MOLECULAR AND CELLULAR BIOLOGY, Aug. 1990, p. 4384–4388
0270-7306/90/084384-05$02.00/0
Copyright © 1990, American Society for Microbiology
Vol. 10, No. 8

A Helix-Loop-Helix Protein Related to the Immunoglobulin E Box-Binding Proteins

CYNTHIA S. CARR AND PHILLIP A. SHARP*

Center for Cancer Research and Department of Biology, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139

Received 22 January 1990/Accepted 26 April 1990

A human cDNA encoding a novel protein in the helix-loop-helix family has been isolated by screening a bacteriophage expression library with a probe containing the binding site for major late transcription factor. The protein encoded by this cDNA, TFEB, probably recognizes E-box sequences in the heavy-chain immunoglobulin enhancer.

A family of sequence-specific DNA-binding factors with a helix-loop-helix structure has been identified (18). Several human genes of this family have been described previously, including E47 and E12, which bind to the immunoglobulin κ light-chain gene enhancer (18); MyoD (31) and myogenin (35), which are specific for muscle cell differentiation; TFE3, which recognizes sequences in the immunoglobulin heavy-chain gene enhancer (3); and the proto-oncogenes c-, N-, and L-*myc* (31). In *Drosophila melanogaster*, members of this family include *daughterless* (6, 9), *hairy* (22), *twist* (32), and the *achaete-scute* complex (1, 6).

We have isolated a partial cDNA segment that encodes another member of the helix-loop-helix family. A sequence in the major late promoter of adenovirus (5′GTAGGCCA CGTGACCGGG3′, base pairs −66 to −49), recognized by major late transcription factor (MLTF) (5, 8, 25), was used to probe a λgt11 expression library made with mRNA from a human B-cell line (BJAB; a kind gift of L. Staudt [30]) by the gene screen method (28). Of 6×10^5 bacteriophage screened, 2 recombinants, shown to be identical and designated λTFEB, demonstrated specific binding.

DNA-binding specificity of the λTFEB protein was tested by probing purified phage plaque replicas with three different DNAs (Fig. 1). The λTFEB protein bound strongly to the MLTF binding-site segment but bound only very weakly to the double point mutant segment and to the polylinker segment alone.

Extracts of lysogens generated from the phage recombinant (14) were analyzed in a Southwestern (DNA-protein) protocol with either the wild-type binding-site DNA probe or the double point mutant DNA probe (Fig. 2). Control extracts containing the λh3 fusion protein, which specifically binds to the H2TF-1/NF-kB-binding site (28), did not bind to the MLTF-binding-site probe. A protein of approximately 180 kilodaltons (kDa) from the λTFEB lysogen extracts bound the wild-type binding-site probe very strongly and the double point mutant probe at least 100-fold less strongly. Since the β-galactosidase portion of the fusion protein is approximately 120 kDa, the cDNA-encoded portion must be approximately 60 kDa.

The sequence-specific binding protein detected by Southwestern analysis could not be solubilized by treatment of λTFEB lysogen extracts with denaturing agents such as guanidine hydrochloride (6 M), deoxycholate (0.5%), octyl-glucoside (1.5%), or potassium thiocyanate (2 M). This problem has not been observed with other β-galactosidase fusion proteins which bind DNA in a sequence-specific manner.

Guanine methylations which interfere with the binding of the MLTF-binding-site probe to the λTFEB fusion protein were determined by using a partially methylated probe in a Southwestern analysis. Bound probe was eluted and cleaved at sites of modification. Comparison of this cleavage pattern with that of a control of input probe shows that the methylation interference pattern for binding to λTFEB recombinant protein is similar to that previously characterized for MLTF protein (7) (Fig. 3C). On the coding strand, modification of any of four guanines interfered with binding; two of these modifications also interfere with the binding of MLTF (7). On the noncoding strand, the same four guanine modifications previously identified as important for the binding of MLTF (7) interfered with binding to the fusion protein.

Two bands complementary to the cDNA were detected by a Southern analysis of X50-7 (human B-cell) DNA in each of three digestions with enzymes which did not cut within the cDNA (data not shown). Thus, the cDNA is encoded by one or a few cellular genes. A 3-kilobase mRNA was detected by a Northern (RNA blotting) protocol on polyadenylated

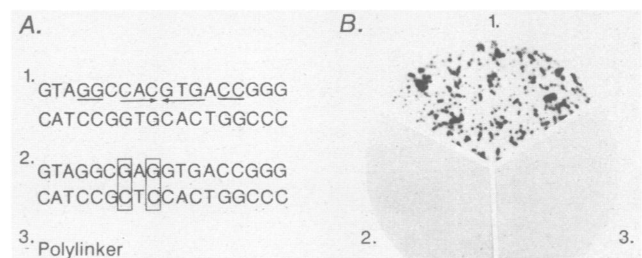

FIG. 1. Specificity of binding of λTFEB-encoded protein. Replica filters from platings of λTFEB recombinant phage were tested for specificity of DNA binding. The probes were cut from pUC13 plasmids containing inserts of either the wild-type binding site or the mutant binding site or no insert. (A) Probe 1, Wild-type MLTF binding site; probe 2, double point mutant, which binds the MLTF with an approximately 150-fold lower affinity; probe 3, pUC polylinker segment into which both the wild-type and the mutant sequences were inserted. (B) Filters which were lifted from plaques that produced the fusion protein were probed with the three different sequences, as described previously (28), with the addition of a renaturation step (33).

* Corresponding author.

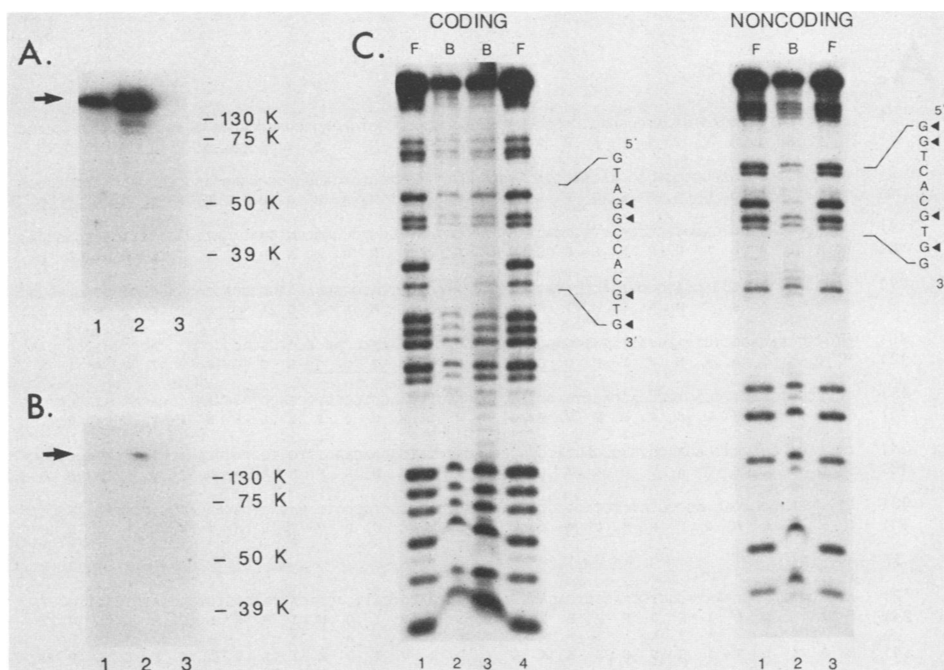

FIG. 2. Southwestern analysis of extracts of the λTFEB lysogen and the methylation interference pattern of the λTFEB fusion protein. The lysates of induced lysogens were resolved by electrophoresis in parallel, and the gels were transferred to nitrocellulose and subjected to Southwestern and Western (immunoblot) procedures. (A) Southwestern blot of wild-type MLTF probe. A Western blot of the same gel developed with antiserum to β-galactosidase (28) showed that the proteins which bound probe were β-galactosidase fusion proteins (data not shown). Lanes 1 and 2, Lysates prepared from two lysogens of the λTFEB phage; lane 3, control lysate prepared from a lysogen of a λh3 phage (28). Protein sizes (in thousands [K]) are indicated to the right of the gel. (B) Nitrocellulose blot treated as in panel A, except that the double point mutant was used in the Southwestern analysis. The arrow in panels A and B points to the 180K protein band identified in this analysis. (C) Probe was partially methylated with dimethylsulfoxide before binding to the Southwestern blot. Bound probe was eluted from the band with NaCl and treated with piperidine before analysis on a sequencing gel (16). Results for the coding and noncoding strands are as shown. Lanes F, Methylated and cleaved input probe which was not selected on the Southwestern blot; lanes B, methylated probe that was eluted from the band in the Southwestern blot and cleaved.

RNAs from several cell lines (data not shown). Compared with the amount of mRNA complementary to an actin cDNA probe, the TFEB mRNA was abundant in epithelial (HeLa) and B (X50-7 and BJAB) cell lines but was 10-fold less abundant in T (Jurkat) cells.

The sequencing (23) of the cDNA insert revealed one open reading frame specifying a polypeptide of 55 kDa which was not closed at either end, so the encoded protein must be greater than 55 kDa. Previous experiments strongly suggest that MLTF is approximately 43 kDa (5, 8, 24–26). Thus, the λTFEB cDNA probably specifies a different protein.

The sequence contains several interesting motifs (Fig. 3). The N-terminal amino acids have a high Gly-Ala content (approximately 50%), a characteristic typical of structural proteins such as keratins. There is a pronounced glutamine-rich (approximately 50%) region from amino acids (aa) 106 to 140 which contains 10 consecutive glutamines. At the carboxyl terminus, there is a proline-rich (approximately 40%) region. The helix-loop-helix homology (18) lies between aa 327 and 392.

Of the helix-loop-helix family of proteins, TFEB is most closely related to the TFE3 protein (3), which binds specifically to the E3 box of the μ-chain immunoglobulin enhancer. The sequence of the E3 box is similar to that of the MLTF binding site, and the distribution of methylation sites which interfere with the binding of TFE3 protein to the E3 box is similar to the distribution of sites which interfere with the binding of λTFEB protein to the MLTF site (Fig. 3C).

For this family of proteins, the sequences responsible for specific DNA recognition are contained within the basic region immediately amino terminal to the helix-loop-helix (11). TFEB and TFE3 amino acid sequences are identical at 29 positions in this region; therefore, it is not surprising that the proteins have similar binding specificities. The helix-loop-helix structure promotes bivalent association of polypeptides as homo- or heterodimers, probably through hydrophobic interactions of coiled-coil domains (19, 20). It is possible that the TFEB protein forms heterodimers with TFE3 or other helix-loop-helix proteins.

The homology between TFE3 and TFEB extends to other regions. Sequences from aa 240 to 320 of TFEB are homologous to sequences from aa 40 to 127 of TFE3. Since this region in TFE3 has been shown to possess transactivation activity, the related region in TFEB may activate transcription (3). Also, these two proteins are homologous through a region containing a potential leucine zipper immediately carboxyl terminal to the helix-loop-helix domain (Fig. 3B), a domain which is found in the c-*myc* protein at a similar position. The intact c-*myc* protein is thought to form a tetramer in solution; deletion of this leucine zipper region yields a dimeric structure for the partial c-*myc* protein (10). It is possible that the leucine zipper sequences of TFE3 and TFEB proteins will also be active in formation of multiple-protein complexes.

The immunoglobulin heavy-chain enhancer contains four related E-box sequences which are specifically bound by proteins present in B cells (13). Nuclear extracts of B cells and non-B cells contain distinguishable factors that bind

A.

```
    1  GAATTCCGGGGTGGAGTGACAGCCGGAGCCCGGGCGCCGGCGCGGGGCTCGGTGACAGCGGAGGCGGCGGCCCGGGCGGGACGAGGGAGC
    1  E  F  R  G  G  V  T  A  G  A  R  A  P  A  R  G  S  V  T  A  E  A  A  A  R  A  G  R  G  S

   91  GGCCGCGGCGCCGACGGCGGGCGGGAAGGGCGGACGGGGCGGGCCTCCCTGGTGGCACGGGGCCGGAGCAGGCCGGGAGCAGGTGGCGGG
   31  G  R  G  A  D  G  G  R  E  G  R  T  G  R  A  S  L  V  A  R  G  R  S  R  P  G  A  G  G  G

  181  CACAGCTTGCGGGCCAGGCACCCGAACTTGCGACAAGTTGCCGGAGCCGGGGCGCGGGCGGCGGACAGATTGACCTTCAGAGCGAGGGAG
   61  H  S  L  R  A  R  H  P  N  L  R  Q  V  A  G  A  G  A  R  A  A  D  R  L  T  F  R  A  R  E

  271  CCAGCGCCGGCAGCCACCATGGCGTCACGCATAGGGTTGCGCATGCAGCTCATGCGGGAGCAGGCGCAGCAGGAGGAGCAGCGGGAGCGC
   91  P  A  P  A  A  T  M  A  S  R  I  G  L  R  M  Q  L  M  R  E  Q  A  Q  Q  E  E  Q  R  E  R

  361  ATGCAGCAACAGGCTGTCATGCATTACATGCAGCAGCAGCAGCAGCAACAGCAGCAGCTCGGAGGGCCGCCCACCCCGGCCATCAAT
  121  M  Q  Q  Q  A  V  M  H  Y  M  Q  Q  Q  Q  Q  Q  Q  Q  Q  Q  L  G  G  P  P  T  P  A  I  N

  451  ACCCCCGTCCACTTCCAGTCGCCACCACCTGTGCCTGGGGAGGTGTTGAAGGTGCAGTCCTACCTGGAGAATCCCACATCCTACCATCTG
  151  T  P  V  H  F  Q  S  P  P  P  V  P  G  E  V  L  K  V  Q  S  Y  L  E  N  P  T  S  Y  H  L

  541  CAGCAGTCGCAGCATCAGAAGGTGCGGGAGTACCTGTCCGAGACCTATGGGAACAAGTTTGCTGCTGCCCACATCAGCCCAGCCCAGGCT
  181  Q  Q  S  Q  H  Q  K  V  R  E  Y  L  S  E  T  Y  G  N  K  F  A  A  A  H  I  S  P  A  Q  A

  631  CTCCGAAACCCCCACCAGCCGCCTCCCCAGGGGTGCGAGCTGGACACGTGCTGTCCTCCTCCGCTGGCAACAGTGCTCCCAATAAGCCCC
  211  L  R  N  P  H  Q  P  P  P  Q  G  C  E  L  D  T  C  C  P  P  P  L  A  T  V  L  P  I  S (P
                                                                                    *     *  *
   37                                                                               P  N  S  P

  721  ATGGCCATGCTGCACATTGGCTCCAACCCTGAGAGGGAGTTGGATGATGTCATTGACAACATTATGCGTCTGACGATGTCTCTTGGCTAC
  241  M  A  M  L  H  I  G  S  N  P  E  R  E  L  D  D  V  I  D  N  I  M  R  L  T  M  S  L  G  Y
       *  *     *     *  *  *        *     *     *  *  *  *  *     *        *        *  *
   41  M  A  L  L  T  I  G  S  S  S  E  K  E  I  D  D  V  I  D  E  I  I  S  L  E  S  S..L  P  G

  811  ATCAATCCTGAAATGCAGATGCCCAACACGCTACCCCTGTCCAGCAGCCACCTGAATGTGTACAGCAGCGACCCCCAGGTCACAGCCTCC
  271  I  N  P  E  M  Q  M  P  N  T  L  P  L  S  S  S  H  L  N  V  Y  S  S  D  P  Q  V  T  A  S
                      *     *     *  *  *     *           *     *  *  *  *           *
   79  G  T  T  G  L  Q  L  P  S  T  L  P  V  S  G  N  L  L  D  V  Y  S  S  -  Q  G  V  A  T  P

  901  CTGGTGGGCGTCACCAGCAGCTCCTGCCCTGCGGACCTGACCCAGAAGCGAGAGCTCACAGATGCTGAGAGCAGGGCCCTGGCCAAGGAG
  301  L  V  G  V  T  S  S  S  C  P  A  D  L  T  Q  K  R  E  L  T) D  A  E  S  R  A  L  A  K  E
                *        *                          *  *  *              *        *  *     *  *
  108  A  I  T  V  S  N  S  C  P  A  E  L  P  N  I  K  R  E  I  S  E  T  E  A  K  A  L  L  K  E

  991  CGGCAGAAGAAAGACAATCACAACTTAATTGAAAGGAGACGAAGGTTCAACATCAATGACCGCATCAAGGAGTTGGGAATGCTGATCCCC
  331  R  Q  K  K  D  N  H  N  L  I  E  R  R  R  R  F  N  I  N  D  R  I  K  E  L  G  M  L  I  P
       *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *     *  *  *
  138  R  Q  K  K  D  N  H  N  L  I  E  R  R  R  R  F  N  I  N  D  R  I  K  E  L  G  T  L  I  P

 1081  AAGGCCAATGACCTGGACGTGCGCTGGAACAAGGGCACCATCCTCAAGGCCTCTGTGGATTACATCCGGAGGATGCAGAAGGACCTGCAA
  361  K  A  N  D  L  D  V  R  W  N  K  G  T  I  L  K  A  S  V  D  Y  I  R  R  M  Q  K  D  L  Q
       *        *           *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *        *  *        *
  168  K  S  S  D  P  E  M  R  W  N  K  G  T  I  L  K  A  S  V  D  Y  I  R  K  L  Q  K  E  Q  Q

 1171  AAGTCCAGGGAGCTGGAGAACCACTCTCGCCGCCTGGAGATGACCAACAAGCAGCTCTGGCTCCGTATCCAGGAGCTGGAGATGCAGGCT
  391  K  S  R  E  L  E  N  H  S  R  R  L  E  M  T  N  K  Q  L  W  L  R  I  Q  E  L  E  M  Q  A
          *        *  *           *     *           *        *     *  *  *  *  *  *  *     *  *
  198  R  S  K  D  L  E  S  R  Q  R  S  L  Q  Q  A  N  R  S  L  Q  L  R  I  Q  E  L  E  L  Q  A

 1261  CGAGTGCACGGCCTCCCTACCACCTCCCCGTCCGGCATGAACATGGCTGAGCTGGCCCAGCAGGTGGTGAAGCAGGAGCTGCCTAGCGAA
  421  R  V  H  G  L  P  T  T  S  P  S  G  M  N  M  A  E  L  A  Q  Q  V  V  K  Q  E  L  P  S  E
             *  *  *  *
  228  Q  I  H  G  L  P

 1351  GAGGGCCCAGGGGAGGCCCTGATGCTGGGGGCTGAGGTCCCTGACCCTGAGCCACTGCCAGCTCTGCCCCCGCAAGCCCCGCTGCCCCTG
  451  E  G  P  G  E  A  L  M  L  G  A  E  V  P  D  P  E  P  L  P  A  L  P  P  Q  A  P  L  P  L

 1441  CCCACCCAGCCACCGTCCCCATTCCATCACCTGGACTTCAGCCACAGCCTGAGCTTTGGGGGCAGGGAGGACGAGGGTCCCCCGGGCTAC
  481  P  T  Q  P  P  S  P  F  H  H  L  D  F  S  H  S  L  S  F  G  G  R  E  D  E  G  P  P  G  Y

 1531  CCCGACCGGAATTC
  511  P  D  R  N
```

FIG. 3. (A) Sequence of the cDNA fragment (GenBank accession number M33782) in λTFEB and its predicted amino acid sequence. The sequence from aa 37 to 233 of the highly homologous λ3 clone protein (encoding TFE3 [3]) is shown below the TFEB sequence, starting at aa 238. Identities are noted by an asterisk. Several features of the TFEB predicted aa sequence are underlined, including the Gly and Ala residues in the high-Gly-Ala region (aa 1 to 60) and the glutamine residues in the glutamine-rich region (aa 103 to 140). The region homologous to the activator region in the λ3-encoded protein (3) is enclosed in parentheses, and the region homologous to the helix-loop-helix motif (or *myc* similarity region) is boxed. The leucines of a leucine zipper motif directly 3′ to the *myc* similarity region and the prolines in the proline-rich region (aa 463 to 515) at the carboxyl end of the clone are also underlined. (B) Comparison of the helix-loop-helix regions of several different proteins. Identities between adjacent sequences are noted by vertical lines. Sequences in the potential leucine zipper are also shown for some proteins. The homologous proteins include L-*myc* (aa 284 to 364 [12]), N-*myc* (aa 388 to 463 [12, 29]), c-*myc* (aa 341 to 429 [2]), TFEB (aa 326 to 425), TFE3 (aa 134 to 233 [3]), *da* (*daughterless*; aa 549 to 645 [6]), E12 (aa 331 to 411 [18]), E47 (aa 331 to 395 [18]), *lyl-1* (aa 132 to 196 [17]), *twist* (aa 352 to 409 [32]), MyoD1 (aa 104 to 167 [31]), and myogenin (aa 76 to 137 [35]). (C) Comparison of the methylation interference patterns of MLTF, TFEB, and TFE3. Pattern 1, MLTF binding to the site in the adenovirus major late promoter, as determined previously by Chodosh et al. (7); pattern 2, TFEB binding to the MLTF site (Fig. 2); pattern 3, TFE3 binding to the μE3 box, as determined previously by Beckmann et al. (3).

B.

Helix Loop Helix Motif — *Leucine Zipper Motif*

Basic Region — Helix 1 — *Loop* — Helix 2

```
L-myc     DTEDVTKRKN HNFLERKRRN DLRSRFLALR DQVPTLASCS KA-PKVVILS RALEYLQALV GAEKRM --A-TELRNL RCRQQQLQKR IAYLSGY
N-myc     DSEDSERRRN HNILERQRRN DLRSSFLTLR DHVPELVKNE KA-AKVVILK KATEYVHSLQ AEEHQ- --LLLEKEKL QARQQQLLKK IEHARTC
c-myc     NTEENVKRRT HNVLERQRRN ELKRSFFALR DQIPELENNE KA-PKVVILK KATAYILSVQ AEEQK- --LISEEDLL RKRREQLKHK LEQLRNSCA
TFEB      LAKERQKKDN HNLIERRRRF NINDRIKELG MLIPKANDLD VRWNKGTILK ASVDYIRRMQ KDLQKS RELENHSRRL EMTNKQLWLR IQELEMQARV HGLP
TFE3      LLKERQKKDN HNLIERRRRF NINDRIKELG TLIPLSSDPQ MRWNLGTILK ASVDYIRKLQ KEQQRS KDLESRQRSL QQANRSLQLR IQELELQAQI HGLP
da        AIREKERRQA NNARERIRIR DINEALKELG RMCMTHLKSD KPQTKLGILN MAVEVIMTLE QQVRER N-LNPKAACL KRREEEKAED GPKLSAQHHM I
E12       AEREKERRVA NNARERLRVR DINEAFKELG RMCQLHLNSE KPQTKLLILH QAVSVILNLE QQVRER N-LNPKAACL KRREEEK
E47       DLRDRERRMA NNARERVRVR DINEAFRELG RMCQMHLKSD KAQTKLLILQ QAVQVILGLE QQVRER
lyl-1     QPQKVARRVF TNSRERWRQQ NVNGAFAELR KLLPTHPPDR KL-SKNEVLR LAMKYIGFLV RLLRDQ
twist     TDEFSNQRVM ANVRERQRTQ SLNDAFKSLQ QIIPTL-PSD KL-SKIQTLK LATRYIDFLC RMLSSS
MyoD      KTTNADRRLA ATMRERRRLS KVNEAFETLK RCT-SSNPNQ RL-PKVEILR NAIRYIEGLQ ALLRDQ
Myogenin  KSVSVDRRRA ATLREKRRLK KVNEAFEALK RST-LLNPNQ RL-PKVEILR HAIQYIERLQ ALLSSL
```

C.

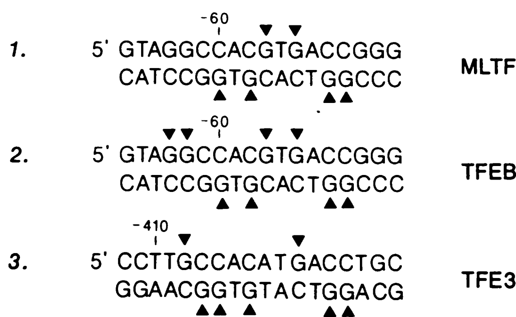

FIG. 3—*Continued.*

each of the boxes with higher affinity than the other three boxes (4, 15, 21, 27, 34), suggesting that there is a family of factors with related binding specificities. The relationships between TFE3 and TFEB proteins and between their binding-site specificities indicate that they are members of this family.

We thank Sharon Jamison for extensive technical assistance during the sequence analysis; Ken LeClair for supplying a Northern blot and for technical advice; Rich Carthew, Lewis Chodosh, and Al Baldwin for inspiration and advice; Ken LeClair, Myles Brown, and Tom Kristie for comments on the manuscript; and Margarita Siafaca for help in typing and preparing the manuscript.

C.S.C. was supported by Public Health Service training grant T32-GM07287 from the National Institutes of Health to the Department of Biology. This work was supported by Public Health Service grant P0-CA14051 from the National Institutes of Health and partially by Cancer Center Support (core) grant P30-CA14051 from the National Cancer Institute and Cooperative Agreement CDR-8803014 from the National Science Foundation to P.A.S.

LITERATURE CITED

1. **Alonso, M. C., and C. V. Cabrera.** 1988. The *achaete-scute* gene complex of *Drosophila melanogaster* comprises four homologous genes. EMBO J. **7**:2585–2591.
2. **Battey, J., C. Moulding, R. Taub, W. Murphy, T. Stewart, H. Potter, G. Lenoir, and P. Leder.** 1983. The human c-*myc* oncogene: structural consequences of translation into the IgH locus in Burkitt lymphomas. Cell **34**:779–787.
3. **Beckmann, H., L.-K. Su, and T. Kadesh.** 1990. TFE3: a helix-loop-helix protein that activates transcription through the immunoglobulin enhancer μE3 motif. Genes Dev. **4**:167–179.
4. **Calame, K., and S. Eaton.** 1988. Transcriptional controlling elements in the immunoglobulin and T cell receptor loci. Adv. Immunol. **43**:235–275.
5. **Carthew, R. W., L. A. Chodosh, and P. A. Sharp.** 1985. An RNA polymerase II transcription factor binds to an upstream element in the adenovirus major late promoter. Cell **43**:439–448.
6. **Caudy, M., H. Vassin, M. Brand, R. Tuma, L. Y. Jan, and Y. N. Jan.** 1988. *daughterless*, a Drosophila gene essential for both neurogenesis and sex determination, has sequence similarities to *myc* and the *achaete-scute* complex. Cell **55**:1061–1067.
7. **Chodosh, L. A., R. W. Carthew, J. G. Morgan, G. R. Crabtree, and P. A. Sharp.** 1987. The adenovirus major late transcription factor activates the rat gamma-fibrinogen promoter. Science **238**:684–688.
8. **Chodosh, L. A., R. W. Carthew, and P. A. Sharp.** 1986. A single polypeptide possesses the binding and transcription activities of the adenovirus major late transcription factor. Mol. Cell. Biol. **6**:4723–4733.
9. **Cline, T. W.** 1989. The affairs of *daughterless* and the promiscuity of developmental regulators. Cell **59**:231–234.
10. **Dang, C. V., M. McGuire, M. Buckmire, and W. M. F. Lee.** 1989. Involvement of the 'leucine zipper' region in the oligomerization and transforming activity of human c-myc protein. Nature (London) **337**:664–666.
11. **Davis, R. L., P.-F. Cheng, A. B. Lassar, and H. Weintraub.** 1990. The MyoD DNA binding domain contains a recognition code for muscle-specific gene activation. Cell **60**:733–746.
12. **DePinho, R. A., K. S. Hatton, A. Tesfaye, G. D. Yancopoulos, and F. W. Alt.** 1987. The human *myc* gene family: structure and activity of L-*myc* and an L-*myc* pseudogene. Genes Dev. **1**:1311–1326.
13. **Ephrussi, A., G. M. Church, S. Tonegawa, and W. Gilbert.** 1985. B lineage-specific interactions of an immunoglobulin enhancer with cellular factors in vivo. Science **227**:134–140.
14. **Huynh, T. V., R. A. Young, and R. W. Davis.** 1985. Constructing

and screening cDNA libraries in λgt10 and λgt11, p. 49–78. *In* D. Glover (ed.), DNA cloning: a practical approach. IRL Press, Oxford.
15. **Lenardo, M., J. W. Pierce, and D. Baltimore.** 1987. Protein-binding sites in Ig gene enhancers determine transcriptional activity and inducibility. Science **236:**1573–1577.
16. **Maxam, A., and W. Gilbert.** 1980. Sequencing end-labeled DNA with base-specific chemical cleavages. Methods Enzymol. **65:** 499–560.
17. **Mellentin, J. D., S. D. Smith, and M. L. Cleary.** 1989. *lyl-1*, a novel gene altered by chromosomal translocation in T cell leukemia, codes for a protein with a helix-loop-helix DNA binding motif. Cell **58:**77–83.
18. **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.
19. **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.
20. **O'Shea, E. K., R. Rutkowski, and P. S. Kim.** 1989. Evidence that the leucine zipper is a coiled coil. Science **243:**538–542.
21. **Peterson, C. L., and K. Calame.** 1989. Proteins binding to site C2 (μE3) in the immunoglobulin heavy-chain enhancer exist in multiple oligomeric forms. Mol. Cell. Biol. **9:**776–786.
22. **Rushlow, C. A., A. Hogan, S. M. Pinchin, K. M. Howe, M. Lardelli, and D. Ish-Horowicz.** 1989. The *drosophila hairy* proteins acts in both segmentation and bristle patterning and shows homology to N-myc. EMBO J. **8:**3095–3103.
23. **Sanger, F., S. Nicklen, and A. R. Coulson.** 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA **74:**5463–5467.
24. **Sawadogo, M.** 1988. Multiple forms of the human gene-specific transcription factor USF. II. DNA binding properties and transcriptional activity of the purified HeLa USF. J. Biol. Chem. **263:**11994–12001.
25. **Sawadogo, M., and R. G. Roeder.** 1985. Interaction of a gene-specific transcription factor with the adenovirus major late promoter upstream of the TATA box region. Cell **43:**165–175.
26. **Sawadogo, M., D. M. W. Van, P. D. Gregor, and R. G. Roeder.** 1988. Multiple forms of the human gene-specific transcription factor USF. I. Complete purification and identification of USF from HeLa cell nuclei. J. Biol. Chem. **263:**11985–11993.
27. **Sen, R., and D. Baltimore.** 1986. Multiple nuclear factors interact with the immunoglobulin enhancer sequences. Cell **46:**705–716.
28. **Singh, H., J. H. Lebowitz, A. S. Baldwin, Jr., and P. A. Sharp.** 1988. Molecular cloning of an enhancer binding protein: isolation by screening of an expression library with a recognition site DNA. Cell **52:**415–423.
29. **Stanton, L. W., M. Schwab, and J. M. Bishop.** 1986. Nucleotide sequence of the human N-*myc* gene. Proc. Natl. Acad. Sci. USA **83:**1772–1776.
30. **Staudt, L. M., R. G. Clerc, H. Singh, J. H. Lebowitz, P. A. Sharp, and D. Baltimore.** 1988. Cloning of a cDNA encoding a B-cell restricted octamer binding factor. Science **241:**577–580.
31. **Tapscott, S. J., R. L. Davis, M. J. Thayer, P.-F. Cheng, H. Weintraub, and A. B. Lassar.** 1988. MyoD1: a nuclear phosphoprotein requiring a myc homology region to convert fibroblasts to myoblasts. Science **242:**405–411.
32. **Thisse, B., C. Stoetzel, C. Gorostiza-Thisse, and F. Perrin-Schmitt.** 1988. Sequence of the *twist* gene and nuclear localization of its protein in endomesodermal cells of early *Drosophila* embryos. EMBO J. **7:**2175–2183.
33. **Vinson, C. R., K. L. LaMarco, P. F. Johnson, W. H. Landschulz, and S. L. McKnight.** 1988. In situ detection of sequence-specific DNA binding activity specified by a recombinant bacteriophage. Genes Dev. **2:**801–806.
34. **Weinberger, J., D. Baltimore, and P. A. Sharp.** 1986. Distinct factors bind to the homologous sequence in the immunoglobulin heavy chain enhancer. Nature (London) **322:**846–848.
35. **Wright, W. E., D. A. Sasson, and V. K. Lin.** 1989. Myogenin, a factor regulating myogenesis, has a domain homologous to MyoD. Cell **56:**607–617.