Proc. Natl. Acad. Sci. USA
Vol. 87, pp. 4722–4726, June 1990
Biochemistry

Mutations that disrupt DNA binding and dimer formation in the E47 helix-loop-helix protein map to distinct domains

(transcriptional enhancers/E2A gene/helix-loop-helix structure)

ANNA VORONOVA AND DAVID BALTIMORE

Whitehead Institute for Biomedical Research, Nine Cambridge Center, Cambridge, MA 02142; and Department of Biology, Massachusetts Institute of Technology, Cambridge, MA 02139

Contributed by David Baltimore, April 2, 1990

ABSTRACT **A common DNA binding and dimerization domain containing an apparent "helix-loop-helix" (HLH) structure was recognized recently in a number of regulatory proteins, including the E47 and E12 proteins that bind to the κE2 motif in immunoglobulin κ gene enhancer. The effect of site-directed mutagenesis on E47 protein multimerization and DNA binding was examined. Mutations in either putative helix domain disrupted protein dimerization and DNA binding. No DNA binding was observed when mutations were introduced in the basic region, but these mutants were able to dimerize. These basic region mutants were not able to bind to DNA as heterodimers with the wild-type E47 proteins, demonstrating that two functional basic regions are required for binding to DNA. Therefore the basic region mutants are "trans-dominant."**

The protein products of the E2A gene, E47 and E12, form homo- and heterodimers that recognize the κE2 site of the immunoglobulin κ gene enhancer (1, 2). The κE2 site plays a critical role in the function of the enhancer (3, 4). A DNA site similar to the κE2 sequence is also found in the insulin gene enhancer (5) and in the muscle creatine kinase gene enhancer (6), and both are important for the tissue-specific expression of their respective genes. The protein products of the three genes that induce myogenesis, MyoD (7, 8), myogenin (9), and *myf-5* (10) all share with E47 a highly conserved 60-amino acid stretch with predicted helix-loop-helix (HLH) structure (2) and act presumably by way of interaction with the κE2-like sequences in the muscle-specific genes. Other members of the "HLH family" have been defined based on amino acid sequence homology to the region of the E47 polypeptide required for specific recognition of DNA and include the *myc* family proteins; proteins important for *Drosophila* development, including the achaete-scute complex (11, 12), daughterless (13), twist (14), and Enhancer of split (15); and the *lyl-1* gene implicated in T-cell acute lymphoblastic leukemia (16). These proteins all contain a basic amino acid region followed by two short putative amphipathic helices separated by an intervening region that might form a loop, the HLH motif. In addition to homodimers, the HLH proteins can also form heterodimers that bind specifically to a common DNA sequence (2). Deletion mutagenesis analysis of the E47/E12 proteins showed that the basic region and HLH domain, which are located at the carboxyl terminus, are sufficient for the DNA binding activity *in vitro* (1, 2). To define further the regions required for specific contact with DNA and for dimer formation, as well as to determine whether the two can be functionally separated, site-directed mutagenesis of the E47 gene was performed. Experimental evidence is presented here that the amphipathic helices mediate complex formation between E47 proteins that results in the appropriate juxtaposition of the DNA binding region of the protein. Furthermore, dimerization is required for DNA binding, which also depends on a specific sequence of basic amino acids, positioned immediately amino-terminal to the HLH motif. In the absence of the DNA binding, however, dimerization still occurs.

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

MATERIALS AND METHODS

Plasmids. Construction of the pE47P and pE47S plasmids encoding two truncated versions of the E47 protein, E47P and E47S, has been described (1).

Site-Directed Mutagenesis of E47. Mutations were constructed by gapped heteroduplex oligonucleotide replacement using pE47S plasmid DNA. The structure of each mutant was confirmed by dideoxynucleotide sequencing of the double-stranded DNA using Sequenase (United States Biochemical).

***In Vitro* Transcription and Translation.** Plasmid DNA was linearized with *Eco*RI and used as a template for *in vitro* transcription with T3 RNA polymerase (Stratagene). Wild-type and mutant proteins were synthesized using rabbit reticulocyte lysates (Promega) in the presence of [^{35}S]methionine (Amersham).

DNA Binding Assay. The *Eco*RI–*Hin*dIII fragment containing the wild-type κE2 sequence was labeled with ^{32}P (50,000 cpm/ng), and 0.5 ng of DNA probe was incubated with 5–10 μl of the reacted or unreacted reticulocyte lysate (detailed in the figure legends). The binding reactions were incubated at room temperature for 15 min and complexes were resolved by electrophoresis as described (1).

Immunoprecipitation. Proteins synthesized *in vitro* in the presence of [^{35}S]methionine (Amersham) were dissolved in ELB buffer (250 mM NaCl/0.1% Nonidet P-40/50 mM Hepes, pH 7.0/5 mM EDTA/1.0 mM phenylmethylsulfonyl fluoride/0.5 mM dithiothreitol) and incubated on ice for 1 hr in the presence of 1 μl of a rabbit polyclonal antibody raised against the carboxyl-terminal 440 amino acids of E12 (2). The immune complexes were isolated by incubation with 25 μl of protein A Sepharose (Sigma), washed four times at 4°C with ELB buffer, and resolved by electrophoresis on $NaDodSO_4$/20% polyacrylamide gels.

RESULTS

Mutagenesis of E47S. The E47 and E12 polypeptides are the protein products resulting from the alternative splicing of the E2A gene and differ only in their carboxyl-terminal exon, which encodes basic and HLH regions (X. H. Sun and D.B., unpublished). Deletion mutagenesis defined a region of the E47 polypeptide containing amino acids 313–495 as a domain sufficient for the specific recognition of the κE2 DNA motif

Abbreviations: HLH, helix-loop-helix; EMSA, electrophoretic mobility shift assay.

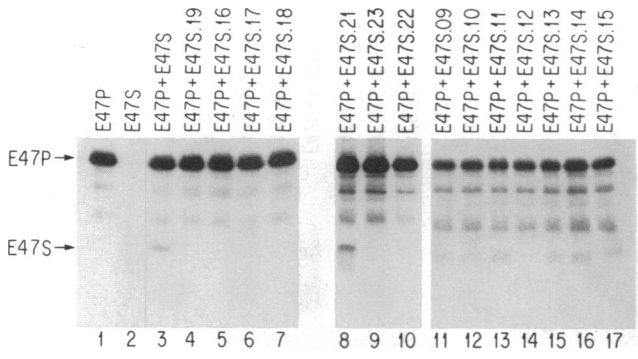

FIG. 1. Immunoprecipitation analysis of the protein complexes with E12-specific antibody. The *in vitro*-translated products were incubated in the absence of DNA as described in the text, and the immunoprecipitates formed with an E12 antibody were analyzed on a NaDodSO$_4$/20% polyacrylamide gel. Mutant sequences are shown in Fig. 2. Positions of E47P and E47S are marked by arrows. The background bands are the products of the E47P *in vitro* translation and are caused either by the premature terminations or protein degradation.

(1). This truncated version of the protein, E47S, can form homodimers as well as heterodimers with other members of the HLH family and can bind specifically to its cognate DNA motif (1, 2). Because the truncated E47S protein does not contain the amino-terminal region shared between E12 and E47 proteins (1), we predicted and confirmed that the E12-specific serum would recognize E47P protein but not its truncated form, E47S (Fig. 1, lanes 1 and 2). However, when the E47S and E47P proteins were combined in the absence of DNA and then subjected to immunoprecipitation, the truncated E47S protein was readily detected in a complex with the E47P protein (Fig. 1, lane 3). This result demonstrates that E47 polypeptides are able to multimerize in the absence of DNA. To define precisely the regions involved in protein complex formation and DNA binding, we introduced site-specific mutations into the E47S polypeptide in the regions of sequence conservation among the members of the HLH family (Fig. 2). Wild-type and mutant proteins were synthesized *in vitro*, and coimmunoprecipitation was used to determine the effect of site-directed mutagenesis on protein complex formation. The ability of the E47S mutant proteins to bind DNA was analyzed by an electrophoretic mobility shift assay (EMSA) with ^{32}P-labeled DNA probe containing the κE2 sequence GGCAGGTGG.

Amphipathic Helices Mediate Dimerization. The positions of the conserved hydrophobic amino acid residues in the amphipathic helices form a four-three repeat, which would bring them to one side of a helix. By analogy with the intermediate filaments, it was proposed that these amino acids allow hydrophobic interaction between the helices during protein dimerization via a coiled-coil structure (1). To test this hypothesis, single and double mutations were introduced at these conserved residues. Mutant protein was synthesized *in vitro*, mixed with the wild-type E47P protein, preincubated in the absence of DNA, and subjected to immunoprecipitation with the E12-specific serum. The one singly mutant protein, E47S.11, was coimmunoprecipitated with E47P (Fig. 1, lane 13). However, all of the double mutants, E47S.12, E47S.16, E47S.17, and E47S.19, were not detected in the immunoprecipitates (Fig. 1), implying that the mutations disrupted dimerization. We next tested the ability of the mutants to bind DNA. When a mixture of wild-type E47S and E47P proteins was used for the DNA binding assay,

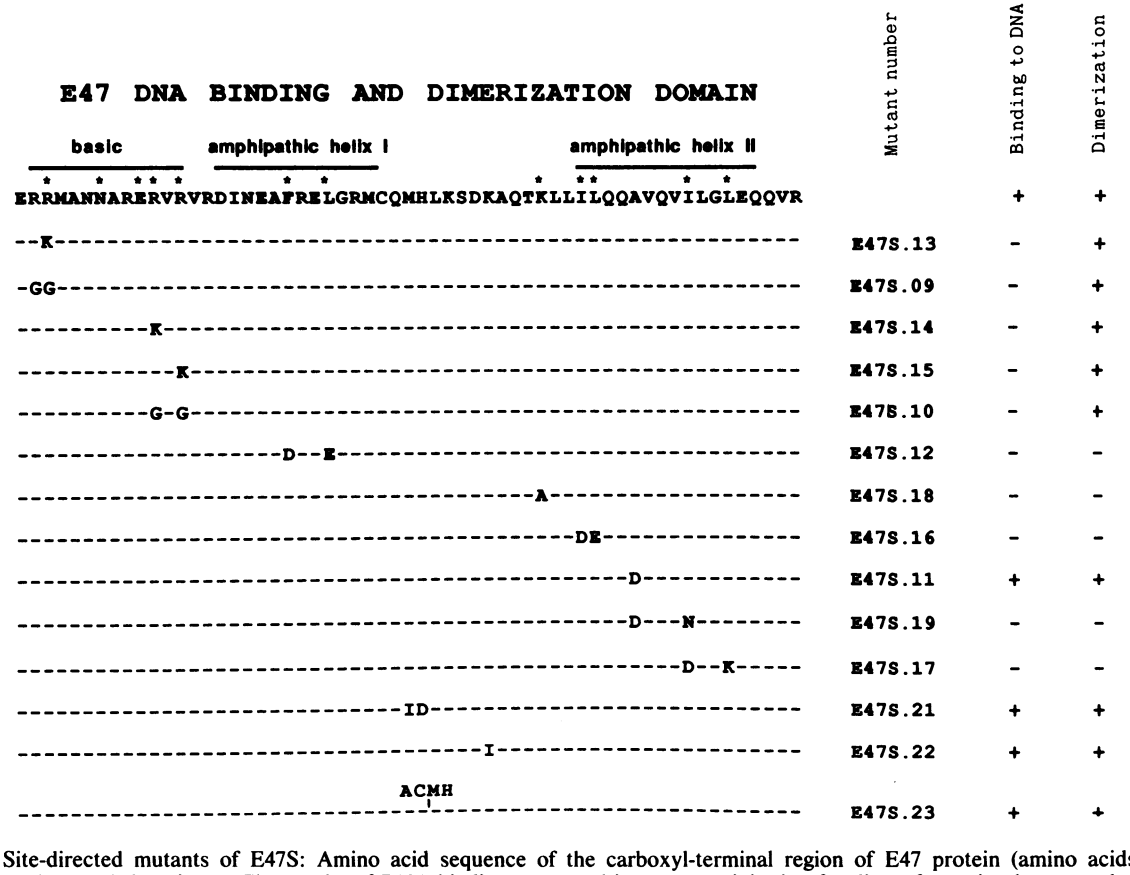

FIG. 2. Site-directed mutants of E47S: Amino acid sequence of the carboxyl-terminal region of E47 protein (amino acids 336–393) encompassing the HLH domain (1). The results of DNA binding assay and immunoprecipitation for dimer formation between the E47P and mutants are shown on the left. Solid lines mark the amphipathic helices and basic domain. Asterisks indicate amino acids identity among the members of the HLH family.

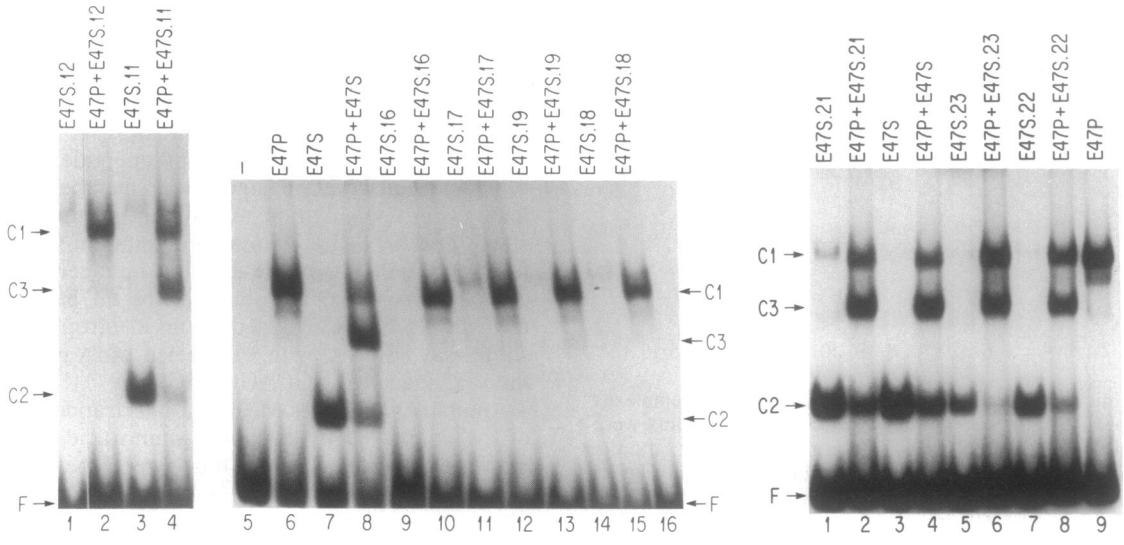

FIG. 3. DNA binding activity of HLH region mutants. The *in vitro*-translated proteins were analyzed by EMSA using a κE2 monomer as a probe. Mutant sequences are shown in Fig. 2. (*Left*) DNA binding by mutants of the helices. (*Right*) DNA binding by loop mutants. C1, C2, and C3 indicate E47P-, E47S-, and E47P/E47S heterodimer-specific complexes. F designates the position of the free DNA probe. Minus sign (−) corresponds to "no RNA" *in vitro* translation reaction mixture.

three complexes, C1, C2, and C3, corresponding to the E47P and E47S homodimers and the E47S/E47P heterodimer, respectively, were formed (Fig. 3 *Left*, lane 8). A double mutation in the amphipathic helix I that replaced phenylalanine and leucine by aspartic and glutamic acids resulted in the absence of a complex with the κE2-specific DNA probe (Fig. 3 *Left*, lane 1). No heterodimer complex was detected when this mutant E47S protein was combined with the wild-type E47P protein (Fig. 3 *Left*, lane 2). A series of double mutations in the amphipathic helix II, E47S.16, E47S.17, and E47S.19, also abolished DNA binding as homo- and heterodimer, whereas the single mutation, E47S.11, resulted in a weaker binding to the specific DNA probe (Fig. 3 *Left*, lanes 3 and 4). Substitution of alanine for the highly conserved lysine positioned amino-terminal in the amphipathic helix II completely disrupted dimerization (Fig. 1, lane 7) and abolished DNA binding (Fig. 1, lanes 15 and 16). Because this charged residue is strongly conserved among the HLH proteins, it is apparently providing stability to the dimer. Therefore we concluded that dimerization is required for DNA binding, since the mutations in the helix I and helix II that disrupted dimerization of the proteins resulted in the inability to bind to DNA.

Loop Region Separates the Two Helices. We next introduced mutations within the loop region to assess its contribution to dimerization and subsequent DNA binding. A few random mutations within the region produced no significant changes in protein dimerization and binding to DNA. A double mutation that replaced methionine and histidine by isoleucine and aspartic acid resulted in essentially wild-type dimerization (Fig. 1, lane 8) and binding to DNA (Fig. 3 *Right*, lanes 1 and 2). Similarly, mutation of a lysine to isoleucine produced no effect on dimerization and the subsequent DNA binding (Figs. 1 and 3 *Right*). Insertion of the four amino acid residues in the middle of the loop (mutant E47S.23) resulted in weaker dimerization that is indicated by a less stable complex with the wild-type E47P during immunoprecipitation (Fig. 1, lane 9). The destabilization of the protein dimer in turn reduced the ability of the mutant to bind DNA as a heterodimer (Fig. 3 *Right*, lane 5). However, the mutant protein formed a strong complex with DNA as a heterodimer (Fig. 3 *Right*, lane 6). It appears, therefore, that the loop region sequence does not play a critical role in E47 protein dimerization and binding to DNA, but it may have a function of providing a structural element separating the two helices.

Effect of the Basic Region on Binding to DNA. A charged region positioned immediately amino-terminal to the HLH domain is structurally conserved among the members of the HLH family (1). A few clusters of basic amino acids—in particular, arginines—are present at the fixed positions within this region, perhaps implicating this domain in DNA binding. We therefore analyzed the effect of the alteration of these residues on E47 protein dimerization and binding to DNA. Double mutations of the highly conserved arginine residues to glycines, E47S.09 and E47S.10, completely abolished DNA binding (Fig. 4, lanes 4 and 6). Surprisingly, DNA binding was absent in all of the mutants with a single substitution of arginine to lysine at each of the three highly conserved positions (Fig. 4, lanes 8, 10, and 12), suggesting an important role of these residues in binding to DNA. All of

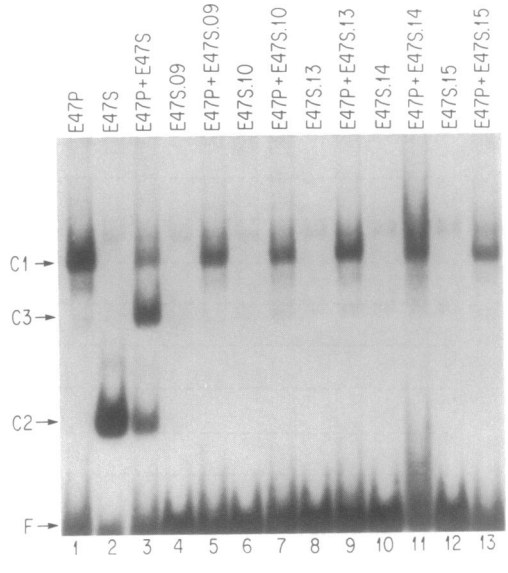

FIG. 4. DNA binding activity of the basic region mutants. The *in vitro*-translated products were analyzed by EMSA using a κE2 monomer as a probe. Mutant sequences are shown in Fig. 2. C1, C2, and C3 indicate E47P-, E47S-, and E47P/E47S heterodimer-specific complexes. F designates the position of the free DNA probe.

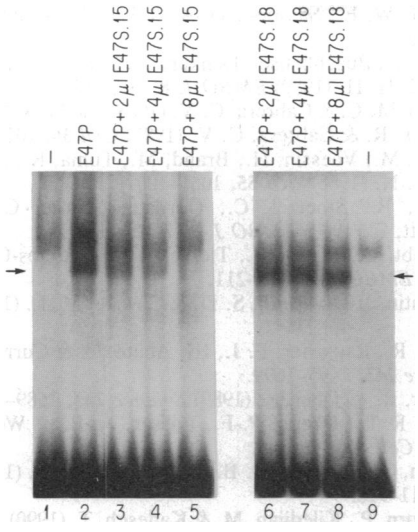

FIG. 5. DNA binding activity of E47P in the presence of E47S mutants. Two microliters of *in vitro* synthesized E47P was mixed with the indicated amount of mutant protein; "no RNA" *in vitro* translation reaction mixture was added to the final volume of 10 μl. Protein binding activity was analyzed by EMSA using a κE2 monomer as a probe. Arrow indicates the position of E47P-specific complex. Minus sign (–) corresponds to "no RNA" *in vitro* translation reaction mixture alone.

these mutations also prevented DNA binding as heterodimers with the wild-type E47P protein (Fig. 4, lanes 9, 11, and 13). However, when the basic region mutant proteins were combined with the wild-type E47P protein and subjected to immunoprecipitation with E12-specific serum, they were readily detected in the precipitates (Fig. 1, lanes 11, 12, and 15–17). Mutations of the basic region therefore did not interfere with the ability of the protein to dimerize, but they did abolish binding to DNA. We propose therefore that the basic region of the protein—in particular, the highly conserved arginine residues present at the fixed positions within the region—are necessary for specific contact with DNA. The DNA binding surface of the protein is apparently formed by dimerization of the polypeptides.

Trans-Dominant Suppression of Binding to DNA. The E47S basic region mutant proteins were able to form dimers with the E47P wild-type protein but completely lacked the ability to bind DNA. These mutations also abolished binding to DNA as heterodimers with the wild-type E47P protein. We therefore assayed the DNA binding activity associated with E47P when increasing amounts of the basic region mutant proteins (mutant E47S.15 is shown here as an example) were added to the reaction mixture (Fig. 5, lanes 3–5). Disappearance of the E47P-specific homodimer complex indicates that these mutants interfere with DNA binding in a dominant fashion. This phenomenon was not observed with the HLH domain mutants (Fig. 5, lanes 6–8), which were shown to be defective in dimerization. Suppression of the binding to DNA occurs when a mutant in the basic (DNA binding) region forms a heterodimer complex with a wild-type protein. This finding provided strong evidence that both polypeptides in a dimer contribute directly to a functional DNA binding domain and that two intact basic regions are required for interaction with DNA.

DISCUSSION

Experiments presented here establish that the HLH region common to a family of genes is responsible for the dimerization of the immunoglobulin κ gene enhancer binding protein, E47. Substituting charged amino acids for pairs of relevant hydrophobic amino acid residues in the proposed amphipathic helices is sufficient to disrupt dimer formation. Complex formation, however, remains unperturbed by mutation of a single hydrophobic residue. This observation is similar to the effect of the single and double mutations of the "leucine zipper" of the Fos-Jun heterodimer (17, 18). In addition to the hydrophobic amino acids, the lysine conserved among the HLH family that is positioned at the amino terminus of helix II is required for dimerization. Substituting alanine for this lysine abolished dimerization and subsequent DNA binding, implying that this charged residue contributes to the interaction between the two polypeptides. All of the mutations that interfered with protein dimerization also abolished binding to DNA. Therefore, interaction of the amphipathic helixes may have a fundamental role in assembling a functional DNA binding surface of protein. The substitutions of residues in the loop region, although targeted randomly, did not produce any significant changes in protein dimerization and DNA binding. Because the region is variable in the amino acid sequence and length among the HLH family, it is unlikely that the loop directly contributes to the dimerization between the HLH proteins. It probably serves as a structural element that separates the two helices and facilitates their spatial juxtaposition during dimerization. Direct structural experiments will be required for a detailed understanding of the complex formation between E47 polypeptides.

Mutagenesis of the basic region found immediately amino-terminal of helix I shows that its highly conserved arginine residues are required for DNA binding by the E47 dimer. This DNA–protein recognition is not due simply to ionic interaction between the charges of the phosphate backbone of the DNA and basic amino acid residues, because even a single conservative substitution of another basic amino acid, lysine, for the relevant arginines still disrupted DNA binding. These results define a specific DNA binding motif for the HLH family of proteins. Similarly, a study of mutants of a muscle differentiation factor, MyoD, defined the HLH domain together with the basic region as necessary and sufficient for DNA binding and myogenesis (19).

A model for a well-studied family of DNA binding proteins that contain the leucine zipper structure was proposed recently (20). This model is based on the recognition of the dyad symmetric DNA binding site by two DNA contact surfaces brought together due to a specific interaction of a dimerization interface of the leucine zipper. Since the DNA target sequence recognized by the E47 protein present in the immunoglobulin gene enhancer has a dyad symmetry structure (3), it is likely that dimerization of the E47 polypeptides juxtaposes their basic domains with the κE2 DNA motif. This model is in good agreement with the finding that although the E47 basic region mutants were able to form dimers with the wild-type E47 protein, the dimers were deficient in binding to DNA. These basic region mutants of E47 are "trans-dominant" in that they interfere with DNA binding by the wild-type E47P protein. *In vivo*, it is possible that E47 usually binds to DNA as a heterodimer with another HLH protein. A negative regulation or modulation of the enhancer could be achieved by dimerization of E47 with a HLH protein lacking a DNA binding domain, similar to the newly reported Id protein (22). The E47 basic region mutants could mimic this effect and therefore could be used for *in vivo* analysis of the biological function of the HLH proteins.

The amino-terminal portion of the E47 and E12 proteins contain a leucine zipper structural motif (1) and a transcription-activating sequence (21). Deletion studies demonstrated that this region of the polypeptides clearly is not required for binding to the κE2 site *in vitro* (1). However it is quite possible that either tissue-specific or ubiquitous factors interact by way of this domain with the E47 protein *in vivo* and regulate the expression of enhancers.

We thank Cornelis Murre, Mark Kamps, Richard Van Etten, and Sankar Ghosh for helpful advice and discussion. We also thank Robert Weinberg and Dan Silver for reading the manuscript. A.V. is supported by a postdoctoral fellowship from The American Cancer Society. This work was supported by U.S. Public Health Service Grant GM 39458.

1. Murre, C., McCaw, P. S. & Baltimore, D. (1989) *Cell* **56,** 777–783.
2. Murre, C., McCaw, P. S., Vaessin, H., Caudy, M., Jan, L. Y., Jan, Y. N., Cabrera, C., Buskin, D. C., Lassar, A. B., Weintraub, H. & Baltimore, D. (1989) *Cell* **58,** 537–544.
3. Sen, R. & Baltimore, D. (1986) *Cell* **46,** 705–716.
4. Lenardo, M., Pierce, J. W. & Baltimore, D. (1987) *Science* **236,** 1573–1577.
5. Moss, L. G., Moss, J. B. & Rutter, W. J. (1988) *Mol. Cell. Biol.* **8,** 2620–2627.
6. Buskin, J. N. & Hauschka, S. D. (1989) *Mol. Cell. Biol.* **9,** 2627–2640.
7. Davis, R. L., Weintraub, H. & Lassar, A. B. (1987) *Cell* **51,** 987–1000.
8. Tapscott, S. J., Davis, R. L., Thayer, M. J., Cheng, P. F., Weintraub, H. & Lassar, A. B. (1988) *Science* **242,** 405–411.
9. Wright, W. E., Sassoon, D. A. & Lin, V. K. (1989) *Cell* **56,** 607–617.
10. Braun, T., Buschhausen-Denker, G., Bober, E., Tannich, E. & Arnold, H. H. (1989) *EMBO J.* **8,** 701–709.
11. Alonso, M. C. & Cabrera, C. V. (1988) *EMBO J.* **7,** 2585–2591.
12. Villares, R. & Cabrera, C. V. (1987) *Cell* **36,** 101–109.
13. Caudy, M., Vaessin, H., Brand, M., Tuma, R., Jan, L. Y. & Jan, Y. N. (1988) *Cell* **55,** 1061–1067.
14. Thisse, B., Stoetzel, C., Gorostiza-Thisse, C. & Perrin-Schmitt, F. (1988) *EMBO J.* **7,** 2175–2183.
15. Klaembt, C., Knust, E., Tietze, K. & Campos-Ortega, J. A. (1989) *EMBO J.* **8,** 203–211.
16. Mellentin, J. D., Smith, S. D. & Clearly, M. L. (1989) *Cell* **58,** 77–83.
17. Gentz, R., Rauscher, F. J., III, Abate, C. & Curran, T. (1989) *Science* **243,** 1695–1699.
18. Turner, R. & Tjian, R. (1989) *Science* **243,** 1689–1694.
19. Davis, R. L., Cheng, P.-F., Lassar, A. B. & Weintraub, H. (1990) *Cell* **60,** 733–746.
20. Vinson, C. R., Sigler, P. B. & McKnight, S. L. (1989) *Science* **246,** 911–916.
21. Henthorn, P., Kiledjian, M. & Kadesch, T. (1990) *Science* **247,** 467–470.
22. Benezra, R., Davis, R. L., Lockshon, D. L. & Weintraub, H. (1990) *Cell* **61,** 49–59.