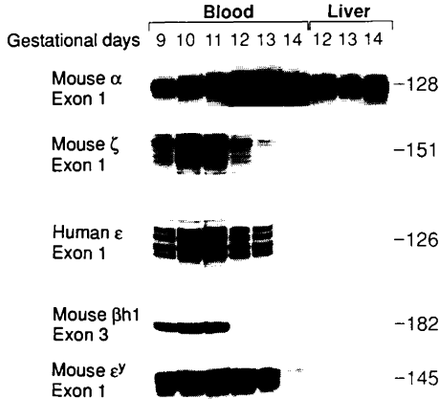

Fig. 4. RNase protection analysis of ϵ-globin RNA from peripheral blood and liver in staged transgenic embryos. The probes used are listed at the left and the size of protected products in nucleotides are indicated at the right. The RNA (100 ng) was prepared as described (*18*) and used for each protection analysis. Probes were from pSP64Mα (*19*), pSP64Mz (*19*), pSP65βh1 (*20*), pSP65Mϵʸ (*19*), and pSP64Hϵ$_x$ (*19*).

two mechanisms that control hemoglobin switching in humans: an autonomous mode, as exemplified by the ϵ gene, and a competitive mode, as illustrated by the γ to β switch. Autonomous and competitive mechanisms have been proposed for hemoglobin switching in the chicken (*15*). The transfected adult β-globin gene is only expressed in definitive cells of the chicken (*15, 16*) and thus appears to be autonomous in its developmental regulation. In contrast, the developmental regulation of the chicken embryonic ϵ genes appears to be competitive, because the restriction of its expression to primitive cells depends on the presence of the β-globin gene on the same plasmid (*15, 16*). Whereas the mechanisms of switching appear to be similar in these two species, the choice of the mechanisms with respect to specific genes is reversed. This is perhaps less surprising than it may seem, because at the divergence of aves and mammalia, only a single β-like globin gene existed (*17*). The duplications that gave rise to the human and chicken β-globin families occurred later and, therefore, independently in the two species.

REFERENCES AND NOTES

1. D. Tuan, W. Solomon, Q. Li, I. M. London, *Proc. Natl. Acad. Sci. U.S.A.* **82**, 6384 (1985); W. C. Forester, C. Thompson, J. T. Elder, M. Groudine, *ibid.* **83**, 1359 (1986).
2. F. Grosveld, G. B. Van Assendelft, D. R. Greaves, G. Kollias, *Cell* **51**, 975 (1987).
3. T. M. Ryan, R. R. Behringer, T. M. Townes, R. D. Palmiter, R. L. Brinster, *Proc. Natl. Acad. Sci. U.S.A.* **86**, 37 (1989); O. Hanscombe *et al.*, *Genes Dev.* **3**, 1572 (1989).
4. D. Talbot *et al.*, *Nature* **338**, 352 (1989).
5. G. B. Van Assendelft, O. Hanscombe, F. Grosveld, D. R. Greaves, *Cell* **56**, 969 (1989).
6. T. Enver, A. J. Ebens, W. C. Forrester, G. Stamatoyannopoulos, *Proc. Natl. Acad. Sci. U.S.A.* **86**, 7033 (1989).
7. R. R. Behringer, T. M. Ryan, R. D. Palmiter, R. Z. Brinster, T. M. Townes, *Genes Dev.* **4**, 380 (1990); T. Enver *et al.*, *Nature* **344**, 309 (1990).
8. F. E. Baralle, C. C. Shoulders, N. J. Proudfoot, *Cell* **21**, 621 (1980).
9. The hypersensitive fragments that constitute the μLCR and their nucleotide sequence coordinates (GenBank HUMBB accession no J00179) are as follows: Hind III 13769 to Nco I 13062; Bgl II 9218 to Hind III 8486; Hind III 5172 to Bal I 4608; Sph I 1702 to Pst I 1182. W. C. Forrester, U. Novak, R. Gelinas, M. Groudine, *Proc. Natl. Acad. Sci. U.S.A.* **86**, 5439 (1989).
10. In this construct, the LCR is linked to the ϵ-globin gene in the reverse orientation genomic configuration. Previous studies have shown that LCR functions in an orientation-independent manner in transgenic mice (*4*).
11. F. Constantini *et al.*, *Cold Spring Harbor Symp. Quant. Biol.* **50**, 361 (1985).
12. Transgenic mice with the human ϵ gene lacking the cis-linked LCR sequences have failed to show expression of the transgene at any developmental stage. Similar results were obtained in transgenic analyses of the human α-globin gene (*3*).
13. The mouse has two major functional β-like embryonic genes (βh1, ϵʸ) with different profiles of expression during the maturation of primitive erythroid cells, βh1 is maximally expressed before ϵʸ. The pattern of human ϵ-globin extinction in the blood of the staged embryos more closely resembled the ϵʸ profile than the βh1 profile. This implied functional relation between human ϵ and mouse ϵʸ is consistent with the evolutionary relation predicted from structural comparisons [A. Hill *et al.*, *J. Biol. Chem.* **259**, 3739 (1984)].
14. Transient assays in erythroid versus nonerythroid cells have revealed a transcriptional silencer in the ϵ promoter region [S.-X. Cao, P. D. Gutman, H. P. G. Dave, A. N. Schechter, *Proc. Natl. Acad. Sci. U.S.A.* **86**, 5306 (1989)].
15. O. R. B. Choi and J. D. Engel, *Cell* **55**, 17 (1988).
16. J. M. Nickol and G. Felsenfeld, *Proc. Natl. Acad. Sci. U.S.A.* **85**, 2548 (1988).
17. M. Goodman, J. Czeiusniak, B. F. Koop, D. A. Tagle, J. L. Slightam, *Cold Spring Harbor Symp. Quant. Biol.* **52**, 875 (1987).
18. J. Karlinsey, G. Stamatoyannopoulos, T. Enver, *Anal. Biochem.* **180**, 303 (1989).
19. M. H. Baron and T. Maniatis, *Cell* **46**, 591 (1986).
20. K. Chada, J. Magram, F. Constantini, *Nature* **319**, 685 (1986).

21. We thank R. Perlmutter for producing the transgenic animals used in this study and D. Chui for the supply of monoclonal antibody to ϵ. Supported by NIH grant DK 3132 and HL 20899.

19 June 1990; accepted 14 September 1990

Sequence-Specific DNA Binding by the c-Myc Protein

T. KEITH BLACKWELL, LEO KRETZNER, ELIZABETH M. BLACKWOOD, ROBERT N. EISENMAN, HAROLD WEINTRAUB

While it has been known for some time that the c-Myc protein binds to random DNA sequences, no sequence-specific binding activity has been detected. At its carboxyl terminus, c-Myc contains a basic–helix-loop-helix (bHLH) motif, which is important for dimerization and specific DNA binding, as demonstrated for other bHLH protein family members. Of those studied, most bHLH proteins bind to sites that contain a CA– –TG consensus. In this study, the technique of selected and amplified binding-sequence (SAAB) imprinting was used to identify a DNA sequence that was recognized by c-Myc. A purified carboxyl-terminal fragment of human c-Myc that contained the bHLH domain bound in vitro in a sequence-specific manner to the sequence, CACGTG. These results suggest that some of the biological functions of Myc family proteins are accomplished by sequence-specific DNA binding that is mediated by the carboxyl-terminal region of the protein.

DESPITE A GREAT DEAL OF RESEARCH indicating that the c-*myc* oncogene functions in cell proliferation and differentiation, the molecular mechanisms of myc function remain unknown. Recent evidence suggests that c-*myc* may be involved in transcription, DNA replication or both (*1*), and thus might be expected to have a sequence-specific DNA-binding activity. However, although c-Myc has been shown to bind to random DNA sequences, no sequence-specific DNA-binding activity has been demonstrated (*2, 3*). c-Myc and other Myc family proteins contain a basic–helix-loop-helix (bHLH) domain, a conserved region that mediates DNA binding by other bHLH proteins (*4, 5*). The bHLH domain consists of the proposed HLH motif (*4*), which mediates homo- and heterooligomerization among certain bHLH family members (*6–8*). Immediately NH_2-terminal to the HLH region is a basic region that contacts DNA (*7, 8*). The c-Myc bHLH region and an immediately COOH-terminal leucine zipper motif (*1*) are essential for the cotransformation properties of myc (*9*). The binding sites that have been identified for other bHLH proteins contain a CA– –TG consensus (*4, 5,*

Division of Basic Sciences, Fred Hutchinson Cancer Research Center, 1124 Columbia Street, Seattle, WA 98104.
T. K. Blackwell and H. Weintraub, Howard Hughes Medical Institute, Fred Hutchinson Cancer Research Center, Seattle, WA 98104.
E. M. Blackwood, Department of Pathology, University of Washington, Seattle, WA 98195.

10, 11). However, bHLH protein complexes differ in their affinities for sequences that contain this motif (*11*), and certain CA– –TG sites that are bound efficiently in vitro by other bHLH proteins are not bound by Myc family members (*4, 6, 12*).

The SAAB imprinting technique (*11*) and other similar strategies (*13*) have made it possible to efficiently identify specific nucleic acid sequences that are bound by a protein complex. In this approach, oligonucleotide templates that contain a region of random DNA sequence are selected for binding to a protein complex of interest, and bound DNA is subsequently amplified by the polymerase chain reaction (PCR). Successive rounds of selection and amplification are used to enrich for templates that bind with the highest affinity. Bound templates can then be analyzed directly and rapidly by sequencing as a pool, providing an "imprint" of protein-DNA binding. SAAB imprinting has been used to explore the DNA-binding sequence preferences of homo- and heterooligomeric bHLH protein complexes (*11*). In this report, we describe its use to identify a CA– –TG site that is bound by c-Myc protein in vitro.

Preparations of purified, full-length c-Myc protein are often insoluble (*3*). Thus, we used a purified preparation of a soluble, bacterially produced glutathione *S*-transferase (GST)-c-myc fusion protein (c-MycC92) that contained the COOH-terminal 92 amino acids of human c-Myc (Fig. 1A) to test for specific DNA binding. The myc sequences of the fusion protein included the bHLH domain and leucine zipper, and therefore should contain the minimal structural requirements for DNA binding. By analogy, a truncated version of the myogenic determination protein, MyoD, that contains only the bHLH region binds specifically to appropriate DNA sequences (*5*). Our experimental DNA template (D6, Fig. 1B) was designed on the basis of a site in the muscle creatine kinase (MCK) enhancer (*14*) to which MyoD binds in vitro (*5, 7*), but had random nucleotide sequences at ten positions within and flanking the CA– –TG consensus.

Binding of c-MycC92 protein to D6 sequences was not readily detectable by the electrophoretic mobility shift assay (EMSA) (Fig. 1C, lane 1). However, the template derived from two successive rounds of SAAB isolation (c-MycD6^2) (*15*) bound to c-MycC92 in two distinct complexes (Fig. 1C, lane 2). The template isolated from the faster migrating complex (c-MycD6^3) (*15*) also bound to c-MycC92, in an apparently analogous set of complexes (Fig. 1C, lane 3). Formation of these complexes with the initial D6 population can be detected, but only after long exposures (*16*).

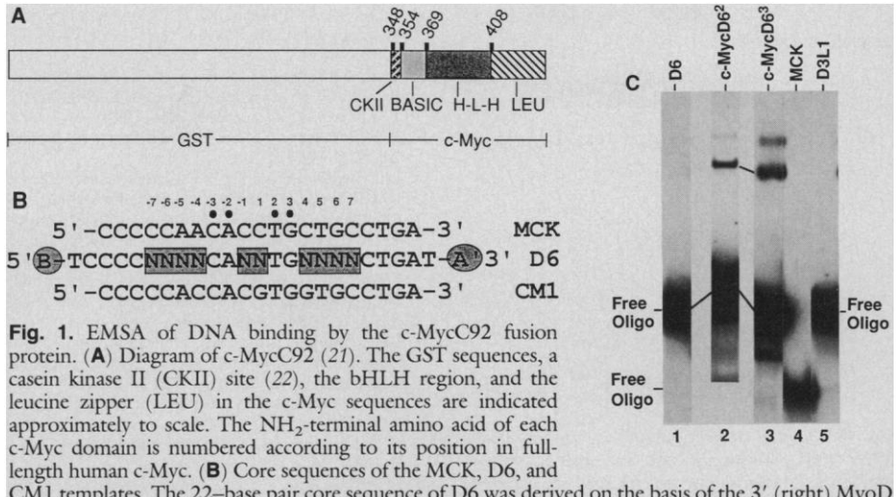

Fig. 1. EMSA of DNA binding by the c-MycC92 fusion protein. (**A**) Diagram of c-MycC92 (*21*). The GST sequences, a casein kinase II (CKII) site (*22*), the bHLH region, and the leucine zipper (LEU) in the c-Myc sequences are indicated approximately to scale. The NH_2-terminal amino acid of each c-Myc domain is numbered according to its position in full-length human c-Myc. (**B**) Core sequences of the MCK, D6, and CM1 templates. The 22–base pair core sequence of D6 was derived on the basis of the 3′ (right) MyoD binding site of the MCK enhancer (read 5′ to 3′ toward the MCK promoter) (*14*), but was synthesized with random sequence at the positions indicated by N. The D6 sequences shown are flanked on the 5′ end by primer B and on the 3′ end by the complement of primer A (indicated as A′) (*23*). The consensus CA– –TG motif is indicated by dots. (**C**) EMSA of c-MycC92 protein binding to the indicated templates (*24*).

We determined the nucleotide sequences of the D6 and c-MycD6^3 template pools (Fig. 2). After PCR amplification of D6, the positions indicated in Fig. 1B were of apparently random sequence (Fig. 2). In contrast, c-MycD6^3 was characterized by a distinct symmetrical preference for C and G at positions −1 and 1, respectively (Fig. 2), and by less distinct preferences at more distal positions. Because successive rounds of SAAB isolation enrich for templates that bind with higher affinities (*11*), we are undertaking further rounds of selection to better define these distal sequence preferences. The template bound in the upper complex (Fig. 1C, lane 2) had a sequence pattern that was identical to that of c-MycD6^3, indicating that both the upper and lower c-MycC92–DNA complexes involved binding to the same sequence. c-MycC92 did not bind to either of two CA– –TG templates of different sequences (Fig. 1C, lanes 4 and 5), which are bound specifically by other factors (*17*), thus confirming that the selection for specific binding sites is inherent in the SAAB-imprinting assay.

An oligonucleotide template (CM1, Fig. 1B) that was synthesized on the basis of the c-MycD6^3 sequence bound to c-MycC92, but not to GST alone (Fig. 3, lanes 1 and 2), indicating that we identified DNA sequences that were specifically bound by the c-Myc portion of the fusion protein, c-MycC92. Concurrent incubation of CM1 with c-MycC92 and an antibody to the 12 COOH-terminal amino acids of c-Myc (anti–c-Myc 12C) (*18*) changed the mobility of the complexes (Fig. 3, lane 3). Addition of the peptide to which the antibody was generated blocked the interaction of the antibody with the protein-DNA complexes, and no complex was observed between CM1 and the antibody alone (Fig. 3, lanes 4 to 6). These findings confirm that the observed complexes contain c-MycC92 protein.

Our results have demonstrated that the bHLH–leucine zipper domain of c-Myc

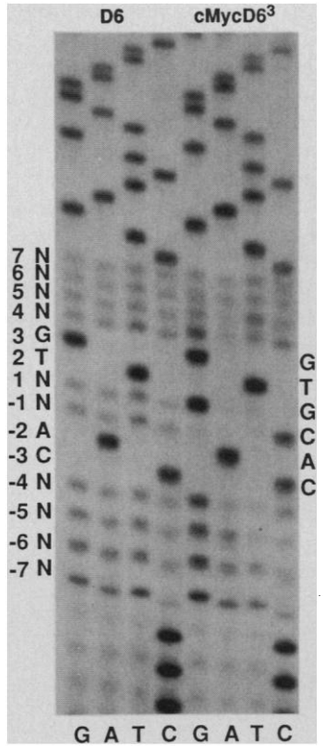

Fig. 2. Nucleotide sequences of selected c-Myc binding sites. Nucleotide sequencing was performed with labeled primer B by the dideoxy method as in (*11*). The D6 sequence shown is of DNA that had undergone 35 cycles of PCR amplification from 1 pg of template (*11*). Experimentally derived sequence preferences at positions −1 and +1 are indicated to the right.

binds to a specific sequence that contains the CA– –TG consensus. The finding that c-MycC92 binds to only a subset of CA– –TG sites is consistent with the idea that different groups of related bHLH proteins might recognize different sets of CA– –TG sites (*11*). Because we have not altered any of the nucleotides in the CA– –TG motif, we cannot say how important any of these four residues is for Myc binding. The c-Myc binding site, CACGTG, identified by our assay is identical to a site in the adenovirus major late promoter that is bound by a cellular bHLH protein, upstream sequence factor (USF) (*19*). USF is similar to Myc family proteins in its basic region and contains an apparent leucine zipper immediately COOH-terminal to the HLH motif (*19*).

Results from SAAB imprint analyses have suggested that bHLH proteins bind to CA– –TG sites in homo- and heterodimeric complexes in which each basic region binds to a half-site that contains one-half of the consensus (*11*). By analogy, we predict that c-Myc would also bind in dimeric complexes; this idea is supported by the symmetry of the CACGTG core sequence of its preferred site. The c-MycC92–DNA complexes (Figs. 1 and 3) might thus represent mono- and multimeric forms of such homodimers. Indeed, c-Myc dimers and tetramers have been observed with relatively high concentrations of bacterially produced proteins (*20*). Formation of dimers and oligomers in vitro is characteristic of other bHLH proteins (*4–8, 11*), but, in the case of c-Myc, may be of special significance because of the potential participation of the leucine zipper in oligomerization.

Our findings demonstrate that c-Myc can act as a sequence-specific DNA-binding protein. Other bHLH proteins appear to function in vivo in heterooligometric complexes (*6, 7, 12*). c-Myc might also bind in vivo to DNA in concert with various oligomerization partners. If so, by analogy to MyoD and its constitutively expressed bHLH partner, E2A, the preferred DNA binding site for such a complex should consist of two half-sites. One half-site should correspond to one half of the sequence described herein, which is recognized by c-Myc, while the other half-site would consist of the sequence preferred by its partner (*11*). If such a partner for c-Myc is identified, the SAAB imprinting technique should be able to test this prediction.

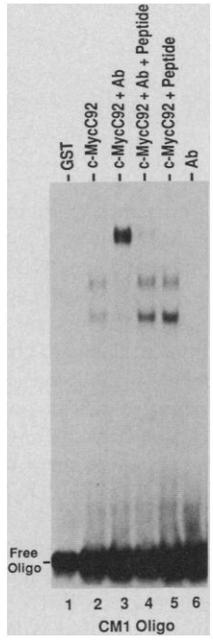

Fig. 3. Binding of c-MycC92 to the CM1 template is specific to c-Myc protein sequences. Binding of the indicated protein and peptide preparations to CM1 template was analyzed by EMSA (*25*). The binding reactions indicated as Ab contained anti–c-Myc 12C (1 μg) (*18*), which was raised against a peptide consisting of the 12 most COOH-terminal amino acids of human c-Myc protein. In the designated reactions, peptide (10 μg) was incubated in the reaction mixture for 5 min at room temperature either with or without the anti–c-Myc 12C prior to addition of DNA and c-MycC92. All other reaction components were mixed simultaneously in each sample.

Note added in proof: G. Prendergast and E. Ziff (personal communication) have found the core sequence of the c-Myc DNA binding site to be the same as the sequence of the site reported herein.

REFERENCES AND NOTES

1. R. G. Collum and F. W. Alt, *Cancer Cells* **2**, 69 (1990); B. Lüscher and R. N. Eisenman, *Genes Dev.*, in press.
2. P. Donner *et al.*, *Nature* **296**, 262 (1982); R. N. Eisenman *et al.*, *Mol. Cell. Biol.* **5**, 114 (1985).
3. R. A. Watt, A. R. Shatzman, R. Rosenberg, *Mol. Cell. Biol.* **5**, 448 (1985).
4. C. Murre, P. S. McCaw, D. Baltimore, *Cell* **56**, 777 (1989).
5. A. B. Lassar *et al.*, *ibid.* **58**, 823 (1989).
6. C. Murre *et al.*, *ibid.*, p. 537.
7. R. L. Davis, P. F. Cheng, A. B. Lassar, H. Weintraub, *ibid.* **60**, 733 (1990).
8. A. Voronova and D. Baltimore, *Proc. Natl. Acad. Sci. U.S.A.* **87**, 4722 (1990).
9. J. Stone *et al.*, *Mol. Cell. Biol.* **7**, 1697 (1987); E. M. Blackwood and R. N. Eisenman, unpublished data.
10. M. Cai and R. W. Davis, *Cell* **61**, 437 (1990).
11. T. K. Blackwell and H. Weintraub *Science* **250**, 1104 (1990).
12. A. B. Lassar and H. Weintraub, unpublished data.
13. C. Tuerk and L. Gold, *Science* **249**, 511 (1990); R. Green, A. D. Ellington, J. W. Szostak, *Nature* **347**, 406 (1990).
14. E. A. Sternberg *et al.*, *Mol. Cell. Biol.* **8**, 2896 (1988); J. N. Buskin and S. D. Hauschka, *ibid.* **9**, 2627 (1990).
15. In each SAAB isolation round, protein-bound DNA templates were isolated by EMSA, eluted from the gel, amplified by PCR (35 cycles), then labeled by PCR as described (*11*). In the first round of EMSA, the D6 template was subjected to electrophoresis and allowed to move 1.5 cm into the gel, and DNA was isolated from the upper 1 cm of the lane. In the second round, the template selected in the first round (c-MycD6) was allowed to migrate about 3 cm into the gel, and two bound complexes were apparent. Both were excised together to yield the c-MycD6^2 template (Fig. 1C, lane 2).
16. T. K. Blackwell, L. Kretzner, E. M. Blackwood, R. N. Eisenman, H. Weintraub, unpublished data.
17. D3L1 (5′-primer B-ACCCCCC(G,A)CCAGTTG-(G,A,T)AGCCTGAT-primer A-3′) specifically binds to multiple factors that are present in a reticulocyte lysate and in various cell types (*11*, unpublished data). Primer A and primer B are given in (*23*).
18. S. R. Hann and R. N. Eisenman, *Mol. Cell. Biol.* **4**, 2486 (1984).
19. P. D. Gregor, M. Sawadogo, R. Roeder, *Genes Dev.* **4**, 730 (1990).
20. C. V. Dang, M. McGuire, M. Buckmire, W. M. F. Lee, *Nature* **337**, 664 (1989).
21. c-MycC92 was expressed from pMC92, in which the Ava II-Eco RI fragment of human c-Myc [S. R. Hann, M. W. King, D. L. Bentley, C. W. Anderson, R. N. Eisenman, *Cell* **52**, 185 (1988)] was subcloned into the pGex-2T [D. B. Smith and K. S. Johnson, *Gene* **67**, 31 (1988)] vector between the Sma I and Eco RI sites. Both c-MycC92 and GST proteins were prepared and purified as in (*5*).
22. B. E. Lüscher, E. A. Kuenzel, E. G. Krebs, R. N. Eisenman, *EMBO J.* **8**, 1111 (1989).
23. Nucleotide sequences of primer A and primer B are 5′-TCCGAATTCCTACAG-3′, and 5′-AGACGG-ATCCATTGCA-3′, respectively.
24. Template preparation, labeling, and EMSA were performed as in (*11*), with 225 ng of c-MycC92 in each binding reaction. Double-stranded D6 and D3L1 templates were synthesized from oligonucleotides with the use of primer A (*23*), and the Klenow fragment of *Escherischia coli* DNA polymerase. The MCK template was generated by the annealing of two complementary oligonucleotides (*5*). These templates were ^{32}P end-labeled with polynucleotide kinase (*7*), and the c-MycD6^2 and c-MycD6^3 templates were labeled as in (*15*).
25. The EMSA was performed essentially as in (*11*), but with 4% polyacrylamide gels. The CM1 template was derived by the annealing of the complementary oligonucleotides, 5′-GATCCCCCCACCACGTGG-TGCCTGA-3′ and 5′-GATCTCAGGCACCACG-TGGTGGGGG-3′, and was ^{32}P end-labeled with polynucleotide kinase (*7*). Each reaction contained BSA (500 ng) and either GST (1.5 μg) or c-MycC92 (700 ng).

26. We thank M. Groudine and D. Turner for critically reading this manuscript. Supported by a Burroughs Wellcome Fund Fellowship from the Life Sciences Research Foundation (T.K.B.); a fellowship from the American Cancer Society (L.K.); the Molecular Training Program in Cancer Research at the University of Washington (NIH grant T32 CA09437) (E.M.B.); and by NIH grant RO1 CA20525 (R.N.E.).

7 September 1990; accepted 10 October 1990