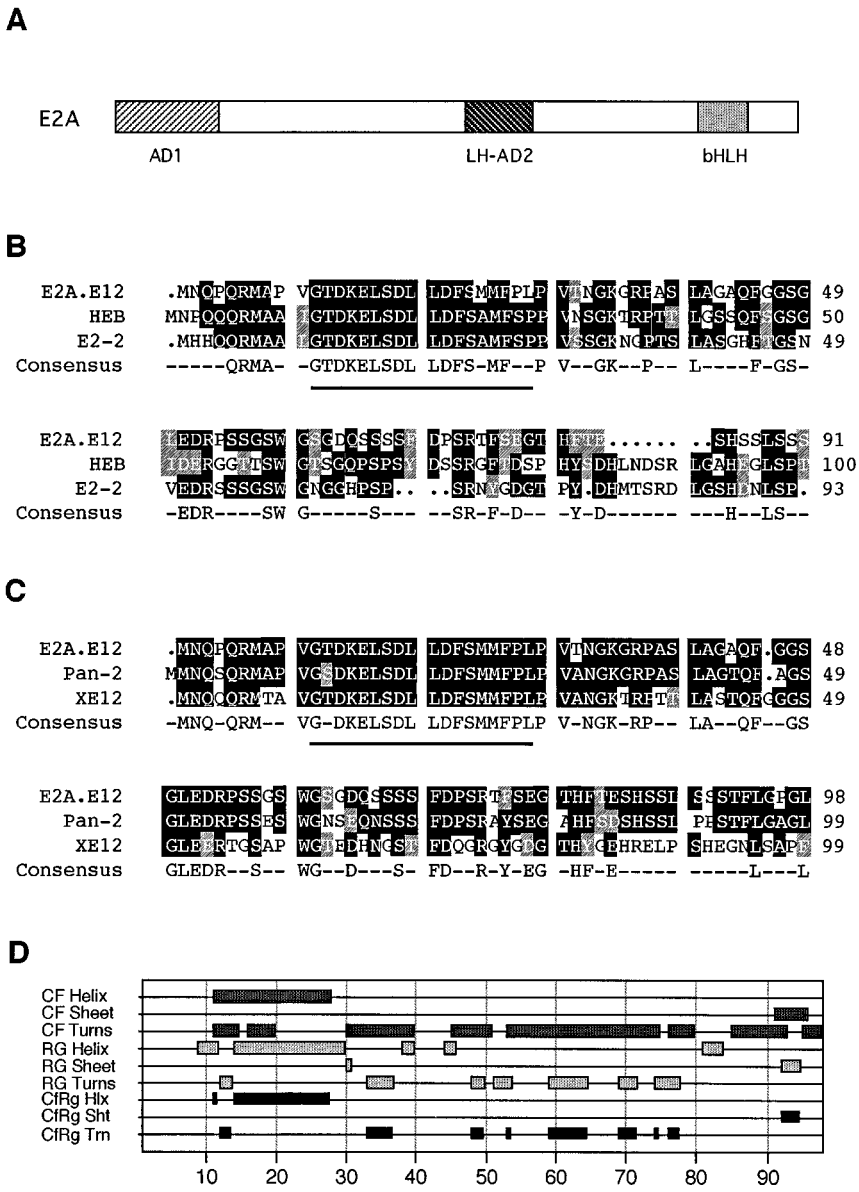

FIG. 1. The extreme N terminus of E2A is highly conserved. (A) Schematic depiction of the *E2A* gene product. The shaded boxes indicate the locations of previously characterized domains. bHLH, basic HLH domain (31); LH/AD2, loop-helix transactivation domain (amino acids 325 to 432) (this corresponds to amino acids 259 to 366 of the E2-5 clone [see text]) (2, 38); AD1, activation domain 1 (amino acids 1 to 99) (this study and reference 2). (B) Multiple sequence alignment of E12 (24), HEB (18), and E2-2 (10, 17) amino acid sequences, created with the PILEUP algorithm (Genetics Computer Group sequence analysis software package, version 7). The consensus sequence was displayed with the PRETTY algorithm. Amino acids identical to those in at least two sequences are shown as white letters in black boxes, and similar residues are shaded. (C) Multiple sequence alignment of E12, Pan-2 (34), and XE12 (39) proteins. The region containing the potential α-helix is underlined. (D) Chou-Fasman (CF) and Robson-Garnier (RG) analyses of the first 99 residues of E2A.E12, displayed with the MacVector sequence analysis software package. Regions of potential secondary structure are indicated by shaded boxes. CfRg indicates where CF and RG analyses agree. The putative helix, as predicted by the Chou-Fasman method, spans amino acids 11 to 28.

E2-2(1–99) domains have similar properties, both domains were fused to the GAL4 DNA-binding domain and transformed into a yeast strain harboring a *GAL1-HIS3* reporter gene. Both E2A and E2-2 could activate transcription, as indicated by growth on a medium lacking histidine and containing 25 mM 3-aminotriazole, a competitive inhibitor of the *HIS3* gene product (14) (Fig. 5). Thus, the conserved region of E2-2, like the AD1 domain of E2A, can function as a transactivator when overexpressed in both yeast and mammalian cells.

CD analysis of AD1. Structure prediction analysis showed that the AD1 domain has the potential to form a helical structure (Fig. 1D). To investigate this possibility, we overexpressed the E2A AD1 domain, containing a histidine tag, in *Escherichia coli* and purified the protein by nickel chelate affinity chromatography. To determine whether the AD1 domain indeed possesses helical properties, we analyzed its secondary structure in aqueous solution by CD spectroscopy. The spectrum displays a negative absorption at 220 nm, which indicates that the protein has helical content in solution at neutral pH (Fig. 6). This structure can be disrupted in the presence of 6 M guanidinium chloride, as indicated by the absence of the minimum at 220 nm (Fig. 6). The structure was fairly constant as

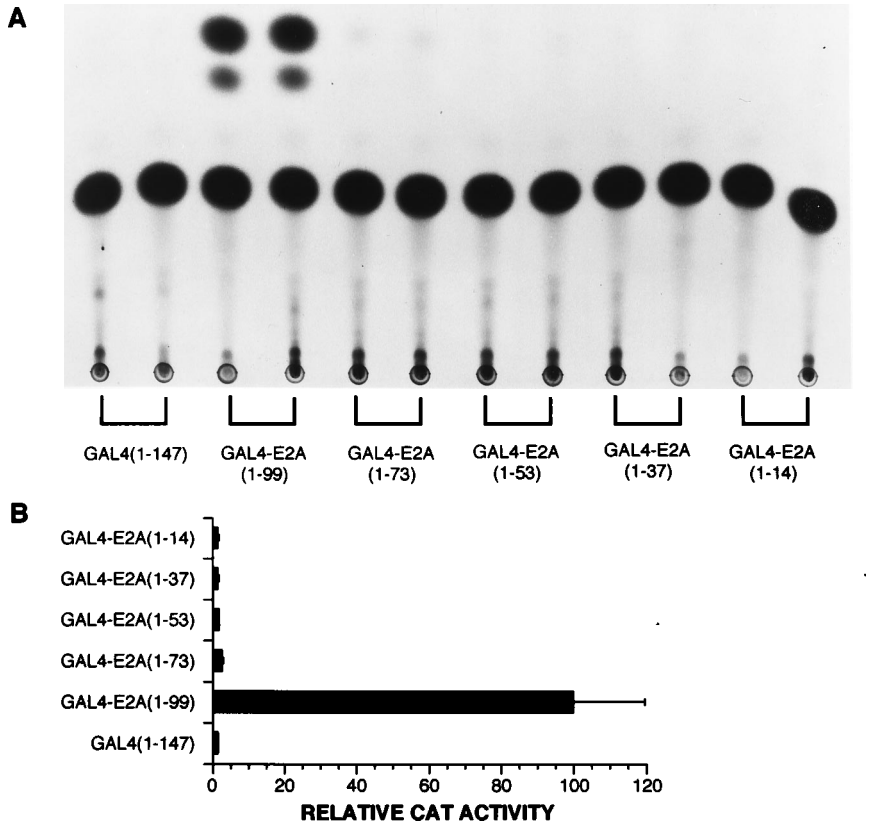

FIG. 2. The putative α-helix is not sufficient for transcriptional activation. (A) A series of C-terminal deletions were fused to the GAL4 DNA-binding domain, and relative transactivation levels of the resulting fusion proteins were assessed in HeLa cells. Expression plasmid (5 μg) was cotransfected along with 10 μg of a CAT reporter plasmid containing five GAL4 binding sites upstream of the adenovirus E1b TATA box. Shown are data from a representative CAT assay done in duplicate. Experiments were performed in duplicate at least two times, with independent assays yielding comparable results. (B) The data shown in panel A were quantitated by liquid scintillation counting, and the averages of duplicate values are expressed as percent relative CAT activity, with GAL4-E2A(1–99) assigned a value of 100%.

a function of temperature, as the CD spectrum did not significantly change whether recorded at 4°C or at 37°C (data not shown). These data suggest that AD1 is structured in aqueous solution. Quantitation of the helical content of AD1 was not warranted because the fusion protein contained an additional 35 residues predicted to be in random-coil conformation.

Since CD spectroscopy provides no information about the location of secondary structure within the protein, we determined the CD spectrum for a 23-amino-acid peptide which spans the predicted helix of E2A AD1. This peptide, designated WT(9–31), has the sequence N-PVGTDKELSDLLDF SMMFPLPVT-C. The WT(9–31) peptide displays little helical character at neutral pH, with only weak minima at 220 and 208 nm (Fig. 7A). In the presence of the nonpolar helix-promoting solvent TFE, however, the peptide adopts a helical structure in solution (Fig. 7A and B). A significant negative absorption occurs at 220 and 208 nm in the CD spectrum of the WT(9–31) peptide when going from 10 to 30% TFE, suggesting that the structure becomes more helical (Fig. 7A). In 50% TFE, the WT(9–31) peptide shows strong bands at 220 and 208 nm and a maximum at 192 nm, highly characteristic of an α-helical conformation (Fig. 7B) (23).

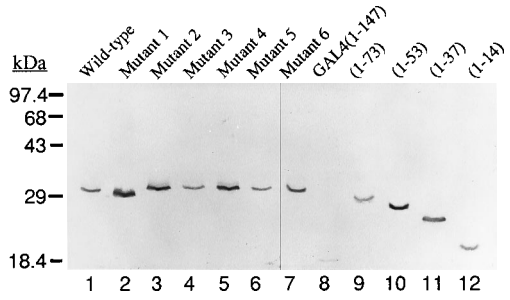

FIG. 3. Western blot analysis showing levels of expression of the various GAL4-E2A fusion proteins in COS cells. Proteins were detected with an anti-GAL4 antibody as described in Materials and Methods.

Next, we determined whether the helical properties of AD1 could be disrupted by the presence of two prolines. A peptide which contained two proline substitutions within the conserved helix was synthesized. The mutant peptide, MutP(9–31), which has the sequence N-PVGTDKELSDLLPFSPMFPLPVT-C, was assayed for its ability to adopt a helical conformation by CD spectroscopy. In the presence of 50% TFE, the MutP(9–31) peptide displayed helical content but significantly less than that of the WT(9–31) peptide under identical conditions (Fig. 7B). The minima at 220 and 208 nm were dramatically reduced in MutP(9–31), as was the maximum at 192 nm (Fig. 7B). Thus, proline substitutions markedly affect the ability of this peptide to adopt an α-helical conformation in solution.

Mutation of the E2A conserved α-helix dramatically reduces transactivation. To determine whether the conserved helix in AD1 is required for transactivation, we targeted this region of E2A for mutational analysis. A series of missense mutations, designed to disrupt the putative helix, were introduced into

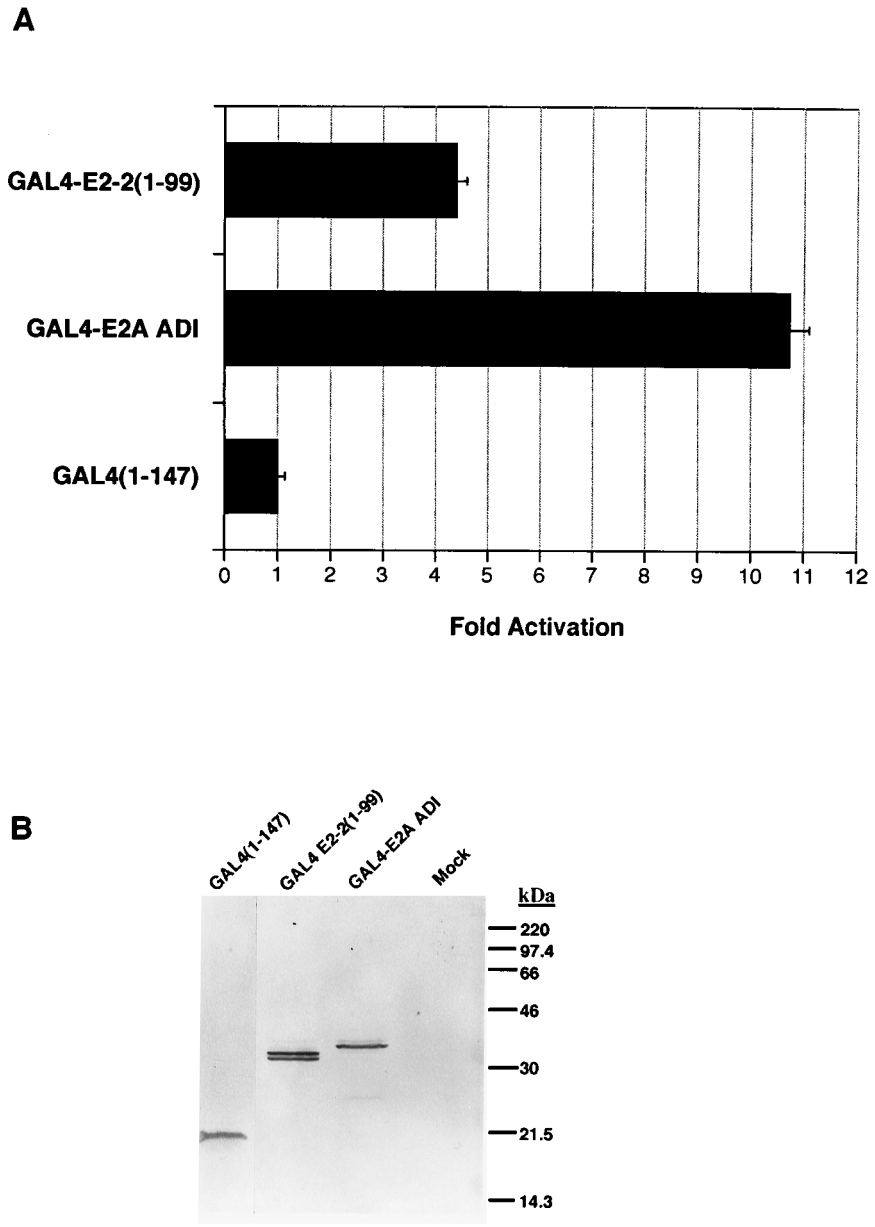

FIG. 4. The conserved region of E2-2 activates transcription in COS cells. (A) Expression plasmid (5 μg) and 5 μg of $(GAL4)_5$E1bCAT reporter were cotransfected into COS cells and assayed for CAT activity 48 h later. Shown is fold activation relative to that by GAL4(1–147). Transfections were performed at least three times. Data shown represent the average of triplicate values for a typical assay. (B) Western blot of GAL4 fusion proteins in COS cells.

AD1 (Fig. 8A; mutants 1, 2, and 4). The mutant proteins were transfected into HeLa cells and assayed for transcriptional activity. Two proline substitutions that clearly affect the ability of the synthetic peptide to adopt a helical conformation were introduced into the α-helix (mutant 4) by site-directed mutagenesis. The proline substitutions resulted in a dramatic decrease in transcriptional activation capability, reducing the activity to approximately 2% of wild-type levels (Fig. 8B and C). Similar effects were observed when two proline substitutions were directed to another part of the putative α-helix (mutant 1), resulting in a 50-fold decrease in transactivation levels (Fig. 8B and C). We also tested the effects of substitution of two conserved nonpolar residues present in the helix. Replacing a leucine with an arginine and a phenylalanine with an arginine (mutant 2) resulted in a remarkable 50-fold decrease in transcriptional activity (Fig. 8B and C). To confirm that the mutant proteins were being synthesized correctly, we performed Western blot analysis. All point mutants were of the expected size and expressed at comparable levels (Fig. 3, lanes 1 to 7). In summary, these data show that the helix present in the E2A AD1 domain is crucial for its ability to activate transcription.

In addition to the conserved α-helix, the AD1 domain contains several amino acids that are extremely well conserved (Fig. 1B and C). A number of amino acid substitutions were designed to test whether these residues play a role in mediating transactivation (Fig. 8A). A serine at position 48 was replaced with an alanine (mutant 5) in order to eliminate a potential casein kinase 2 phosphorylation site that is conserved in both HEB and E2-2. This mutation resulted in a modest but repro-

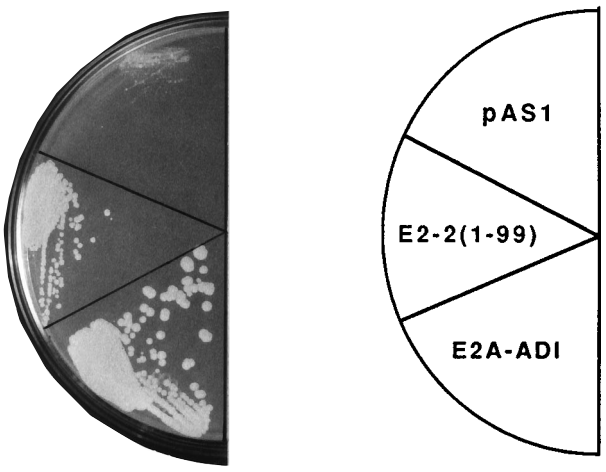

FIG. 5. E2A AD1 and E2-2(1–99) activate transcription in *S. cerevisiae*. Sequences encoding the conserved N-terminal domains of E2-2 or E2A were cloned into the expression vector pAS1CYH2 to create an in-frame fusion with the GAL4 DNA-binding domain. The indicated constructs were transformed into yeast strain y190, which contains an integrated *GAL1-HIS3* reporter gene (14). Transformants were selected on SC-Trp (3) plates and streaked out on SC-His plates containing 25 mM 3-aminotriazole (3AT). Growth on this selective medium indicates transcriptional activation of the *HIS3* reporter gene (14). pAS1 was included as a negative control. The plate was photographed after 6 days at 30°C.

ducible 25% decrease in activity. When the serines at positions 67 and 68 were substituted with alanine residues (mutant 6), transactivation levels were reduced to 44% (Fig. 8B and C). We also targeted highly conserved serine and tryptophan residues for mutagenesis (mutant 3) and found that activation was decreased by 51%. Thus, substitutions of conserved residues outside the putative helix of AD1 have only modest effects on transcriptional activity.

The AD1 domain is functionally conserved in mammalian cells and *Saccharomyces cerevisiae*. To test whether the E2A AD1 domain has the ability to activate transcription in the yeast *S. cerevisiae*, we transformed wild-type and mutant GAL4-AD1 expression constructs into a *GAL1-lacZ*-bearing reporter yeast strain (Fig. 9). The AD1 domain strongly activated the *lacZ* gene in yeast cells (Fig. 9). However, substitutions of residues in the putative helix of AD1 had a dramatic impact on the transactivation capability of AD1 in yeast cells, similar to that described above for HeLa cells. Specifically, substitution of proline residues into the α-helix (mutants 1 and 4) severely compromised transactivation in yeast cells, reducing activity by as much as 16-fold (Fig. 9). When conserved hydrophobic residues were replaced with arginines (mutant 2), transactivation dropped to background levels (Fig. 9). However, substitution of conserved residues located outside the proposed helix (mutants 3, 5, and 6) had only marginal impacts on transactivation capability (Fig. 9). Western blot analysis of yeast extracts made from strains harboring the various GAL4-AD1 expression vectors revealed that all proteins were expressed and were of the expected size (data not shown). These data suggest an interaction of AD1 with a common target that is conserved between mammals and *S. cerevisiae* (Fig. 8B and C and 9).

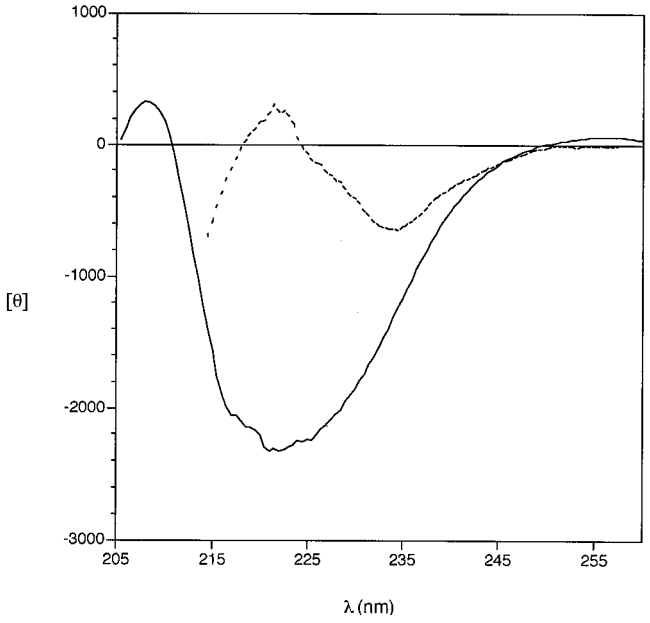

FIG. 6. CD spectroscopy reveals that E2A AD1 is structured. Shown are CD spectra of E2A AD1 6×His-tagged protein in the native (solid line) and denatured (dotted line) states. The protein concentration was 71 μM in a 0.2-cm cell.

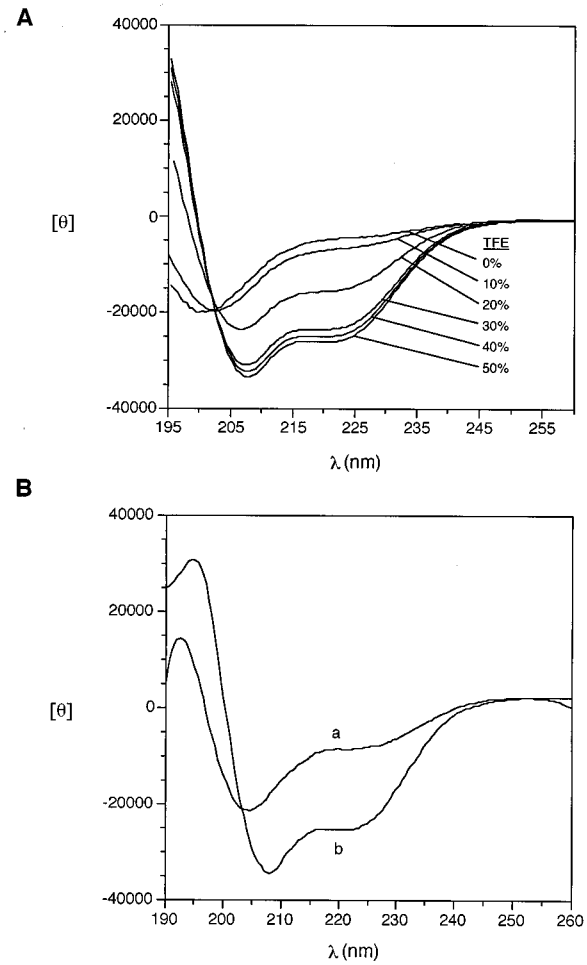

FIG. 7. A peptide containing the highly conserved residues (9 to 31) of E2A forms an α-helix in the presence of TFE. (A) CD spectra of the wild-type [WT(9–31)] peptide in the indicated concentrations of TFE. The peptide concentration was 30 μM in a 0.2-cm cell. (B) CD spectra of MutP(9–31) (a) and WT(9–31) (b) peptides in 50% TFE. Peptide concentrations were 30.7 μM for WT(9–31) and 31.2 μM for MutP(9–31). Spectra were taken in a 0.2-cm cell.

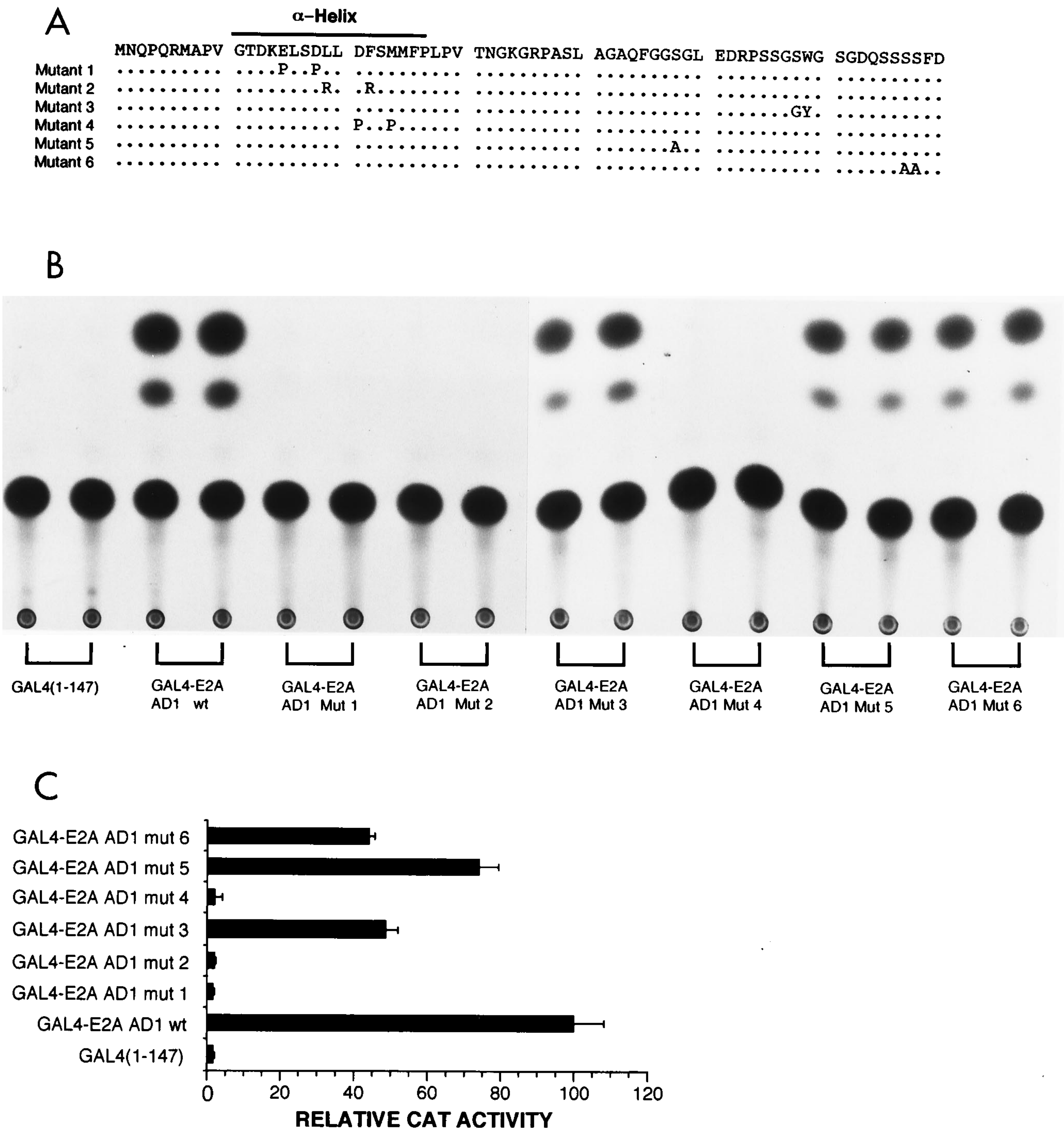

FIG. 8. The α-helix present in AD1 is crucial for transactivation. (A) Summary of the amino acid substitutions used in this study. The top row shows the first 70 residues of E2A.E12. Subsequent rows show the amino acid substitutions made in each AD1 point mutant. The black bar indicates the region of the proposed α-helix. (B) Five micrograms of expression plasmid was cotransfected along with 10 μg of $(GAL4)_5$E1bCAT reporter into HeLa cells. A representative CAT assay is shown. Transfections were performed in duplicate at least three times, and typical results are shown. (C) Quantitation of the data shown in panel B by liquid scintillation counting.

DISCUSSION

Activation of gene expression in eukaryotic cells requires proteins that have a modular structure. They contain a DNA-binding domain and a domain which allows them to activate transcription. We have characterized an activation domain, AD1, which has the unusual property of being conserved in a particular class of HLH proteins, the E proteins. AD1 is restricted to and conserved in a ubiquitously expressed class of HLH proteins. We attempted to find additional proteins that contain this conserved domain; however, no other proteins with significant homologies were detected in the SwissProt or GenBank database. Transcriptional activation domains have been widely characterized by their richness in certain amino acids, for example, proline- and glutamine-rich activation do-

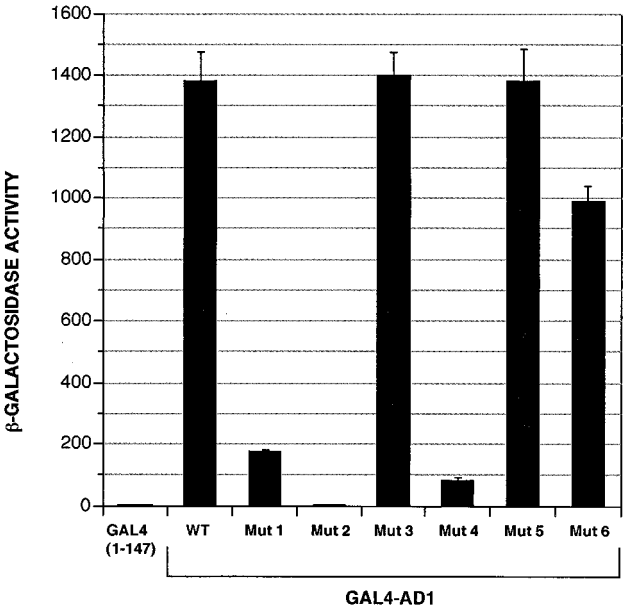

FIG. 9. The AD1 domain of E2A activates transcription in *S. cerevisiae*. The indicator yeast strain yWAM2 (46), which contains an integrated *GAL1-lacZ* reporter gene, was transformed with the indicated GAL4-AD1 expression constructs and assayed for β-galactosidase activity. The assay was performed in triplicate three times. Values shown represent the average of three independent transformants from a representative assay. β-Galactosidase activity is expressed in units.

mains (27). Few transactivation domains, however, have been characterized at the structural level. One particular class of transactivation domain, the acidic activation domain (AAD), has been characterized by CD spectroscopy. The AADs present in GAL4 and GCN4 form β-sheets under slightly acidic conditions (45). However, other acidic activation domains, such as the VP16 activation domain, have been shown to be largely unstructured at neutral pH in aqueous solution (13, 36). More recently, a transactivation domain present in the glucocorticoid receptor was shown to contain helical properties (12). Proline substitutions in the helix abolished transactivation, consistent with a role of the helix in transactivation (12).

A consensus sequence for the acidic class of activation domains, the AAD motif, has recently been identified (43). Interestingly, the AD1 helix contains the highly conserved dipeptide aspartic acid/glutamic acid-phenylalanine found in a number of AADs (Fig. 10). Although the significance of this motif is not yet understood, it suggests that the AD1 domain may be related to the AADs. Notably, the AD1 mutant 2 (Fig. 8A), which contains a phenylalanine-to-arginine substitution within this conserved dipeptide motif, has no activity in either yeast or mammalian cells. Although more extensive mutagenesis is required, these results suggests that the phenylalanine may play a critical role in mediating transactivation. AD1 is not highly acidic but has other features in common with the AADs. For example, AD1 is a potent transcriptional activator in yeast cells, a characteristic of many AADs. Like the human glucocorticoid receptor (hGR) τ1 (12) and p65 TA_1 (43) AADs, AD1 can form an α-helix in the presence of TFE. In all three cases, introduction of prolines into the helix severely impaired transactivation.

E Proteins	E2A (18-25)	DLLDFSMM
	HEB (19-26)	DLLDFSAM
	E2-2 (18-25)	DLLDFSAM
AADs	c-Fos (268-275)	PFDDFLFP
	LAP (83-90)	AHHDFLSD
	RelA (530-537)	GDEDFSSI
	RelA (538-545)	ADMDFSAL
	E2F (387-394)	VREDFSGL
	E2F (395-402)	LPEEFISL
	E2F (425-432)	FDCDFGDL
	VP16 (438-445)	ALDDFDLD
	VP16 (469-476)	DMADFEFE
	hGR (195-202)	QDLEFSSG
	E1A (140-147)	EGEEFVLD
	C/EBP (72-79)	FNDEFLAD

FIG. 10. The E proteins share a sequence motif with the acidic class of activators. The boxed region represents the previously identified dipeptide consensus aspartic acid/glutamic acid-phenylalanine present in a number of AADs (43). The portion of E2A AD1 containing the potential helix and the highly conserved regions of E2-2 and HEB are shown aligned with this consensus motif.

CD analysis strongly suggests that AD1 contains a helix which resides in the conserved N-terminal portion of E2A. The presence of a helix is further suggested by mutational analysis of AD1. Proline substitutions in the helix reduce the helical character of the peptide and, in addition, severely reduce transactivation both in mammalian cells and in *S. cerevisiae*. Further mutagenesis indicates essential roles for the conserved residues present in the helix, as substitutions in a number of residues reduce the transactivation abilities of AD1. In summary, we postulate that in the cell, the unstructured region present in AD1 becomes helical upon interaction with its target molecule, as previously suggested for the AAD present in p65 (43).

It is remarkable that the activities of the AD1 helix mutants in HeLa cells compare so closely with the results in yeast cells, strongly suggesting a common mechanism of transactivation. Perhaps the AD1 domain interacts with an ancient target molecule whose function has been conserved between *S. cerevisiae* and mammalian organisms. While performing the transactivation studies in *S. cerevisiae*, we noticed that overexpression of the GAL4-AD1 chimera conferred a slow-growth phenotype (26a). This toxicity was not restricted to the E2A AD1 domain but was also caused by GAL4–E2-2(1–99). Interestingly, the slow-growth phenotype correlated with transactivation capability; transformants harboring mutants that were defective for activation grew normally, suggesting a mechanism by which the AD1 is interacting with a limiting component(s) of the transcriptional machinery that has been conserved throughout evolution.

It is intriguing that the E2A N-terminal domain is fused to two heterologous DNA-binding domains in pro-B ALL and in pre-B ALL. This raises the question: what is the role of E2A in transformation? A first approach to this question has recently been demonstrated. AD1 has been shown to be required for the transforming activity of the E2A-Pbx1 and E2A-HLF chimeras (28, 49). Recently, a role for E2A in growth control has also been shown (37). A region within the first 174 amino acids of E2A, which overlaps the AD1 domain, can cause growth inhibition when expressed in NIH 3T3 cells (37). Understanding how the AD1 domain functions in regulating transcription and generating neoplasia are important questions, the answers to which await the identification of a molecular target.

ACKNOWLEDGMENTS

We thank Randall R. Reed for the yeast strain yWAM2, Rene Bernards for the GAL4 antibody, and Stephen J. Elledge for yeast strains and plasmids. We are grateful to Maarten H. Stuiver and to members of the Murre lab for many valuable discussions. We thank Merritt Maduke for quantitative amino acid analysis and Mike Hunter and Betsy Komives for assistance with peptide purification.

This work was supported by the National Institutes of Health, the Council for Tobacco Research, and the Edward Mallinkrodt, Jr., Foundation.

REFERENCES

1. **Altschul, S. F., W. Gish, W. Miller, E. W. Myers, and D. J. Lipman.** 1990. Basic local alignment search tool. J. Mol. Biol. **215:**403–410.
2. **Aronheim, A., R. Shiran, A. Rosen, and M. D. Walker.** 1993. The *E2A* gene product contains two separable and functionally distinct transcription activation domains. Proc. Natl. Acad. Sci. USA **90:**8063–8067.
3. **Ausubel, F., R. Brent, R. Kingston, D. Moore, J. Seidman, J. Smith, and K. Struhl.** 1991. Current protocols in molecular biology. John Wiley, New York.
4. **Bain, G., E. Maandag, D. Izon, D. Amsen, A. Kruisbeek, B. Weintraub, I. Krop, M. Schlissel, A. Feeney, M. van Roon, M. van der Valk, H. te Riele, A. Berns, and C. Murre.** 1994. E2A proteins are required for proper B cell development and initiation of immunoglobulin gene rearrangements. Cell **79:**885–892.
5. **Caudy, M., E. H. Grell, C. Dambly-Chaudiere, A. Ghysen, L. Y. Jan, and Y. N. Jan.** 1988. The maternal sex determination gene *daughterless* has zygotic activity necessary for the formation of peripheral neurons in *Drosophila*. Genes Dev. **2:**843–852.
6. **Caudy, M., H. Vaessin, M. Brand, R. Tuma, L. Y. Jan, and Y. N. Jan.** 1988. *daughterless*, a gene essential for both neurogenesis and sex determination in Drosophila, has sequence similarities to *myc* and the *achaete-scute* complex. Cell **55:**1061–1067.
7. **Chou, P. Y., and G. D. Fasman.** 1974. Prediction of protein conformation. Biochemistry **13:**222–245.
8. **Cline, T. W.** 1989. The affairs of *daughterless* and the promiscuity of developmental regulators. Cell **59:**231–234.
9. **Cordle, S., E. Henderson, H. Masuoka, P. Weil, and R. Stein.** 1991. Pancreatic beta-cell-type-specific transcription of the insulin gene is mediated by basic helix-loop-helix DNA-binding proteins. Mol. Cell. Biol. **11:**1734–1738.
10. **Corneliussen, B., A. Thornell, B. Hallberg, and T. Grundstrom.** 1991. Helix-loop-helix transcriptional activators bind to a sequence in glucocorticoid response elements of retrovirus enhancers. J. Virol. **65:**6084–6093.
11. **Cronmiller, C., P. Schedl, and T. W. Cline.** 1988. Molecular characterization of *daughterless*, a Drosophila sex determination gene with multiple roles in development. Genes Dev. **2:**1666–1676.
12. **Dahlman-Wright, K., H. Baumann, I. McEwan, T. Almlof, A. Wright, J.-A. Gustafsson, and T. Hard.** 1995. Structural characterization of a minimal functional transactivation domain from the human glucocorticoid receptor. Proc. Natl. Acad. Sci. USA **92:**1699–1703.
13. **Donaldson, L., and J. Capone.** 1992. Purification and characterization of the carboxyl-terminal transactivation domain of Vmw65 from herpes simplex virus type 1. J. Biol. Chem. **267:**1411–1414.
14. **Durfee, T., K. Becherer, P.-L. Chen, S.-H. Yeh, Y. Yang, A. Kilburn, W.-H. Lee, and S. Elledge.** 1993. The retinoblastoma protein associates with the protein phosphatase type 1 catalytic subunit. Genes Dev. **7:**555–569.
15. **Garnier, J., O. Osguthorpe, and B. Robson.** 1978. Analysis of the accuracy and implications of simple methods for predicting the secondary structure of globular proteins. J. Mol. Biol. **120:**97–120.
16. **Gill, S., and P. H. von Hippel.** 1989. Calculation of protein extinction coefficients from amino acid sequence data. Anal. Biochem. **182:**319–326.
17. **Henthorn, P., M. Kiledjian, and T. Kadesch.** 1990. Two distinct transcription factors that bind the immunoglobulin enhancer μE5/κE2 motif. Science **247:** 467–470.
18. **Hu, J.-S., E. N. Olson, and R. E. Kingston.** 1992. HEB, a helix-loop-helix protein related to E2A and ITF2 that can modulate the DNA-binding ability of myogenic regulatory factors. Mol. Cell. Biol. **12:**1031–1042.
19. **Hunger, S., K. Ohyashiki, K. Toyama, and M. Cleary.** 1992. HLF, a novel hepatic bZIP protein, shows altered DNA-binding properties following fusion to E2A in t(17;19) acute lymphoblastic leukemia. Genes Dev. **6:**1608–1620.
20. **Inaba, T., W. M. Roberts, L. H. Shapiro, K. W. Jolly, S. C. Raimondi, S. D. Smith, and A. T. Look.** 1992. Fusion of the leucine zipper gene HLF to the E2A gene in human acute B-lineage leukemia. Science **257:**531–534.
21. **Ito, H., Y. Fukuda, K. Murata, and A. Kimura.** 1983. Transformation of intact yeast cells treated with alkali cations. J. Bacteriol. **153:**163–168.
22. **Johnson, W. C., Jr.** 1985. Circular dichroism and its empirical application to biopolymers. Methods Biochem. Anal. **31:**61–163.
23. **Johnson, W. C., Jr.** 1990. Protein secondary structure and circular dichroism: a practical guide. Proteins Struct. Funct. Genet. **7:**205–214.
24. **Kamps, M. P., C. Murre, X.-H. Sun, and D. Baltimore.** 1990. A new homeobox gene contributes the DNA binding domain of the t(1;19) translocation protein in pre-B ALL. Cell **60:**547–555.
25. **Lassar, A. B., R. L. Davis, W. E. Wright, T. Kadesch, C. Murre, A. Voronova, D. Baltimore, and H. Weintraub.** 1991. Functional activity of myogenic HLH proteins requires hetero-oligomerization with E12/E47-like proteins in vivo. Cell **66:**305–315.
26. **Ma, J., and M. Ptashne.** 1987. A new class of yeast transcriptional activators. Cell **51:**113–119.

26a. **Massari, M. E., and C. Murre.** Unpublished data.

27. **Mitchell, P., and R. Tjian.** 1989. Transcriptional regulation in mammalian cells by sequence-specific DNA-binding proteins. Science **245:**371–378.
28. **Monica, K., D. LeBrun, D. Dedera, R. Brown, and M. Cleary.** 1994. Transformation properties of the E2a-Pbx1 chimeric oncoprotein: fusion with E2A is essential, but the Pbx1 homeodomain is dispensable. Mol. Cell. Biol. **14:** 8304–8314.
29. **Murre, C., G. Bain, M. A. van Dijk, I. Engel, B. A. Furnari, M. E. Massari, J. R. Matthews, M. W. Quong, R. R. Rivera, and M. H. Stuiver.** 1994. Structure and function of helix-loop-helix proteins. Biochim. Biophys. Acta **1218:**129–135.
30. **Murre, C., and D. Baltimore.** 1992. The helix-loop-helix motif: structure and function, p. 861–879. *In* S. L. McKnight and K. R. Yamamoto (ed.), Transcriptional regulation. Cold Spring Harbor Laboratory Press, Plainview, N.Y.
31. **Murre, C., P. S. McCaw, and D. Baltimore.** 1989. A new DNA binding and dimerization motif in immunoglobulin enhancer binding, *daughterless*, *MyoD*, and *myc* proteins. Cell **56:**777–783.
32. **Murre, C., P. S. McCaw, H. Vaessin, M. Caudy, L. Y. Jan, Y. N. Jan, C. V. Cabrera, J. N. Buskin, S. D. Hauschka, A. B. Lassar, H. Weintraub, and D. Baltimore.** 1989. Interactions between heterologous helix-loop-helix proteins generate complexes that bind specifically to a common DNA sequence. Cell **58:**537–544.
33. **Murre, C., A. Voronova, and D. Baltimore.** 1991. B-cell and myocyte-specific E2-box-binding factors contain E12/E47-like subunits. Mol. Cell. Biol. **11:** 1156–1160.
34. **Nelson, C., L. P. Shen, A. Meister, E. Fodor, and W. J. Rutter.** 1990. Pan: a transcriptional regulator that binds chymotrypsin, insulin and AP-4 enhancer motifs. Genes Dev. **4:**1035–1044.
35. **Nourse, J., J. D. Mellentin, N. Galili, J. Wilkinson, E. Stanbridge, S. D. Smith, and M. L. Cleary.** 1990. Chromosomal translocation t(1;19) results in synthesis of a homeobox fusion mRNA that codes for a potential chimeric transcription factor. Cell **60:**535–545.
36. **O'Hare, P., and G. Williams.** 1992. Structural studies of the acidic transactivation domain of the Vmw65 protein of herpes simplex virus using 1H NMR. Biochemistry **31:**4150–4156.
37. **Peverali, F., T. Ramqvist, R. Saffrich, R. Pepperkok, M. Barone, and L. Philipson.** 1994. Regulation of G1 progression by E2A and Id helix-loop-helix proteins. EMBO J. **13:**4291–4301.
38. **Quong, M. W., M. E. Massari, R. Zwart, and C. Murre.** 1993. A new transcriptional activation motif restricted to a class of helix-loop-helix proteins is functionally conserved in both yeast and mammalian cells. Mol. Cell. Biol. **13:**792–800.
39. **Rashbass, J., M. V. Taylor, and J. B. Gurdon.** 1992. The DNA-binding protein E12 co-operates with XMyoD in the activation of muscle-specific gene expression in *Xenopus* embryos. EMBO J. **11:**2981–2990.
40. **Sambrook, J., E. F. Fritsch, and T. Maniatis.** 1989. Molecular cloning: a laboratory manual, 2nd ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.
41. **Sawada, S., and D. R. Littman.** 1993. A heterodimer of HEB and an E12-related protein interacts with the CD4 enhancer and regulates its activity in T-cell lines. Mol. Cell. Biol. **13:**5620–5628.
42. **Schlissel, M., A. Voronova, and D. Baltimore.** 1991. Helix-loop-helix transcription factor-E47 activates germ-line immunoglobulin heavy-chain gene transcription and rearrangement in a pre-T-cell line. Genes Dev. **5:**1367–1376.
43. **Schmitz, M., M. dos Santos Silva, H. Altmann, M. Czisch, T. Holak, and P. Baeuerle.** 1994. Structural and functional analysis of the NF-κB p65 C terminus. J. Biol. Chem. **269:**25613–25620.
44. **Sun, X.-H., and D. Baltimore.** 1991. An inhibitory domain of E12 transcription factor prevents DNA binding in E12 homodimers but not in E12 heterodimers. Cell **64:**459–470.
45. **Van Hoy, M., K. Leuther, T. Kodadek, and S. Johnston.** 1993. The acidic activation domains of the GCN4 and GAL4 proteins are not α helical but form β sheets. Cell **72:**587–594.
46. **Wang, M. M., and R. R. Reed.** 1993. Molecular cloning of the olfactory neuronal transcription factor Olf-1 by genetic selection in yeast. Nature (London) **364:**121–126.
47. **Wulbeck, C., C. Fromental-Ramain, and J. Campos-Ortega.** 1994. The HLH domain of a zebrafish HE12 homologue can partially substitute for functions of the HLH domain of Drosophila DAUGHTERLESS. Mech. Dev. **46:**73–85.
48. **Yaffe, M., and G. Schatz.** 1984. Two nuclear mutations that block mitochondrial protein import in yeast. Proc. Natl. Acad. Sci. USA **81:**4819–4823.
49. **Yoshihara, T., T. Inaba, L. H. Shapiro, J.-Y. Kato, and A. T. Look.** 1995. E2A-HLF-mediated cell transformation requires both the *trans*-activation domains of E2A and the leucine zipper dimerization domain of HLF. Mol. Cell. Biol. **15:**3247–3255.
50. **Zhuang, Y., P. Soriano, and H. Weintraub.** 1994. The helix-loop-helix gene E2A is required for B cell formation. Cell **79:**875–884.