Proc. Natl. Acad. Sci. USA
Vol. 88, pp. 6162–6166, July 1991
Genetics

Determination of the c-MYC DNA-binding site

(oncogene/transcription factor TFEB/helix–loop–helix motif)

THANOS D. HALAZONETIS AND AMANY N. KANDIL

Department of Cancer Research, Merck Sharp & Dohme Research Laboratories, West Point, PA 19486

Communicated by Philip Leder, March 29, 1991

ABSTRACT **The carboxyl terminus of the protein encoded by the c-*MYC* protooncogene has similarity to the helix–loop–helix family of DNA-binding proteins and recognizes a six-nucleotide-long DNA sequence. We have used *in vitro*-translated c-MYC protein to further define its DNA-binding specificity. The hexanucleotide originally identified is necessary for DNA binding by c-MYC, but not sufficient; the c-MYC target site is a 12-nucleotide-long palindrome. This site is present within regulatory regions of genes that are expressed during cell growth. Point mutations within the helix–loop–helix motif of c-MYC abolish DNA-binding and transforming activities, indicating that c-MYC acts as a DNA-binding protein to transform cells. c-MYC may transform cells by activating transcription of genes required for cell division.**

Constitutive expression of the c-*MYC* oncogene leads to cellular transformation *in vitro* and the development of tumors in experimental animals (1, 2). Furthermore human tumors are often associated with mutations that activate this gene (3–6).

Within the 90 carboxyl-terminal amino acids of c-MYC, the protein product of the c-*MYC* oncogene, two motifs can be recognized; the helix–loop–helix (HLH) motif, which is preceded by a region rich in basic amino acids, and the leucine zipper motif, which immediately follows the HLH motif (7, 8). Similar motifs are found in sequence-specific DNA-binding proteins, where they serve to mediate protein dimerization and subsequent DNA binding (7–11).

The presence of these motifs in c-MYC suggests that c-MYC has sequence-specific DNA-binding activity. Indeed Blackwell *et al.* (12) expressed in *Escherichia coli* a fusion protein of glutathione *S*-transferase with the carboxyl terminus of c-MYC, and Prendergast and Ziff (13) *in vitro* translated an E12 HLH protein, which had 11 amino acids of the basic region substituted by the corresponding amino acids of c-MYC. Both the above fusion proteins recognize the DNA sequence CACGTG. To address whether the DNA-binding site of c-MYC extends beyond these six nucleotides, we incubated *in vitro*-translated c-MYC protein with a repertoire of DNA sequences and identified a 12-nucleotide-long sequence that is both necessary and sufficient for DNA binding by c-MYC. Furthermore, point mutations within the c-MYC DNA-binding domain abolish both DNA binding and transforming activities, indicating that c-MYC acts as a sequence-specific DNA-binding protein to transform cells.

MATERIALS AND METHODS

Recombinant Plasmids. Classical recombinant technology was used (14). Plasmid pGEMmycB encodes a full-length human c-MYC protein (Fig. 1*A*). It is based on the pGEM4 vector (Promega). First the *Eco*RI–*Bam*HI fragment of the polylinker was replaced by synthetic oligonucleotides that provide a protein translation initiation codon preceded by a consensus translation initiation sequence (15). The c-*MYC* coding sequence was cloned as two separate fragments. The *Bam*HI–*Cla* I fragment was derived from plasmid pOTS-myc (16), whereas the *Cla* I–*Xba* I fragment was from M13mp11mycRsa. The latter is an M13mp11 vector with a 980-base-pair (bp) genomic *Rsa* I fragment containing the third exon of c-*MYC* subcloned into the *Sma* I site of the polylinker. The *Bam*HI site of the M13mp11 polylinker was previously destroyed.

Plasmid pGEMmyc3X encodes the carboxyl-terminal 178 amino acids of c-MYC. It was derived from pGEMmycB by substituting the *Bam*HI–*Cla* I c-*MYC* fragment with adaptor oligonucleotides (GATCCTGCTCGAGGAAT, upper strand; CGATTCCTCGAGCAG, lower strand).

Plasmid pGEMmycPp encodes the carboxyl-terminal 94 amino acids of c-MYC. This plasmid was derived from pGEMmycB by substituting the *Bam*HI–*Ppu*MI c-*MYC* fragment with adaptor oligonucleotides (GATCCTCTAGACAG, upper strand; GACCTGTCTAGAG, lower strand).

Site-directed mutagenesis was performed as described (17) using M13mp11mycRsa as the template. The mutant c-*MYC* fragments were subcloned into either pGEMmyc3X or pSV7hummyc as *Cla* I–*Bcl* I fragments.

Plasmid pGEMtfebPX encodes a fragment of the human transcription factor TFEB protein that contains the HLH motif and the adjacent leucine zipper. A *Pst* I–*Xho* I fragment of the *TFEB* gene was isolated from plasmid pUC1-4tfeb and cloned into pGEM4, which had the *Eco*RI–*Bam*HI fragment of the polylinker replaced by the oligonucleotides that provide an initiation codon, as described above for pGEMmycB. Adaptor oligonucleotides were used to bridge the *Bam*HI site of the vector to the *Pst* I site of the insert (15). The *Xho* I site of the *TFEB* insert was ligated to the *Sal* I site of the pGEM4 vector.

DNA-Binding Assay. This assay was done as described (15), except that poly(dI·dC) was substituted with 100 ng per reaction of the following single-stranded oligonucleotide: CAGGAAGCAGGTCATGTGGCAAGGCTATT.

Rat Embryo Fibroblast Transformation Assay. This assay was done as described (18), except that the cells were transfected as described by Chen and Okayama (19).

RESULTS

***In Vitro* Translation of c-MYC.** Several DNA-binding proteins, including members of the HLH family, bind DNA with specificity when expressed in an *in vitro* translation system (7, 9, 20). The coding sequence of human c-*MYC* was therefore cloned into the pGEM4 vector (Promega) to generate *in vitro*-transcribed mRNA, which was then translated by using a rabbit reticulocyte lysate (Fig. 1*A*). Truncated versions were also generated. One of these, Myc3X, contained the amino acid sequences encoded by the third exon of c-*MYC*, whereas a further truncated protein, MycPp,

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.

Abbreviation: HLH, helix–loop–helix.

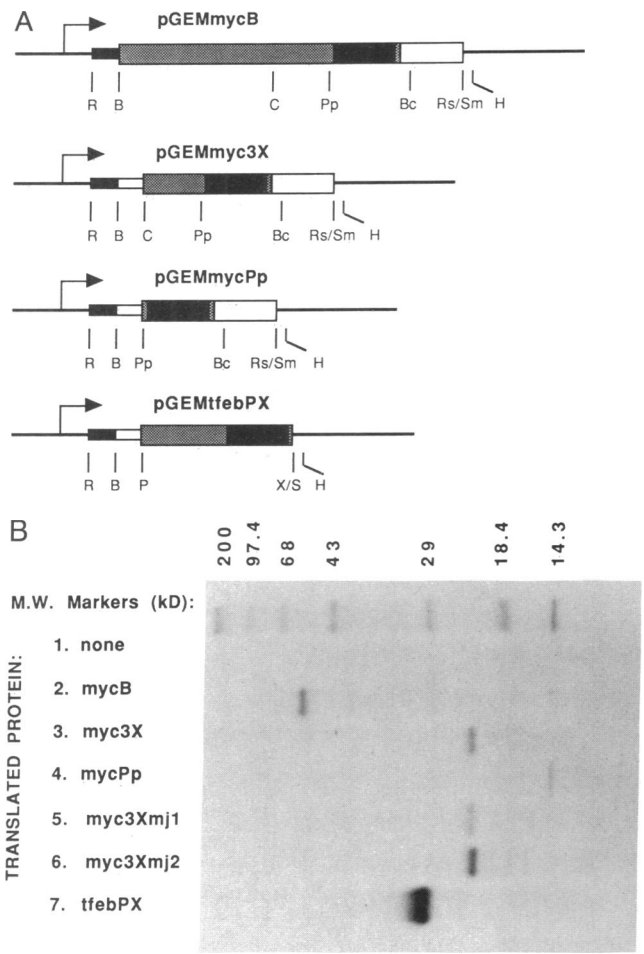

FIG. 1. Plasmids used in this study and *in vitro*-translated c-MYC and TFEB proteins. (*A*) Structure of plasmids pGEMmycB, pGEMmyc3X, pGEMmycPp, and pGEMtfebPX. Thick lines represent pGEM4 (vector) DNA, thin filled rectangles represent synthetic oligonucleotides with an ATG initiation codon, thin clear rectangles represent adaptor oligonucleotides, and thick rectangles represent the c-*MYC* or *TFEB* inserts. Coding sequences are stippled. The filled area within each coding sequence represents the HLH motif and adjacent leucine zipper. Position of the SP6 promoter and direction of transcription are indicated by horizontal arrows. Relevant restriction endonuclease cleavage sites are shown as follows: B, *Bam*HI; Bc, *Bcl* I; C, *Cla* I; R, *Eco*RI; H, *Hin*dIII; Pp, *Ppu*MI; P, *Pst* I; Rs, *Rsa* I; S, *Sal* I; Sm, *Sma* I; and X, *Xho* I. (*B*) SDS/PAGE of *in vitro*-translated proteins. Positions of standard molecular size protein markers indicate direction of electrophoretic migration.

consisted only of the HLH DNA-binding motif and the adjacent leucine zipper (Fig. 1*A*).

A plasmid encoding a partial fragment of human TFEB, a putative transcription factor containing a HLH DNA-binding motif highly homologous to that of c-MYC was also constructed (Fig. 1*A*). TFEB was used as a positive control for DNA binding as its DNA-recognition site is known (21).

All proteins were efficiently translated *in vitro* and migrated on an SDS/polyacrylamide gel in accordance with their molecular weights (Fig. 1*B*).

Sequence-Specific DNA-Binding Mediated by the HLH Domain of c-MYC. The consensus binding site of the HLH class of DNA-binding proteins, referred to as E-box, is NNN-CANNTGNNN, where N can be any nucleotide (7, 9, 10, 21–25). DNA-binding specificity within the HLH family is imparted by the two nucleotides between the invariant CA and TG, hereafter referred to as inner nucleotides, and by the nucleotides flanking the central six nucleotides, hereafter referred to as outer nucleotides.

Taking into account the observations that the binding site of c-MYC has CG as the inner nucleotides (12, 13), we synthesized a battery of palindromic oligonucleotides with a variety of nucleotide combinations flanking the sequence CACGTG. These DNAs were examined for binding to *in vitro*-translated MycPp. Of all the DNAs tested MycPp recognized with highest affinity the sequence GAC-CACGTG-GTC (Fig. 2*A*). The DNAs with GAG and TAG as the outer nucleotides were recognized with lower affinity, whereas the other DNAs examined did not bind to MycPp (Fig. 2*A*).

To determine whether the outer nucleotides GAC represent the optimum target site for c-MYC we substituted each of these nucleotides one at a time with every other possible nucleotide. Two such variants, GAG and GAT, had already been tested (Fig. 2*A*). Of all additional variants tested only GCC failed to bind to MycPp (Fig. 2*B*; note that a rabbit reticulocyte lysate protein formed a complex with the GCC variant DNA, but this complex could be observed even when the reticulocyte lysate was not programmed by c-MYC RNA, data not shown). We conclude that c-MYC recognizes GAC as the outer nucleotides with high affinity but will bind equally well when a mismatch occurs at one of these positions, except for thymidine at the position closest to the central core CACGTG and cytosine at the middle position of the outer nucleotide triplet. Mismatch at two positions of the outer nucleotides is usually not tolerated because the oligonucleotide with CAA as the outer nucleotides did not bind to MycPp (Fig. 2*A*). The inability of this sequence to be recognized over the sequence with GAA as the outer nucleotides (Fig. 2*B*) argues that the target site of c-MYC is 12, rather than 10 nucleotides long.

To further demonstrate the importance of specific outer nucleotides for recognition by c-MYC, we examined the ability of the DNA with GAT as the outer nucleotides (nonspecific DNA) to compete for specific protein–DNA binding. Whereas 50-fold excess of nonlabeled specific DNA completely competed with binding of MycPp to the labeled DNA, the same excess of nonspecific DNA competed only slightly (Fig. 2*C*).

All above examined DNAs contained CG as the inner nucleotides. To assess whether c-MYC had a specific sequence requirement for the inner nucleotides, we generated a variant of the high-affinity recognition site GACCACGTG-GTC, which has the inner nucleotides CG mutated to GC. MycPp failed to bind to this DNA (Fig. 2*D*, lanes 1 and 2). Furthermore 200-fold excess of this variant DNA competed only slightly with MycPp-specific DNA binding, whereas the same excess of specific DNA competed completely (Fig. 2*D*, lanes 5–8).

The general HLH protein consensus recognition sequence can be considered to be formed of two half-sites, each recognized by one subunit of a dimeric complex. We examined whether altering the spacing of the half-sites would eliminate c-MYC DNA binding. A variant of the high-affinity recognition site was generated that had the two inner nucleotides CG replaced by four: CGCG. MycPp failed to bind to this DNA (Fig. 2*D*, lanes 3 and 4).

Recognition of Distinct DNA Sequences by c-MYC and the Transcription Factor TFEB. TFEB is a HLH protein with high homology to c-MYC over the basic region (21), the region that most likely mediates sequence-specific interactions with DNA (10, 11). The DNA-binding site of TFEB protein has been defined as GGC-CACGTG-ACC (21). To determine whether TFEB and c-MYC have identical DNA-binding specificities, we examined DNA binding of TFEB to the panel of palindromic oligonucleotides that vary the outer nucleotide sequences. In contrast to MycPp, TFEB protein required GGT and GAT as outer nucleotide triplets for high-affinity DNA binding (Fig. 3).

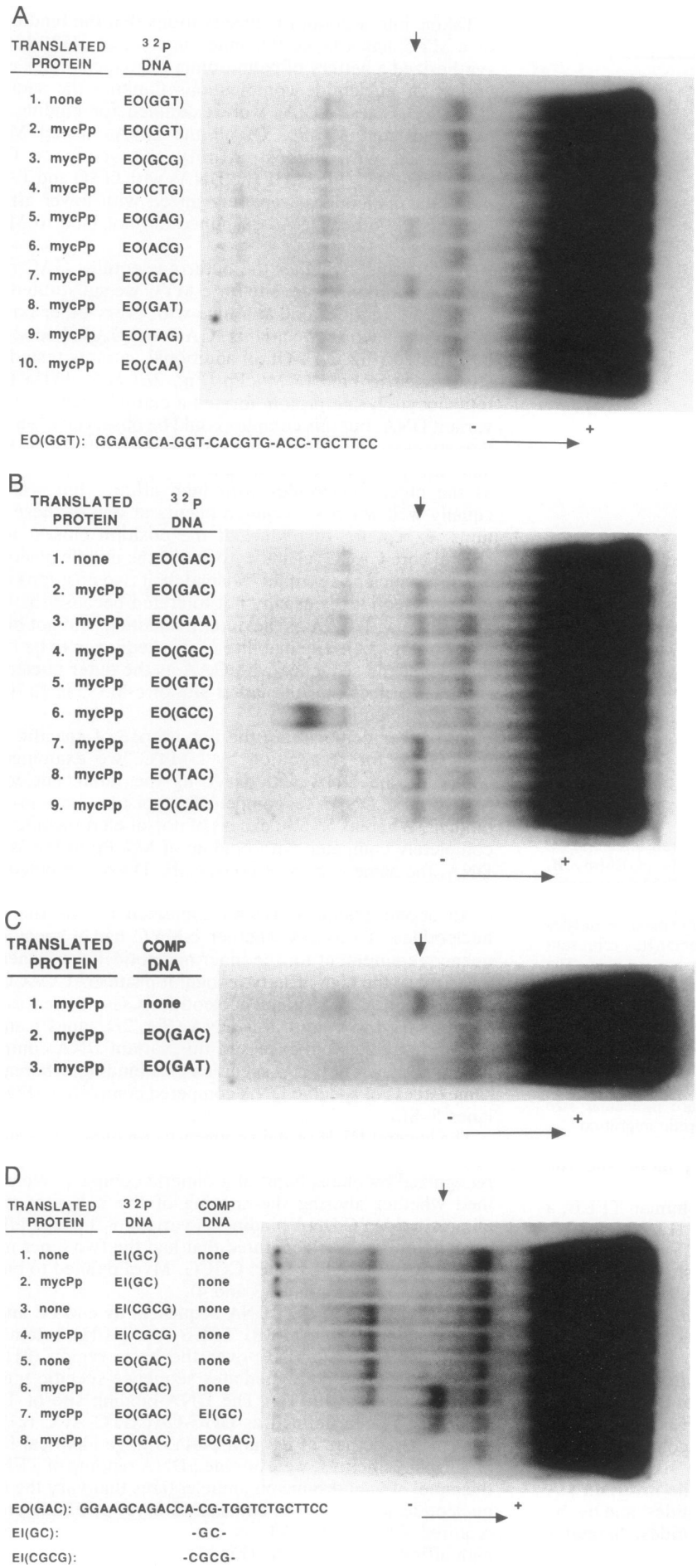

FIG. 2. Determination of DNA-binding specificity of c-MYC. (*A* and *B*) *In vitro*-translated MycPp was assayed for binding to a battery of ^{32}P-labeled oligonucleotides. Sequence of oligonucleotide EO(GGT) is indicated; it is a complete palindrome with the sequence CACGTG at its center. Outer nucleotides (GGT) are indicated by name of this synthetic DNA. All other oligonucleotides are palindromes, identical to EO(GGT), except for outer nucleotides. Names of these synthetic DNAs indicate sequences of outer nucleotides. (*C*) *In vitro*-translated MycPp was assayed for binding to ^{32}P-labeled oligonucleotide EO(GAC) in the presence of excess nonlabeled competitor DNAs. (*D*) *In vitro*-translated MycPp was assayed for binding to the ^{32}P-labeled oligonucleotides EI(GC) and EI(CGCG), which differ from EO(GAC) in the inner nucleotides, as well as for binding to ^{32}P-labeled EO(GAC) in the presence of excess nonlabeled competitor DNAs. Vertical arrows indicate protein–DNA complexes, whereas horizontal arrows indicate direction of electrophoretic migration.

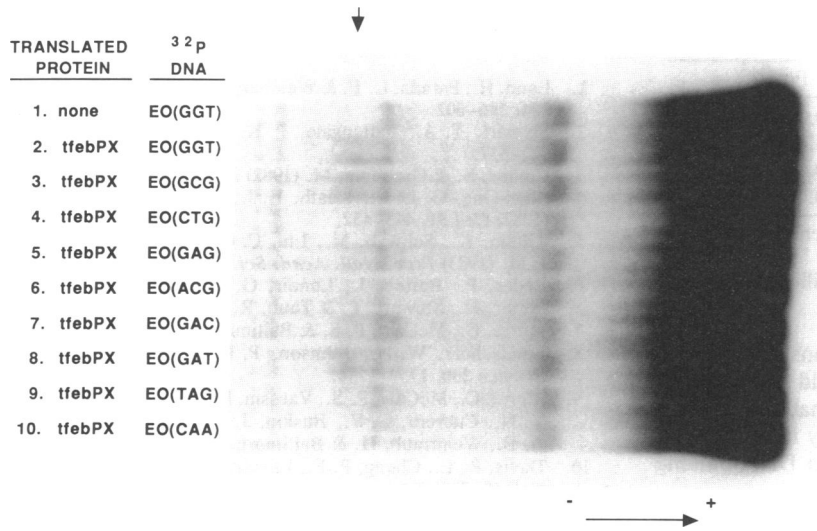

FIG. 3. DNA-binding specificity of transcription factor TFEB. *In vitro*-translated TFEB was assayed for binding to the same battery of ^{32}P-labeled oligonucleotides as in Fig. 2*A*. The affinity of TFEB protein to the DNA with GCG as the outer nucleotides cannot be evaluated from these experiments, as this DNA complexes with a rabbit reticulocyte lysate protein that migrates where a TFEB–DNA complex would be predicted (compare with Fig. 2*A*).

Sequence-Specific DNA Binding by Full-Length c-MYC. MycPp is a truncated version of c-MYC containing only the basic region, HLH, and leucine zipper elements. We examined whether bigger fragments of c-MYC protein would also bind DNA. Myc3X contains the amino acids encoded by the whole c-*MYC* third exon, whereas MycB is a full-length c-MYC protein (Figs. 1 *A* and *B*). Both exhibited specific DNA-binding activity (Fig. 4, lanes 2 and 5).

Point Mutations Within the HLH Motif of c-MYC Abolish Both DNA Binding and Transforming Activity. To assess whether DNA binding is required for the transforming activity of c-MYC, we introduced point mutations within the protein that would abolish DNA binding. Mutant mj1 has four amino acid substitutions within the first α-helix of the HLH motif, whereas mj2 has five substitutions within the upstream basic region (Fig. 4). c-MYC proteins containing these mutations were efficiently translated *in vitro* (Fig. 1*B*) but failed to bind DNA (Fig. 4). The same mutations inactivated the transforming activity of c-*MYC* (Table 1).

DISCUSSION

Determination of the DNA-Binding Site of c-MYC. The c-MYC protein has similarity to the HLH family of DNA-binding proteins. The consensus binding site of the latter, referred to as E-box, is CANNTG, suggesting that c-MYC might bind a related DNA sequence. Indeed Blackwell *et al.* (12) and Prendergast and Ziff (13) have demonstrated that the binding site of c-MYC includes the hexanucleotide CACGTG. We have been able to define the complete binding site of c-MYC as GACCACGTGGTC by using *in vitro*-translated protein.

c-MYC recognizes the same inner nucleotides, CG, as the HLH protein TFEB (21). The similarity in DNA-recognition sequences reflects the high homology of the basic regions of these proteins. Nevertheless, c-MYC and transcription factor TFEB recognize different outer nucleotides, suggesting that they operate in distinct pathways in the cell.

In our search for DNA sequences recognized by c-MYC we have not used every possible combination of outer nucleotide sequences, and thus the question arises whether we have identified the optimum DNA target site. Substituting the outer nucleotides of the optimum site one at a time with every other possible nucleotide did not lead to appreciably increased binding affinity. Furthermore, c-MYC binds the target sequence we have identified as efficiently as transcription factor TFEB binds its optimum recognition site when equal concentrations of the two proteins are used (compare Figs. 2*A* and 3).

Dimerization is a prerequisite for DNA binding by the HLH family of proteins; we therefore speculate that *in vitro*-translated c-MYC binds DNA as a dimer. This hypothesis is consistent with the inability of the protein to bind DNA when spacing of the half-sites is altered. c-MYC protein produced in *E. coli* forms dimers and tetramers (26), whereas *in vitro*-translated c-MYC is a monomer (27). The presence of specific DNA in our assay may drive dimerization of *in vitro*-translated c-MYC.

Linkage of DNA Binding and Transforming Activities of c-MYC. Mutagenesis of the HLH motif of c-MYC abolishes transforming activity (28). However, in the absence of a

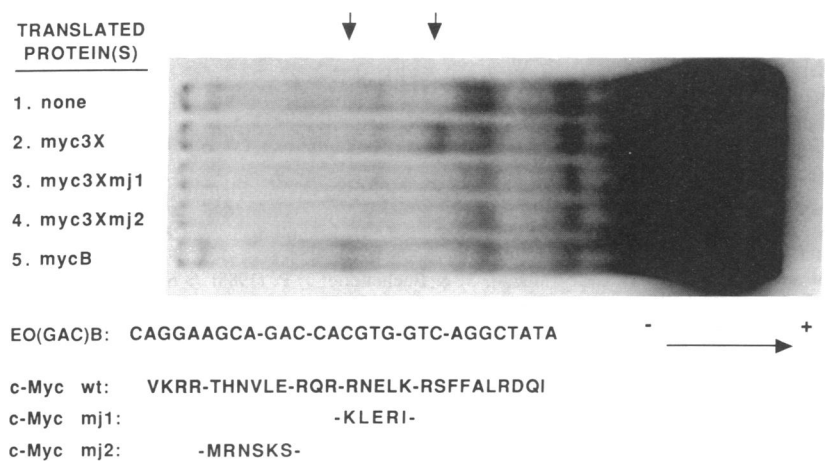

FIG. 4. Specific DNA-binding activity of full-length c-MYC protein and of c-MYC HLH mutants. *In vitro*-translated proteins were assayed for binding to the ^{32}P-labeled oligonucleotide EO(GAC)B. Vertical arrows indicate MycB and Myc3X–DNA complexes. Amino acid substitutions (single-letter code) in c-MYC mutants mj1 and mj2 are shown in comparison with wild-type (wt) sequence of the basic region and adjacent first helix of HLH motif (amino acid positions 354–381 of human c-MYC).

Table 1. Transforming activity of wild-type and mutant c-*MYC* genes

Transfected DNA	Foci*
pSV7neo + pEJ.6.6 (*ras*)	0.0 ± 0.0
pSV7hummyc wt + pEJ.6.6 (*ras*)	7.6 ± 0.6
pSV7hummyc mj1 + pEJ.6.6 (*ras*)	0.0 ± 0.0
pSV7hummyc mj2 + pEJ.6.6 (*ras*)	0.0 ± 0.0

Data represent means ± 1 SD of three independent transfections. wt, Wild type.
*Results are expressed as number of foci per million transfected cells.

DNA-binding assay for c-MYC, the relationship between DNA binding and transforming activities could not be established. Using point mutations, we have eliminated the DNA-binding activity of c-MYC and concomitantly its transforming activity, suggesting that c-MYC acts as a DNA-binding protein to transform cells.

Presence of c-MYC-Binding Sites Within Regulatory Regions of Genes That Are Induced During Cell Growth. The biochemical action of c-MYC after DNA binding has not been established. There is evidence, however, that c-MYC may activate gene transcription (29, 30). Determination of the c-MYC binding site as a 12-nucleotide-long sequence allowed us to identify putative gene targets because the frequency of random occurrence of a sequence of such length is low.

The c-*SIS* protooncogene encodes the B polypeptide of platelet-derived growth factor. Approximately 270 nucleotides upstream of the transcription initiation site, a 12-nucleotide element fits the c-MYC consensus binding site (31). This element is conserved across species (32) and maps within an enhancer region (33).

The mouse *H19* gene encodes a protein of unknown function. The gene was identified by virtue of its coordinate regulation with the α-fetoprotein gene and its inducibility during liver regeneration (34). One of the enhancers of this gene located at the 3′ end contains a c-MYC DNA-binding site (35).

The gene that encodes nucleolin, a major nucleolar protein involved in the synthesis and assembly of ribosomes, contains c-MYC DNA-binding sites within the first intron that are conserved across species (36, 37). These regions have been speculated to be involved in gene transcription because they are bordered by G+C-rich sequences (36).

The U3B RNA gene encodes a small nuclear RNA localized predominantly in the nucleolus that probably functions in pre-rRNA processing (38). The gene is transcribed by RNA polymerase II and contains a c-MYC DNA-binding site within its 5′ flanking sequences (39). The presence of c-MYC DNA-binding sites within both the nucleolin and U3B RNA genes suggests that c-MYC may be involved in regulating ribosome synthesis, a process clearly related to cell growth. Interestingly, cells transformed by c-*MYC* but not by other oncogenes have large nucleoli (40).

The identification of c-MYC binding sites within regulatory regions of genes induced during cell growth suggests that c-MYC functions to regulate cell growth as a factor activating transcription. This hypothesis can now be experimentally tested.

T.D.H. is grateful to P. Leder, in whose laboratory some plasmids were constructed, for continued support and friendship. Both authors thank A. Oliff, J. Shafer, L. Davis, M. Marshall, D. Pompliano, M. Greenberg, K. Georgopoulos, and P. Gentile for support and helpful discussions, G. Prendergast for communicating unpublished observations, P. Sharp for plasmid pUC1-4tfeb, and M. Rosenberg for plasmid pOTS-myc.

1. Land, H., Parada, L. F. & Weinberg, R. A. (1983) *Nature (London)* **304**, 596–602.
2. Stewart, T. A., Pattengale, P. K. & Leder, P. (1984) *Cell* **38**, 627–637.
3. Collins, S. & Groudine, M. (1982) *Nature (London)* **298**, 679–681.
4. Shen-Ong, G. L. C., Keath, E. J., Piccoli, S. P. & Cole, M. D. (1982) *Cell* **31**, 443–452.
5. Alitalo, K., Schwab, M., Lin, C. C., Varmus, H. E. & Bishop, J. M. (1983) *Proc. Natl. Acad. Sci. USA* **80**, 1707–1711.
6. Leder, P., Battey, J., Lenoir, G., Moulding, C., Murphy, W., Potter, H., Stewart, T. & Taub, R. (1983) *Science* **222**, 765–771.
7. Murre, C., McCaw, P. S. & Baltimore, D. (1989) *Cell* **56**, 777–783.
8. Landschulz, W. H., Johnson, P. F. & McKnight, S. L. (1988) *Science* **240**, 1759–1764.
9. Murre, C., McCaw, P. S., Vaessin, H., Caudy, M., Jan, L. Y., Jan, Y. N., Cabrera, C. V., Buskin, J. N., Hauschka, S. D., Lassar, A. B., Weintraub, H. & Baltimore, D. (1989) *Cell* **58**, 537–544.
10. Davis, R. L., Cheng, P.-F., Lassar, A. B. & Weintraub, H. (1990) *Cell* **60**, 733–746.
11. Voronova, A. & Baltimore, D. (1990) *Proc. Natl. Acad. Sci. USA* **87**, 4722–4726.
12. Blackwell, T. K., Kretzner, L., Blackwood, E. M., Eisenman, R. N. & Weintraub, H. (1990) *Science* **250**, 1149–1151.
13. Prendergast, G. C. & Ziff, E. B. (1991) *Science* **251**, 186–189.
14. Ausubel, F. M., Brent, R., Kingston, R. E., Moore, D. D., Seidman, J. G., Smith, J. A. & Struhl, K. (1987) *Current Protocols in Molecular Biology* (Greene/Wiley, New York).
15. Halazonetis, T. D., Georgopoulos, K., Greenberg, M. E. & Leder, P. (1988) *Cell* **55**, 917–924.
16. Watt, R. A., Shatzman, A. R. & Rosenberg, M. (1985) *Mol. Cell. Biol.* **5**, 448–456.
17. Sarid, J., Halazonetis, T. D., Murphy, W. & Leder, P. (1987) *Proc. Natl. Acad. Sci. USA* **84**, 170–173.
18. Halazonetis, T. D., Daugherty, C. & Leder, P. (1988) *Mol. Cell. Biol.* **8**, 1845–1848.
19. Chen, C. & Okayama, H. (1987) *Mol. Cell. Biol.* **7**, 2745–2752.
20. Hope, I. A. & Struhl, K. (1985) *Cell* **43**, 177–188.
21. Carr, C. S. & Sharp, P. A. (1990) *Mol. Cell. Biol.* **10**, 4384–4388.
22. Beckmann, H., Su, L.-K. & Kadesch, T. (1990) *Genes Dev.* **4**, 167–179.
23. Cai, M. & Davis, R. W. (1990) *Cell* **61**, 437–446.
24. Gregor, P. D., Sawadogo, M. & Roeder, R. G. (1990) *Genes Dev.* **4**, 1730–1740.
25. Mellor, J., Jiang, W., Funk, M., Rathjen, J., Barnes, C. A., Hinz, T., Hegemann, J. H. & Philippsen, P. (1990) *EMBO J.* **9**, 4017–4026.
26. Dang, C. V., McGuire, M., Buckmire, M. & Lee, W. M. F. (1989) *Nature (London)* **337**, 664–666.
27. Smith, M. J., Charron-Prochownik, D. C. & Prochownik, E. V. (1990) *Mol. Cell. Biol.* **10**, 5333–5339.
28. Stone, J., De Lange, T., Ramsay, G., Jakobovits, E., Bishop, J. M., Varmus, H. & Lee, W. (1987) *Mol. Cell. Biol.* **7**, 1697–1709.
29. Kingston, R. E., Baldwin, A. S., Jr., & Sharp, P. A. (1984) *Nature (London)* **312**, 280–282.
30. Kato, G. J., Barrett, J., Villa-Garcia, M. & Dang, C. V. (1990) *Mol. Cell. Biol.* **10**, 5914–5920.
31. Rao, C. D., Pech, M., Robbins, K. C. & Aaronson, S. A. (1988) *Mol. Cell. Biol.* **8**, 284–292.
32. Van den Ouweland, A. M. W., Roebroek, A. J. M., Schalken, J. A., Claesen, C. A. A., Bloemers, H. P. J. & Van de Ven, W. J. M. (1986) *Nucleic Acids Res.* **14**, 765–778.
33. Ratner, L. (1989) *Nucleic Acids Res.* **17**, 4101–4115.
34. Pachnis, V., Belayew, A. & Tilghman, S. M. (1984) *Proc. Natl. Acad. Sci. USA* **81**, 5523–5527.
35. Yoo-Warren, H., Pachnis, V., Ingram, R. S. & Tilghman, S. M. (1988) *Mol. Cell. Biol.* **8**, 4707–4715.
36. Bourbon, H.-M., Prudhomme, M. & Amalric, F. (1988) *Gene* **68**, 73–84.
37. Srivastava, M., McBride, O. W., Fleming, P. J., Pollard, H. B. & Burns, A. L. (1990) *J. Biol. Chem.* **265**, 14922–14931.
38. Busch, H., Reddy, R., Rothblum, L. & Choi, Y. C. (1982) *Annu. Rev. Biochem.* **51**, 617–654.
39. Mazan, S. & Bachellerie, J.-P. (1988) *J. Biol. Chem.* **263**, 19461–19467.
40. Palmieri, S., Kahn, P. & Graf, T. (1983) *EMBO J.* **2**, 2385–2389.