MOLECULAR AND CELLULAR BIOLOGY, Nov. 1991, p. 5603–5611
0270-7306/91/115603-09$02.00/0

Vol. 11, No. 11

Id Proteins Id1 and Id2 Selectively Inhibit DNA Binding by One Class of Helix-Loop-Helix Proteins

XIAO-HONG SUN,[1]† NEAL G. COPELAND,[2] NANCY A. JENKINS,[2] AND DAVID BALTIMORE[1,3]*

Whitehead Institute for Biomedical Research, 9 Cambridge Center, Cambridge, Massachusetts 02142[1]; Rockfeller University, 1230 York Avenue, New York, New York 10021[3]; and Mammalian Genetics Laboratory, ABL-Basic Research Program, NCI-Frederick Cancer Research and Development Center, Frederick, Maryland 21702[2]*

Received 11 March 1991/Accepted 16 August 1991

The DNA binding activities of some basic region and putative helix-loop-helix (bHLH)-containing transcriptional factors can be inhibited by the Id protein. Because Id contains the HLH motif for dimerization but not the basic amino acid region for DNA binding, heterodimers of Id with bHLH transcriptional factors may not bind to DNA. We have isolated and characterized the gene and cDNA clones for a new Id protein, designated Id2. The Id2 protein contains a helix-loop-helix motif similar to that of the previously described Id protein (referred to here as Id1), but the two proteins are different elsewhere. Id1 and Id2 are encoded by two unlinked genes, as shown by chromosome mapping. The two Id proteins have similar inhibitory activities. They selectively bind to and inhibit the function of one set of bHLH proteins, typified by E2A.E47 and E2B.m3, but not that of the other set, including TFE3, USF, and AP4. The Id proteins also homodimerize poorly. Expression of both *Id* genes is down-regulated during differentiation in a variety of cell types.

A large family of proteins, mostly transcriptional activators, share a common DNA binding motif, thought to be composed of a basic amino acid region and a helix-loop-helix region (bHLH) (29). The bHLH proteins form homo- and heterodimers through the helix-loop-helix regions and bind to specific DNA sequences with the basic region (14, 45). These proteins have been shown to play important and specific roles in the differentiation of specific cell types. For example, the bHLH proteins MyoD, myogenin, MRF4, and Myf5 (8, 15, 16, 35, 46), control aspects of gene expression in differentiating muscle cells, while E12 and E47, two differentially spliced products of the *E2A* gene (30, 42), have multiple transcriptional activating functions in differentiating B lymphocytes, muscle cells, pancreatic β-cells, nerve cells, and perhaps others (29). It is thus important to understand specific regulation of the functions of these proteins. Some regulation is achieved by cell-type- and stage-specific expression of the bHLH proteins, as in the muscle-specific expression of MyoD (15). In other cases the function of expressed bHLH proteins is controlled by their intrinsic inability to bind to DNA as in E12 homodimers (42). Additional mechanisms must exist to regulate the activities of certain constitutively expressed and yet active DNA binding proteins such as the E47 homodimer.

The discovery of the Id protein provided a new perspective for considering the regulation of bHLH proteins (6). Id protein contains a helix-loop-helix region similar to that of the bHLH proteins and can form heterodimers with some of them. But these heterodimers cannot bind to DNA, because the Id protein lacks the apparent basic region responsible for DNA binding; Id therefore negatively regulates the DNA binding capacity of bHLH proteins. Benezra et al. (6) have shown that the Id protein not only inhibits the DNA binding activities of E12, E47, and MyoD homo- and heterodimers in vitro, but also inhibits the transcriptional activation by MyoD when Id is expressed in cells with the MyoD gene and a reporter gene, demonstrating that Id is an effective negative regulator acting directly on the bHLH proteins. Id also has a *Drosophila* homolog, extramachrochaetae (*emc*), whose mutant phenotype suggests that *emc* inhibits the functions of the daughterless and achaete-scute proteins (17, 19), two bHLH proteins involved in sex determination and neurogenesis in *Drosophila melanogaster* (2, 12, 13, 44).

To more fully understand the role of Id in the regulation of bHLH proteins in various cellular processes, we have addressed the following questions: Is there more than one Id protein? Many bHLH proteins exist; are all inhibited by the Id proteins? Finally, are the genes for the Id proteins regulated? For example, in what cells and at what developmental stages is an Id expressed? To begin to define these issues, we report here a genomic clone and its cDNA clones which encode a new Id protein, designated Id2. Id2 is very similar in the helix-loop-helix motif to the first Id protein, referred to as Id1, but the two Id proteins are very different elsewhere. We then investigated the inhibitory specificities of Id1 and Id2 on bHLH proteins and found that while both Id1 and Id2 inhibit DNA binding by certain bHLH proteins like E2A.E47 and another close relative, E2B.m3, neither of them inhibits the binding by other bHLH proteins such as c-Myc, TFE3, USF, and AP4. We also directly examined heterodimer formation of the two Id proteins with various bHLH proteins as well as homodimer formation between them. The expression of the two *Id* genes is down-regulated in parallel during differentiation of many different cell types, which is consistent with a role for both as negative regulators that inhibit the function of bHLH proteins until cells have initiated a program of terminal differentiation.

MATERIALS AND METHODS

Cell lines, RNA isolation, and analysis. The mouse B lymphoid cell lines HAFTL (1), PD31, and WEHI-231 were maintained in RPMI medium containing 10% fetal calf se-

* Corresponding author.
† Present address: Department of Cell Biology, New York University Medical Center, New York, NY 10016.

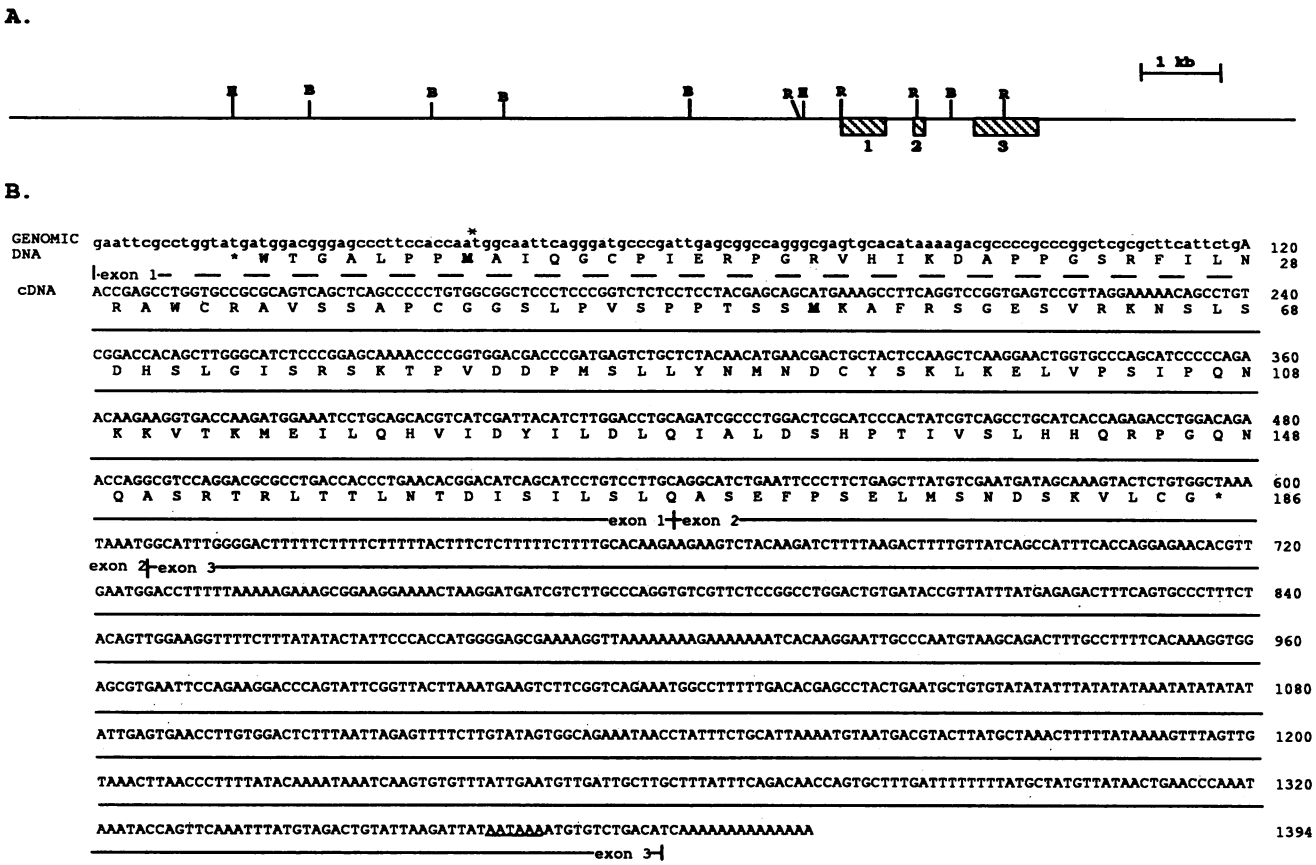

FIG. 1. Structure of the *Id-2* gene and its cDNA. (A) Restriction map of *Id-2* locus. The thin line represents the locus, and exons are shown in hatched boxes and numbered 1, 2, and 3. H, *Hin*dIII; R, *Eco*RI; B, *Bam*HI. (B) DNA sequence of a chimeric Id2 clone. The sequence of an *Eco*RI genomic fragment overlaps with the sequence of a partial cDNA between nucleotide 120 and 542. The first 119 bp of the sequence was from the genomic fragment shown in lowercase characters and speculated to be part of the exon 1 (underlined with a dashed line). Each exon is bracketed with horizontal and vertical lines and labeled at both ends of the exon. The polyadenylation signal is underlined. The two potential initiation methionine residues are shown in bold, and the first ATG is marked with an asterisk.

rum. LyD9 cells (33) were grown in Iscove modified Dulbecco's medium supplemented with 5% fetal calf serum and 10% WEHI-3 cell supernatant as a source of interleukin-3, 2 μM glutamine, and 50 μM β-mercaptoethanol. Murine erythroleukemia (MEL) cells (cell line 745-PC3-A21) (34) were grown to the density of 2×10^5 cells per ml in Dulbecco modified Eagle medium with 10% fetal calf serum and induced to differentiate for 48 h at 37°C with 1.5% dimethyl sulfoxide. The undifferentiated MEL cell sample was grown the same way without dimethyl sulfoxide treatment. The mouse F9 embryonal carcinoma cell line was grown in Dulbecco modified Eagle medium with 10% calf serum and treated for 120 h with 1 μM retinoic acid, 250 μM cyclic AMP, and 250 μM theophylline to differentiate the cells into parietal endoderm-like cells (41). The F9 stem cells were treated with the vehicles of the drugs, which were ethanol and water.

RNA was isolated from these cells by the guanidinium-$CsCl_2$ method as previously described (4) and analyzed by Northern (RNA) blot analysis (4). About 5 μg of each RNA was analyzed by electrophoresis through a 1% agarose gel containing ethidium bromide, and the gel was photographed with type 55 Polaroid film, which also gave a negative film. The intensities of the rRNA bands were quantitated by scanning the negative film with an ultrascan laser densitometer 2202 (LKB; Pharmacia), and the values were used to normalize the concentrations of all RNA samples against that of the LyD9 RNA. Fifteen micrograms of each RNA was then analyzed by electrophoresis through a 1% formaldehyde gel as previously described (4). The filter was hybridized with either a 900-bp fragment containing Id1 cDNA or a 900-bp fragment of *Id-2* genomic DNA containing 540 bp of exon 1 (Fig. 1B) at high stringency (50% formamide–5× SSC [1× SSC is 0.15 M NaCl plus 0.015 M sodium citrate] at 42°C).

Cloning and analysis of *Id-2* genomic and cDNA clones. A mouse genomic DNA library was constructed by cloning *Sau*3AI partial digestion fragments into the λ phage EMBL4 vector. The library was screened with the 900-bp Id1 cDNA probe at high stringency (50% formamide–5× SSC at 42°C). A mouse thymus cDNA library cloned in the pcDNA II plasmid vector (Invitrogen) was screened with a 900-bp *Eco*RI genomic fragment of the *Id-2* gene at high stringency. DNA sequencing was performed by using the dideoxynucleotide chain termination method (37) with single-stranded plasmids as templates (4) prepared from the pcDNA II plasmids for cDNA clones and from the Bluescript plasmids (Stratagene) for genomic DNA subclones. The restriction map of the *Id-2* gene λ clone was obtained by Southern analysis (4) with the Id1 cDNA and the Id2 cDNA as probes.

A chimeric Id2 clone was constructed by replacing a 400-bp *Eco*RI-*Cla*I genomic DNA fragment containing exon 1 with a 280-bp *Xho*I-*Cla*I cDNA fragment in the cId2-k cDNA clone (Fig. 1B).

Chromosome mapping by interspecific backcross. Interspecific backcross progeny were generated by mating (C57BL/6J × *Mus spretus*)F_1 females and C57BL/6J males as previously described (11). A total of 205 N_2 progeny were obtained; a random subset of these N_2 mice were used to map the *Id-2* locus. DNA isolation and Southern blot analysis were performed essentially as previously described (25). All blots were prepared with Zetabind nylon membrane (AMF-Cuno). The Id2 cDNA probe (nucleotides 342 to 1394 [Fig. 1B]) was labeled with [α-^{32}P]dCTP by using a nick translation kit (Boehringer Mannheim). Washing was done to a final stringency of 0.2× SSC–0.1% sodium dodecyl sulfate (SDS) at 65°C. Fragments of 7.6 and 5.8 kb were detected in *Xba*I-digested C57BL/6J DNA; fragments of 9.0 and 5.8 kb were detected in *Xba*I-digested *M. spretus* DNA. A description of the probes and restriction fragment length polymorphisms for the proopiomelanocortin-1 (*Pomc-1*), chromosome 12 DNA segment Nyu2 (*D12Nyu2*), the N-myc proto-oncogene (*Nmyc*), and laminin B1-1 (*Lamb1-1*) have been reported previously (26, 32, 39). Recombination distances were calculated as previously described (20) by using the computer program SPRETUS MADNESS developed by D. Dave (Data Management Services, Inc., Frederick, Md.) and A. M. Buchberg (ABL-Basic Research Program, National Cancer Institute, Frederick, Md.). Gene order was determined by minimizing the number of recombination events required to explain the allele distribution patterns.

In vitro transcription and translation. To generate Id1 and Id2 proteins by in vitro translation, the DNA templates were first prepared by polymerase chain reaction (PCR) (36). The primers used for PCR were as follows:

Id1 sense, GCAATTAACCCTCACTAAAGGGTACCGTACAACCTTTCTCCAACTTC

Id1 antisense, GGCTGGAGTCCATCTGGTCCCTCAGTGC

Id2 sense, GCAATTAACCCTCACTAAAGGGTACCATGGCAATTCAGGGATGC

Id2 antisense, GGCGGATCCTTATTTAGCCACAGAGTAC

The sense primers of Id1 and Id2 both contain sequences for the T3 promoter (nucleotide 1 to 19) and transcription initiation site (nucleotide 20). The additional sequence in the Id1 primer includes Id1 5′ untranslated sequence (nucleotides 27 to 47 of the primer), and that in the Id2 primer contains the consensus sequence for translation initiation (nucleotides 24 to 29 of the primer) and the Id2 coding sequence beginning at the second codon (Fig. 1B). The antisense primers of Id1 and Id2 hybridize to the sequences near the translation termination codon. Each pair of primers was used in a PCR with an Id1 cDNA clone (6) or the chimeric Id2 clone as a template (Fig. 1B). The PCR products were extracted with phenol-chloroform and precipitated with ethanol. An aliquot of the products (1/10 to 1/5) was used as a template for the in vitro transcription reaction by using T3 polymerase as described by the vendor (Boehringer Mannheim). The RNAs were extracted with phenol-chloroform and precipitated with ethanol. Ten percent of each product was then used for the in vitro translation reaction with rabbit reticulocyte lysates and [^{35}S]methionine as described by the vendor (Promega). One microliter of each 50-μl translation reaction mix was analyzed by electrophoresis through a SDS-polyacrylamide gel and autoradiographed to estimate the molar amounts of protein produced by comparing the intensities of the protein bands corrected by the number of methionines present in each protein.

To obtain the DNA templates for TFE3, USF, and AP4, a pair of primers was designed for each protein in the same way as those for Id2 protein, i.e., the sense primer contains the T3 promoter, transcription start site, translation signal, and specific coding sequence at the indicated position, and the antisense primer has the sequence at the indicated positions (see Fig. 5A). These primers were then used in PCR with a pool of HeLa cell cDNAs as a template. The PCR products were purified from a low-melting-point agarose gel and used as templates for an additional round of PCR. The products were then used for the in vitro transcription and translation reactions as described above.

To produce full-length MyoD and truncated forms of E47, m3 and c-*myc*, linearized plasmid DNAs, pMyoD (pV2C11b) (15), pE47S (30), pM3, and pMyc were used as templates to produce RNAs by in vitro transcription reactions. The plasmids, pM3 and pMyc, were constructed by cloning PCR products, containing the indicated m3 and c-*myc* sequences (see Fig. 4A and 5A) into the *Nco*I and *Hin*dIII sites of a Bluescript-based translation vector (Stratagene; R. Andino, Whitehead Institute for Biomedical Research). The PCR products were obtained by using primers which contain the sense and antisense coding sequences with an *Nco*I site and an *Hin*dIII site at the ends. The RNA products were translated as described above.

DNA binding inhibition assays. In vitro-translated Id proteins were first mixed with various DNA binding bHLH proteins at the indicated molar ratio in a final volume of 10 μl equalized with a mock-translated lysate and incubated at 37°C for 20 min. Ten microliters of DNA binding mixes containing 20 mM Tris (pH 7.5), 100 mM NaCl, 2 mM dithiothreitol, 2 mM EDTA, 10% glycerol, 1 μg of poly(dI-dC), and ^{32}P-labeled probes (10^4 cpm) containing the appropriate binding sites was then added to the proteins, and the mixtures were incubated for 15 min at room temperature. The reactions were analyzed by electrophoresis through a 4% polyacrylamide gel in 0.5× TBE buffer (0.045 mM Tris, 0.045 mM boric acid, 0.05 mM EDTA).

Interaction of Ids with bHLH proteins. To generate Id fusion proteins with glutathione S-transferase (GST), a PCR-amplified fragment containing full-length Id1 or Id2 with artificial *Bam*HI sites on both ends was cloned into the *Bam*HI site in the expression vecter pGex-2T (40). The GST fusion proteins with C-terminal portions of E2A.E12 (amino acid 493 to 654) and E2A.E47 (amino acid 525 to 609) (27) were constructed by cloning, into the *Sma*I and *Eco*RI sites of pGex-2T, a blunt *Not*I-*Eco*RI fragment of pE12 (30) and a blunt *Nco*I-*Eco*RI fragment of pE47S (30). These plasmids were transformed into NB42, an *Escherichia coli* strain with the *lon* mutation genotype, for protein production.

Log-phase *E. coli* cultures were induced with 0.4 mM isopropyl-β-D-thiogalactopyranoside (IPTG) for 1 h, and crude extracts were prepared as previously described (42). To purify the proteins, 1 ml of 50% slurry of glutathione-agarose (Sigma) was added to each 15-ml extract and mixed at room temperature for 30 min. The agarose beads were washed twice with phosphate-buffered saline (PBS) containing 1% Nonidet P-40 (NP-40). The amount of protein bound per volume of beads was equalized by adding unbound glutathione-agarose.

The interaction of Id proteins with different bHLH proteins was examined by mixing the agarose beads which were bound to different fusion proteins with ^{35}S-labeled in vitro-translated proteins at 37°C for 20 min and then by washing

four times with PBS containing 0.1% NP-40. Radiolabeled proteins brought down by the beads were analyzed by electrophoresis through SDS-polyacrylamide gels. The gels were then treated with Biofluor and fluorographed.

RESULTS AND DISCUSSION

Isolation and characterization of the *Id-2* gene and its cDNA clones. In the process of isolating the *Id-1* gene from a mouse genomic library, we also obtained another λ phage clone, λId2-1, which hybridized weakly to the Id1 cDNA probe. By further restriction mapping and DNA sequencing analyses, we localized the homologous sequence to a 900-bp *Eco*RI fragment (Fig. 1A). This fragment contains an open reading frame of 222 amino acids starting at an AUG codon (see asterisk in Fig. 1B), part of which encodes the helix-loop-helix motif homologous to that in Id1. The reading frame extends only a few amino acids upstream from the AUG codon, suggesting that this AUG codon could be a translation initiation codon. To determine whether this reading frame is transcribed, we screened a mouse thymus cDNA library by using this *Eco*RI fragment as a probe and obtained several cDNA clones. The largest clone, cId2-k, was sequenced. Although the cDNA encodes an AUG that could serve as an initiation codon, the reading frame remains open upstream from the AUG to the 5′ end, suggesting that the cDNA might not be complete at the 5′ end because of a potential premature termination of reverse transcription at a GC-rich stretch (nucleotide 89 to 110) (Fig. 1B). Comparing the cDNA sequence with the genomic DNA sequence, we found that the 5′ portion of the partial cDNA is identical to the stretch of the genomic sequence from nucleotide 120 to 542 and contains the same open reading frame (Fig. 1B). Searching the genomic sequence from nucleotide 119 upstream to nucleotide 38, where the possible AUG initiation codon is located, we could not find any sequence resembling a putative 3′ splice acceptor site, suggesting that this genomic segment is part of exon 1. We thus predict that the full-length cDNA would contain this segment and that the reading frame of *Id-2* begins at the putative initiating AUG in the genomic sequence. However, this remains to be confirmed by determining the molecular weight of the natural Id2 protein once an anti-Id2 antibody is available. By comparing the genomic and cDNA sequences, it is evident that the 3′ portion of the cDNA is composed of two exons, exon 2 and exon 3 (Fig. 1B), and they were mapped as shown in Fig. 1A.

Compared with Id1, Id2 has a very similar helix-loop-helix motif (79% homology, and most of the changes are conservative) (Fig. 2). Both Id proteins lack a conserved alanine residue in helix II (Fig. 2), which might influence their dimerization properties. Id2 also does not contain the conserved basic amino acid region thought to be responsible for DNA binding in many bHLH proteins. It is thus very likely that Id2 can interact with the bHLH proteins and inhibit DNA binding just as Id1 does. However, whether the minor amino acid differences might give the two proteins different affinities for different bHLH proteins remains to be determined. Outside of the helix-loop-helix motif, the two proteins are very different except for two short stretches of weak homology (Fig. 2). It is interesting to note that these two homology stretches are rich in serine and threonine, which may be of regulatory importance for mechanisms like phosphorylation. Several known phosphorylation sites such as those for cyclic AMP-dependent kinase, protein kinase C, and casein kinase 2, were also found in Id2.

```
Id1 ..............................MKVASGSAAAAAG
                                     ..:::|:. :...
Id2 MAIQGCPIERPGRVHIKDAPPGSRFILNRAWCRAVSSAPCGGSLPVSP
                .         .         .         .
Id1 PSCSLKAGRTAGEVVL.GLSEQSVAISRCAGTRLPALLDEQQVNVLLY
    |.:|:|| |.::.|   :|||::|::|||: ..  .::       |||
Id2 PTSSMKAFRSGESVRKNSLSDHSLGISRSKTPVDDPM.......SLLY
                .         .  *      .         .
Id1 DMNGCYSRLKELVPTLPQNRKVSKVEILQHVIDYIRDLQLELNSESEV
    :||:|||:|||||||.:|||:||.|:|||||||||| |||:.|:|...:
Id2 NMNDCYSKLKELVPSIPQNKKVTKMEILQHVIDYILDLQIALDSHPTI
                .         .         .         .
Id1 GT....TGGRGLPVRAPLSTLNGEISALAAEAACVPADDRILCR.....
    ..     .|.. : |..|.|||.:||  |. :|. .|.: .   .
Id2 VSLHHQRPGQNQASRTRLTTLNTDISILSLQASEFPSELMSNDSKVLCG
```

FIG. 2. Amino acid sequence comparison between Id1 and Id2. The homologous helix-loop-helix motifs are underlined and two serine- and threonine-enriched weak homologies are boxed. The valine shown by an asterisk has been found to be an alanine in other bHLH proteins.

Chromosome mapping of the *Id-2* locus. The mouse chromosomal location of the *Id-2* locus was determined by interspecific backcross analysis with progeny derived from matings of (C57BL/6J × *M. spretus*)F_1 × C57BL/6J mice. This interspecific backcross mapping panel has been typed for more than 675 loci that are well distributed among all the autosomes as well as the X chromosome. C57BL/6J and *M. spretus* DNAs were digested with several enzymes and analyzed by Southern blot hybridization for informative restriction fragment length polymorphisms by using the Id2 cDNA probe. A 9.0-kb *M. spretus*-specific *Xba*I restriction fragment length polymorphism was used to follow the segregation of *Id-2* locus in backcross mice. The mapping results indicated that *Id-2* is located in the proximal region of mouse chromosome 12 tightly linked to *Pomc-1* and N-*myc*. The additional loci included in this analyses are *D12Nyu2* and *Lamb1-1*.

We have compared the interspecific backcross map of chromosome 12 with the December, 1990, chromosome 12 linkage map compiled by Davisson et al. and provided from GBASE, a computerized data base maintained at the Jackson Laboratory, Bar Harbor, Maine (data not shown). This map is based largely on genetic crosses among laboratory mouse strains and recombinant inbred strains. These comparative studies indicated that the interspecific backcross map is generally colinear with the composite map and that *Id-2* is located in a region of chromosome 12 that lacks any known mutations with effects suggestive of an alteration in *Id-2* locus. Finally, the proximal region of mouse chromosome 12 in which *Id-2* is located shares regions of homology with human chromosomes 2 and 7 (Fig. 3), suggesting that the human homolog of *Id-2* may reside on one of these chromosomes in humans.

Assignment of *Id-2* to the proximal region of mouse chromosome 12 clearly demonstrates that the locus is unlinked to *Id-1*, which has been assigned to mouse chromosome 2 (24a). But the two *Id* genes are probably products of a duplication event from a common ancestor, and rearrangement has occurred during evolution.

DNA binding inhibition by Id1 and Id2. To test whether Id2 could inhibit DNA binding and whether it behaves similarly to Id1, we first chose two DNA binding bHLH proteins, E2A.E47 and E2B.m3, as targets. E2A.E47 (30) binds strongly to the E2 sequence in the enhancer of immunoglobulin κ light chain gene (κE2 sequence) (18). Benezra et al. (6) have shown that Id1 inhibited DNA binding by E47 very efficiently. E2B.m3 is a mouse cDNA clone isolated by

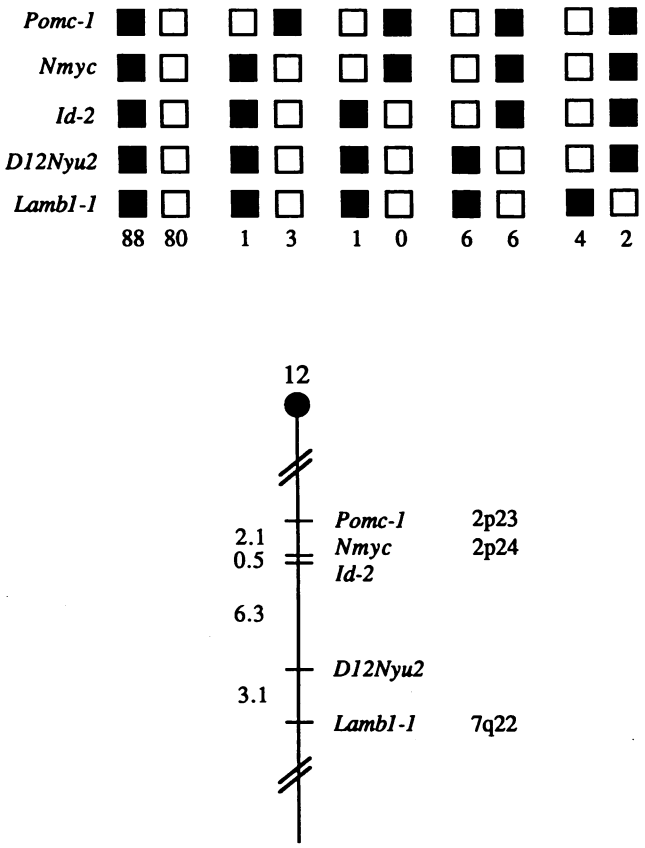

FIG. 3. Position of the *Id-2* locus on mouse chromosome 12. *Id-2* was placed on mouse chromosome 12 by interspecific backcross analysis. The segregation patterns of *Id-2* and flanking genes in 191 backcross animals that were typed in common for *Id-2* is shown at the top of the figure. For individual pairs of loci, more than 191 animals were typed. Each column represents the chromosome identified in the backcross progeny that was inherited from the (C57BL/6J × *M. spretus*)F_1 parent. The shaded boxes represent the presence of a C57BL/6J allele, and white boxes represent the presence of an *M. spretus* allele. The number of offspring inheriting each type of chromosome is listed at the bottom of each column. A partial chromosome 12 linkage map showing the location of *Id2* in relation to linked genes is shown at the bottom. Recombination distances between loci in centimorgans and the standard errors are shown to the left of the chromosome as follows: *Pomc-1*–2.1 ± 1.0–*Nmyc*–0.5 ± 0.5–*Id-2*, 6.3 ± 1.8–*D12Nyu2*–3.1 ± 1.3–*Lambl-1*. The positions of the *Pomc-1*, *Nmyc*, and *Lambl-1* loci on human chromosomes are shown to the right.

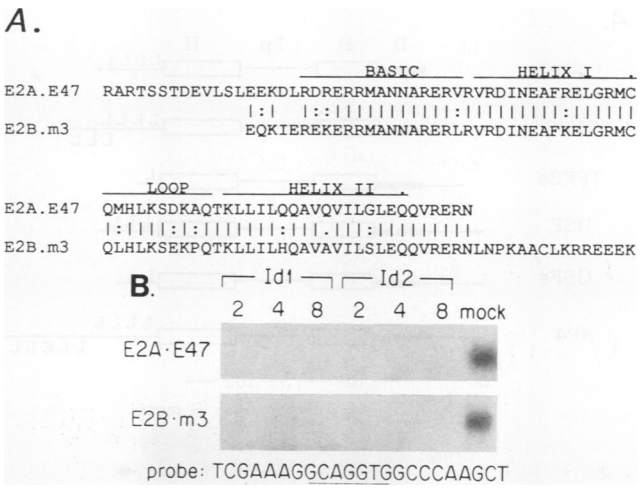

FIG. 4. DNA binding inhibition by Id1 and Id2. (A) Amino acid sequences of truncated E2A.E47 and E2B.m3 proteins used in the binding assays. The two sequences are aligned to show homology. The basic amino acid region and helix-loop-helix region are indicated. (B) DNA binding by E47 and m3 in the presence of Id1 or Id2. The molar ratio of Id to DNA binding proteins are indicated on the top of each lane. mock, binding reactions performed by preincubation with a mock translation mix. The probe used is the κE2 site as underlined.

homology to E47 (Fig. 4A), but m3 is apparently encoded by a different gene, named *E2B*, because its sequence dramatically diverges from that of E47 outside of the bHLH region (43). E2B.m3 also shares extended homology to the human ITF2 protein (22), but it is not clear whether m3 is the mouse homolog of ITF2, because the divergence between the two is unusually high. E2B.m3 is expressed ubiquitously (data not shown) and binds to the κE2 sequence efficiently as determined by electrophoretic mobility shift assay (Fig. 4B, lane mock). Both the DNA binding bHLH proteins and inhibitory Id proteins were synthesized in vitro by using rabbit reticulocyte lysates, and the DNA binding activity was monitored by electrophoretic mobility shift assay. When Id1 or Id2 was preincubated with the DNA binding E47 or m3 proteins at molar excess as indicated (Fig. 4B), the strong DNA binding to the κE2 sequence by these proteins was completely abolished. Apparent inhibition was also observed even when Id1 or Id2 was mixed with the binding proteins at an equal molar ratio (data not shown), and Id1 and Id2 did not show any significant difference in their abilities to inhibit DNA binding by either E47 or m3.

We then asked whether some bHLH protein might not be inhibited by either Id1 or Id2. With the same protocol used for E47 and m3, we examined truncated forms of c-*myc*, TFE3, USF, and AP4 (Fig. 5A). The c-*myc* oncogene is involved in cell proliferation and differentiation (3), and its DNA binding site has recently been identified as CACGTG (7). The TFE3 protein is an apparently ubiquitous transcription activator that binds to the E3 site (CACATG) of the immunoglobulin heavy chain gene (5). The USF protein (upstream stimulatory factor) is a ubiquitous cellular factor involved in efficient transcription in vitro of the adenovirus major late promoter, which binds to the UAE site (CACGTG) of the adenovirus major late promoter (21). Both TFE3 and USF can also bind strongly to the oligonucleotide containing the binding site for c-*myc*, but not to the κE2 sequence (Fig. 5B). The AP4 protein is also a ubiquitous transcription activator binding to the AP4 site (CAGCTG) of simian virus 40 enhancer (24), but it does not bind strongly to the κE2 sequence, the μE3 sequence, or the *myc*-binding site (data not shown). These proteins were produced by in vitro translation as detailed in Materials and Methods and bound to the *myc*-binding sequence probe (for c-*myc*, TFE3, and USF) or the AP4 site probe after preincubation with excess amounts of Id1 and Id2. DNA binding was then analyzed by electrophoretic mobility shift assay, and no inhibition of DNA binding by either of the Id proteins was observed for any of these bHLH proteins (Fig. 5B).

Thus, we have divided the bHLH protein family into two subsets on the basis of their susceptibilities to Id inhibition. The first set of bHLH proteins whose DNA binding can be inhibited by both Id proteins includes E2A.E47 and E2B.m3.

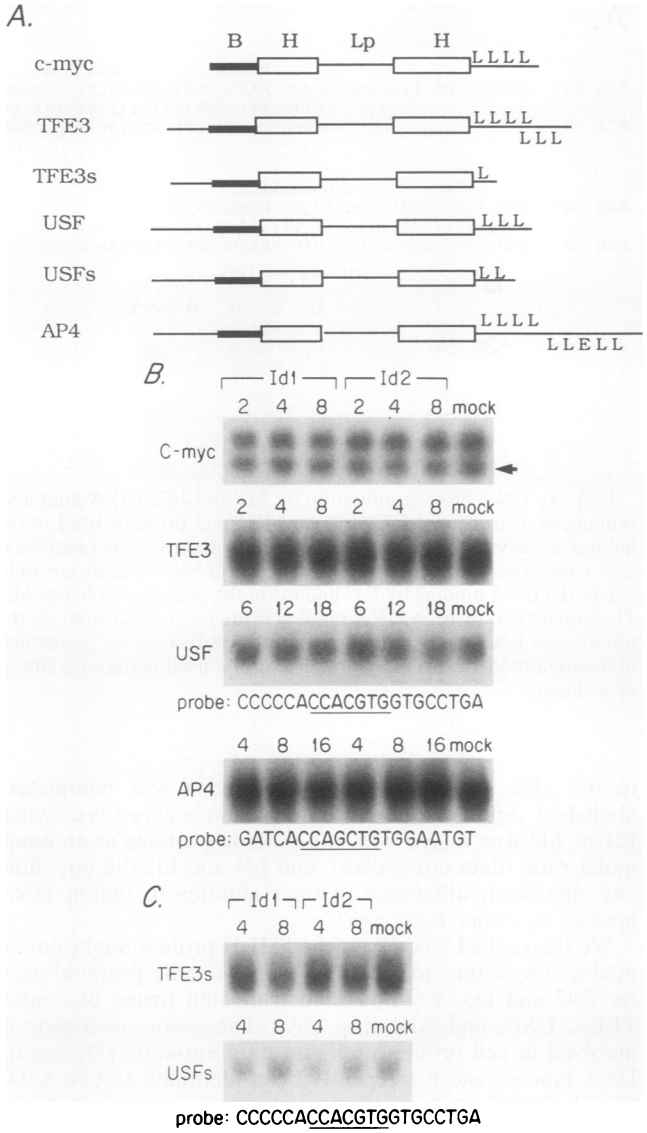

FIG. 5. Lack of DNA binding inhibition by Id1 and Id2 on bHLH leucine zipper proteins. (A) Schematic diagram of truncated DNA binding proteins used in the test. The basic amino acid regions are represented by a black bar. H, helices, shown in open boxes; LP, loops, shown as a thin line. The leucine zippers located C terminal to helix II are illustrated by an array of L's. The precise amino acid positions where the proteins were truncated are as follows: c-*myc*, 354 to 439 (3); TFE3, 118 to 240 (5); TFE3s, 118 to 204 (5); USF, 173 to 285 (21); USFs, 173 to 284 (21); and AP4, 7 to 193 (24). (B) DNA binding by the indicated proteins in the presence or absence of Id1 or Id2. The molar ratios of Id to DNA binding proteins are indicated on the top of each lane. mock, binding reactions performed by preincubation with a mock translation mix. The specific DNA complexes with c-*myc* are indicated by an arrow. The probe for c-*myc*, TFE3, and USF is shown below the complexes, and the core sequence is underlined. AP4 was bound to the AP4 site shown below. (C) DNA binding by TFEs and UFSs in the presence or absence of Id1 or Id2 as described in panel B.

DNA binding by the E2A.E12 and MyoD proteins as homodimers or as a heterodimer was also shown to be abolished by Id1 (6). This set of proteins falls into previously defined class I or class II of the bHLH proteins (31), which mutually interact to form heterodimers.

The second set of bHLH proteins, which were not affected by the Id proteins, belongs to the previously defined class III. They share a common structural feature that is not present in the proteins of the first set, namely, at least one leucine zipper motif (28) immediately C terminal to the second helix of the helix-loop-helix motif (Fig. 5A). The leucine zippers of USF and AP4 have been shown to strengthen the interaction between the two molecules in a homodimer (21, 24). To test whether the inability of the Id proteins to inhibit the bHLH proteins was due to strong interactions of the bHLH homodimers facilitated by the leucine zippers, we shortened the TFE3 and USF proteins from their C termini to remove the leucine zippers, generating two proteins, TFE3s and USFs (Fig. 5A). DNA binding by TFE3s and USFs in the presence of the Id proteins was tested as described above, and still no inhibition was observed (Fig. 5C). These results suggest that the reason that the Id proteins fail to inhibit the binding by this set of proteins is not simply because of strong homodimeric interaction of the bHLH proteins but is most likely because of incompatibility of the helix-loop-helix motifs in the Id proteins and the bHLH proteins. The Id proteins are probably dedicated to repress class I proteins and its associated class II proteins, but there may be different protein molecules serving as Id for proteins of class III proteins, or some of the proteins may not need to be repressed by this mechanism because they are general transcription factors active in all cells in which they are expressed.

Interaction of Id1 and Id2 with bHLH proteins. Id proteins are thought to prevent bHLH proteins from binding to DNA through formation of inactive heterodimers. To test this hypothesis, we directly examined the interaction of Ids with various bHLH proteins. We first generated GST fusion proteins of Id1 and Id2 by overexpression in *E. coli* and showed that these fusion proteins were able to inhibit DNA binding by E47 (data not shown). The proteins were then purified by binding to glutathione-agarose beads. To detect heterodimer formation, the beads bound with GST-Id fusion proteins were incubated with ^{35}S-labeled in vitro-translated bHLH proteins used in the binding assays (Fig. 4 and 5). The radioactive proteins which were brought down by the beads, presumably through interacting with the Id proteins, were analyzed by electrophoresis through SDS-polyacrylamide gels. As shown in Fig. 6A, while GST itself did not complex with any bHLH proteins (lanes 1), both GST-Id1 and GST-Id2 (lanes 2 and 3, respectively) can interact efficiently with E2A.E47s and E2B.m3 and to a lesser extent with MyoD. Both Id fusion proteins interacted very poorly with USFs, TFE3, and AP4. Id1 and Id2 also did not form homodimers efficiently (Fig. 6A). Using a similar approach as described above, we also analyzed the interaction of ^{35}S-labeled in vitro-translated Id1 and Id2 with GST-E12 and GST-E47 fusion proteins and found that both E12 and E47 can form complexes with the two Id proteins (Fig. 6B, lanes c and d). Because these experiments were performed with excess GST fusion proteins to allow complete heterodimerization, the efficiency of heterodimer formation can be estimated qualitatively by comparing the amount of protein precipitated (lanes 1, 2, and 3 of Fig. 6A or lanes b, c, and d of Fig. 6B) with the input amount (lane 4 of Fig. 6A and lane a of Fig. 6B show 50% of the input).

These results provide a molecular basis for inhibition of DNA binding by Id proteins through formation of heterodimers. We found that the bHLH proteins, whose binding activities could be inhibited, formed heterodimers with Id proteins, but those whose activities were not inhibited did

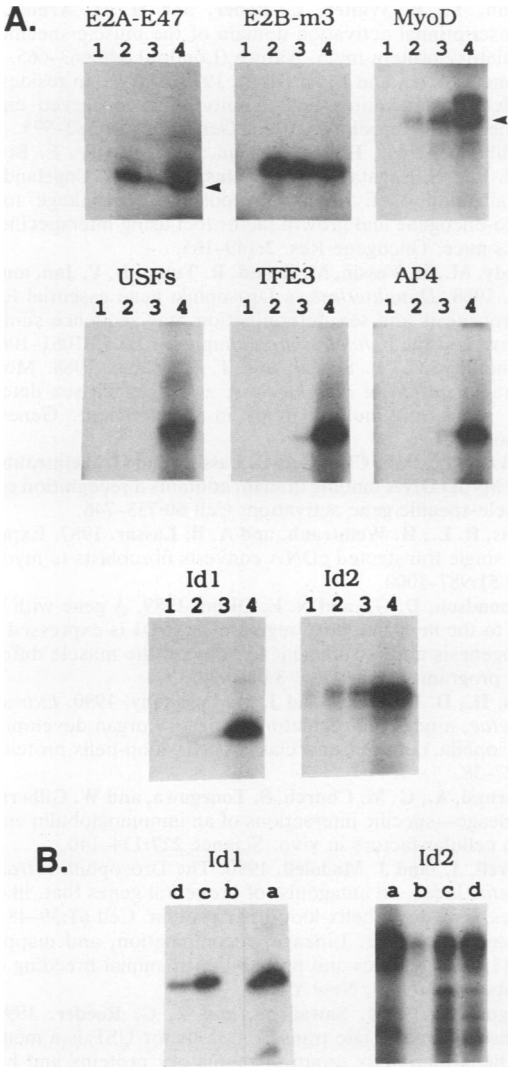

FIG. 6. Interaction of Id1 and Id2 with bHLH proteins. (A) bHLH proteins mixed with Id1 and Id2-GST fusion proteins. ^{35}S-labeled in vitro-translated bHLH proteins, as indicated on the top of each panel, were incubated with glutathione-agarose beads bound to equivalent amounts of GST (lane 1), GST-Id1 (lane 2), and GST-Id2 (lane 3). Lane 4, total translation extract of each indicated protein at an amount approximately 50% of what was used for lanes 1 to 3. Translated E47S and MyoD proteins are pointed out by arrows. (B) Id1 and Id2 proteins mixed with GST-E12 and GST-E47 fusion proteins. ^{35}S-labeled in vitro-translated Id1 and Id2, as indicated on the top of each panel, were incubated with glutathione-agarose beads bound to equivalent amounts of GST (lane b), GST-E12 (lane c), and GST-E47 (lane d). Lane a contained total translation extract of Id1 or Id2 at an amount approximately 50% of what was used for lanes b to d.

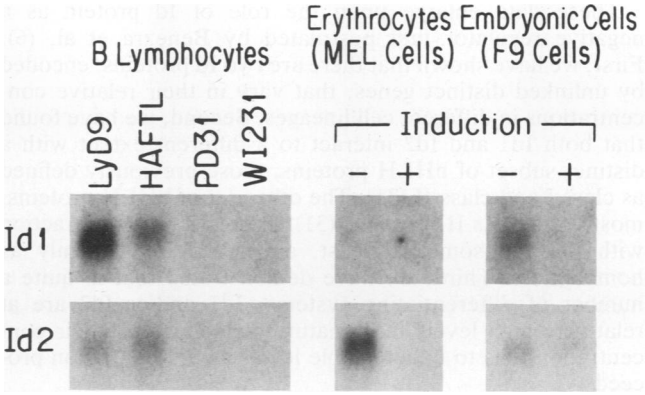

FIG. 7. Northern blot analysis of Id1 and Id2. Equal amounts of total RNA were loaded on the gel as described in Materials and Methods. Id1 mRNA was estimated to be 1.2 kb (6), and Id2 mRNA appears to be 1.6 kb.

not interact with the Id proteins. These results also strengthen our conclusion that the primary targets of the Id proteins Id1 and Id2 were products of the E2A and E2B genes or the class I bHLH proteins. The class II bHLH proteins may also complex with Id proteins, but the interaction is less efficient, as in the case of MyoD which both we and Benezra et al. (6) found to form less complex with Id1 or Id2. We also found that *tal*, another class II bHLH protein (23), did not interact with Id1 and interacted very weakly if at all with Id2 (data not shown). Considering that several control mechanisms, such as tissue-specific gene expression (29) and the dependence of their DNA binding activities on formation of heterodimers with the class I proteins (23, 42), already exist for the class II proteins, it may be less important for class II proteins to have the Id proteins as additional regulators. In contrast, the constitutively expressed class I proteins need to be inhibited by the Id proteins at a defined developmental stage or cell type. Moreover, sequestering the class I proteins could indirectly inhibit class II proteins, whose function depends on the class I proteins. The fact that the Id proteins exist mostly as monomers should also make them more effective inhibitors for the class I proteins.

Expression of the *Id-1* and *Id-2* genes. The Id proteins are clearly efficient negative regulators for a specific set of bHLH proteins. One might then expect that the regulation of expression of the *Id-1* and *Id-2* genes would correlate with their functional roles. Because the Id proteins inhibit the function of bHLH proteins involved in cellular differentiation, we examined the levels of Id1 and Id2 mRNAs in several cell types at different stages of differentiation by Northern blot analysis (Fig. 7). In general, Id1 and Id2 were expressed at different levels in different cell types, but their expression was down-regulated by cell differentiation in all of the cell types examined. For B lymphocytes, we analyzed RNAs from cell lines whose developmental stages have been defined on the basis of the patterns of transcription and rearrangement of the immunoglobulin genes and found that the level of *Id-1* is higher than that of *Id-2* and the expression of both genes was decreased as differentiation proceeded from pro-B cells (LyD9), to pre-B cells (HAFTL and PD31), to mature B cells (WEHI-231). Similar results were also observed when the levels of RNA were measured with PCR assays (38). In MEL cells, Id1 was hardly detected while Id2 was very abundant, but the expression of Id2 was dramatically shut off after the cells were induced to differentiate with 1.5% dimethyl sulfoxide for 48 h. (Benezra et al. [6] had previously clearly detected Id1 in MEL cells; the difference from our result might be due to variations of sources of MEL cells or simply due to the different sensitivities of the analyses.) Induction of mouse F9 embryonal carcinoma cells also reduced the expression of both *Id-1* and *Id-2* genes. The expression of *Id-1* was also shut off in F3 aza-myoblasts upon differentiation (6).

These data enlarge upon the role of Id protein as a negative regulator first postulated by Benezra et al. (6). First, we have shown that there are two Id proteins, encoded by unlinked distinct genes, that vary in their relative concentrations in different cell lineages. Second, we have found that both Id1 and Id2 interact to a different extent with a distinct subset of bHLH proteins, those previously defined as class I and class II (31). The other set of bHLH proteins, mostly the class III proteins (31), are resistant to interaction with Id, and some, at least, appear to exist mainly as homodimers. Third, we have demonstrated that in quite a number of differentiating systems, Id1 and/or Id2 are at relatively high levels in immature cells and that their concentration falls to undetectable levels as differentiation proceeds.

ACKNOWLEDGMENTS

We thank R. Benezra and H. Weintraub for the Id1 probe, J. A. Goliger and L. Gudas for the RNA samples from F9 cells, S. Luo and D. Page for the mouse genomic library, Y. Ichikawa and S. Tonegawa for the thymus cDNA library, R. Andino and M. Schlissel for advice, S. Ghosh for reading the manuscript, and D. Swing, D. J. Gilbert, and B. Cho for excellent technical assistance in chromosome mapping.

X.-H.S. was supported by a postdoctoral fellowship from the Cancer Research Institute. This work was supported in part by Public Health Service grant GM 39458 (to D.B.) and by the National Cancer Institute, DHHS, under contract N01-CO-74101 with ABL.

ADDENDUM

Since the submission of the manuscript, Christy et al. reported the cloning of an Id-like protein named HLH462 (Proc. Natl. Acad. Sci. USA **88**:1815–1819, 1991). This protein shares amino acid sequence homology with both Id1 and Id2 in the helix-loop-helix region with some divergence in the loop. HLH462 also contains one of the serine- and threonine-rich homology regions discussed in the text. Because Id1 and Id2 are more closely related, HLH462 should probably be considered Id3.

REFERENCES

1. **Alessandrini, A., J. H. Pierce, D. Baltimore, and S. V. Desiderio.** 1987. Continuing rearrangement of immunoglobulin and T-cell receptor genes in a Ha-ras-transformed lympoid progenitor cell line. Proc. Natl. Acad. Sci. USA **84**:1799–1803.
2. **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.
3. **Alt, F. W., R. DePinho, K. Zimmerman, E. Legouy, K. Hatton, P. Ferrier, A. Tesfaye, G. Yancopoulos, and P. Nisen.** 1986. The human *myc* gene family. Cold Spring Harbor Symp. Quant. Biol. **51**:931–941.
4. **Ausubel, F. M., R. Brent, R. E. Kingston, D. D. Moore, J. G. Seidman, J. A. Smith, and K. Struhl (ed.).** 1987. Current protocols in molecular biology. John Wiley & Sons, New York.
5. **Beckmann, H., L.-K. Su, and T. Kadesch.** 1990. TFE3: a helix-loop-helix protein that activates transcription through the immunoglobulin enhancer μE3 motif. Genes Dev. **4**:167–179.
6. **Benezra, R., R. L. Davis, D. Lockshon, D. L. Turner, and H. Weintraub.** 1990. The protein Id: a negative regulator of helix-loop-helix DNA binding proteins. Cell **61**:49–59.
7. **Blackwell, T. K., L. Kretzner, E. M. Blackwood, R. N. Eisenman, and H. Weintraub.** 1990. Sequence-specific DNA binding by the c-myc protein. Science **250**:1149–1151.
8. **Braun, T., D. G. Buschhausen, E. Bober, E. Tannich, and H. H. Arnold.** 1989. A novel human muscle factor related to but distinct from *MyoD*1 induces myogenic conversion in 10T1/2 fibroblasts. EMBO J. **8**:701–709.
9. **Braun, T., B. Winter, E. Bober, and H. H. Arnold.** 1990. Transcriptional activation domain of the muscle-specific gene-regulatory protein myf5. Nature (London) **346**:663–665.
10. **Brennan, T. J., and E. N. Olson.** 1990. Myogenin resides in the nucleus and acquires high affinity of a conserved enhancer element on heterodimerization. Genes Dev. **4**:582–595.
11. **Buchberg, A. M., H. G. Bedigian, B. A. Taylor, E. Brownell, J. N. Ihle, S. Nagata, N. A. Jenkins, and N. G. Copeland.** 1988. Localization of Evi-2 to chromosome 11: linkage to other proto-oncogene and growth factor loci using interspecific backcross mice. Oncogene Res. **2**:149–165.
12. **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.
13. **Cronmiller, C., P. Schedl, and T. W. Cline.** 1988. Molecular characterization of *daughterless*, a Drosophila sex determination gene with multiple roles in development. Genes Dev. **2**:1666–1676.
14. **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.
15. **Davis, R. L., H. Weintraub, and A. B. Lassar.** 1987. Expression of a single transfected cDNA converts fibroblasts to myoblasts. Cell **51**:987–1000.
16. **Edmondson, D. G., and N. E. Olson.** 1989. A gene with homology to the *myc* similarity region of *MyoD*1 is expressed during myogenesis and is sufficient to activate the muscle differentiation program. Genes Dev. **3**:628–640.
17. **Ellis, H., D. R. Spann, and J. W. Posakony.** 1990. *Extramacrochaetae*, a negative regulator of sensory organ development in Drosophila, defines a new class of helix-loop-helix proteins. Cell **61**:27–38.
18. **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.
19. **Garrell, J., and J. Modolell.** 1990. The Drosophila *extramacrochaetae* locus, an antagonist of proneural genes that, like these genes, encodes a helix-loop-helix protein. Cell **61**:39–48.
20. **Green, E. L.** 1981. Linkage, recombination, and mapping, p. 77–113. *In* Genetics and probability in animal breeding experiments. Macmillan, New York.
21. **Gregor, P. D., M. Sawadogo, and R. G. Roeder.** 1990. The adenovirus major late transcription factor USF is a member of the helix-loop-helix group of regulatory proteins and binds to DNA as a dimer. Genes Dev. **4**:1730–1740.
22. **Henthrorn, P., R. McCarrick-Walmsley, and T. Kadesch.** 1990. Sequence of the cDNA encoding ITF-2, a positive-acting transcription factor. Nucleic Acids Res. **18**:678.
23. **Hsu, H.-L., J.-T. Cheng, Q. Chen, and R. Baer.** 1991. Enhancer-binding activity of the *tal-1* oncoprotein in association with the E47/E12 helix-loop-helix proteins. Mol. Cell. Biol. **11**:3037–3042.
24. **Hu, Y.-F., B. Luscher, A. Admon, N. Mermod, and R. Tjian.** 1990. Transcription factor AP-4 contains multiple dimerization domains that regulate dimer specificity. Genes Dev. **4**:1741–1752.

24a. **Jenkins, N. A., and N. G. Copeland.** Unpublished data.

25. **Jenkins, N. A., N. G. Copeland, B. A. Taylor, and B. K. Lee.** 1982. Organization, distribution, and stability of endogenous ecotropic murine leukemia virus DNA sequences in chromosomes of *Mus musculus*. J. Virol. **43**:26–36.
26. **Justice, M. J., C. M. Silan, J. D. Ceci, A. M. Buchberg, N. G. Copeland, and N. A. Jenkins.** 1990. A molecular genetic linkage map of mouse chromosome 13 anchored by the beige (*bg*) and satin (*sa*) loci. Genomics **6**:341–351.
27. **Kamps, M. P., C. Murre, X.-H. Sun, and D. Baltimore.** 1990. A new homeobox gene contributes the DNA binding domain of the t(1;19) translocation protein in pre-B ALL. Cell **60**:547–555.
28. **Landschulz, W. H., P. F. Johnson, and S. L. McKnight.** 1988. The leucine zipper: a hypothetical structure common to a new class of DNA binding proteins. Science **240**:1759–1764.
29. **Murre, C., and D. Baltimore.** The helix-loop-helix motif: struc-

ture and function. *In* K. Yamamoto and S. McKnight (ed.), Transcriptional regulation, in press. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.

30. **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.
31. **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.
32. **Oettinger, H. F., H. Streeter, E. Lose, N. G. Copeland, D. J. Gilbert, M. J. Justice, N. A. Jenkins, T. K. Mohandas, and M. Bernfield.** Submitted for publication.
33. **Palacios, R., H. Karasuyama, and A. Rolink.** 1987. Ly1+ pro-B lymphocyte clones: phenotype, growth requirements and differentiation in vitro and in vivo. EMBO J. **6**:3687–3693.
34. **Patel, V. P., and H. F. Lodish.** 1987. A fibronectin matrix is required for differentiation of murine erythroleukemia cells into reticulocytes. J. Cell Biol. **105**:3105–3118.
35. **Rhodes, S. J., and S. F. Konieczny.** 1989. Identification of MRF4: a new member of the muscle regulatory factor gene family. Genes Dev. **3**:2050–2061.
36. **Saiki, R. K., D. H. Gelfand, S. Stoffel, S. J. Scharf, R. Higuchi, G. T. Horn, K. B. Mullis, and H. A. Erlich.** 1988. Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. Science **239**:487–494.
37. **Sanger, F., S. Nicklen, and A. R. Coulson.** 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA **74**:5463–5467.
38. **Schlissel, M., and D. Baltimore.** Unpublished data.
39. **Siracusa, L. D., N. A. Jenkins, and N. G. Copeland.** 1991. Identification and applications of repetitive probes for gene mapping in the mouse. Genetics **127**:169–179.
40. **Smith, D. B., and K. S. Johnson.** 1988. Single-step purification of polypeptides expressed in *Escherichia coli* as fusions with glutathione S-transferase. Gene **67**:31–40.
41. **Stricklend, S., and V. Mahdadi.** 1978. The induction of differentiation in teratocarcinoma stem cells by retinoic acid. Cell **15**:393–403.
42. **Sun, X.-H., and D. Baltimore.** 1991. An inhibitory domain of E12 transcription factor prevents DNA binding in E12 homodimers but not in E12 heterodimers. Cell **64**:459–470.
43. **Sun, X.-H., and D. Baltimore.** Unpublished data.
44. **Villares, R., and C. V. Cabrera.** 1987. The *achaete-scute* gene complex of D. melanogaster: conserved domains in a subset of genes required for neurogenesis and their homology to *myc*. Cell **50**:415–424.
45. **Voronova, A., and D. Baltimore.** 1990. Mutations that disrupt DNA binding and dimer formation in the E47 helix-loop-helix protein map to distinct domains. Proc. Natl. Acad. Sci. USA **87**:4722–4726.
46. **Wright, W. E., D. A. Sassoon, and V. K. Lin.** 1989. Myogenin, a factor regulating myogenesis, has a domain homologous to *MyoD*. Cell **56**:607–617.