MOLECULAR AND CELLULAR BIOLOGY, Feb. 1994, p. 1245–1255
0270-7306/94/$04.00+0

Vol. 14, No. 2

HEN1 Encodes a 20-Kilodalton Phosphoprotein That Binds an Extended E-Box Motif as a Homodimer

LAMORNA BROWN AND RICHARD BAER*

Department of Microbiology, University of Texas Southwestern Medical Center, Dallas, Texas 75235

Received 26 July 1993/Returned for modification 29 September 1993/Accepted 22 October 1993

***HEN1* and *HEN2* encode neuron-specific polypeptides that contain the basic helix-loop-helix (bHLH) motif, a protein dimerization and DNA-binding domain common to several known transcription factors. We now describe characteristics of the HEN1 gene product that are consistent with its postulated role as a transcription factor that functions during development of the mammalian nervous system. Thus, transcription of the *HEN1* gene is activated upon the induction of neural differentiation in PC12 cells by nerve growth factor. *HEN1* encodes a 20-kDa polypeptide (pp20^{HEN1}) that is phosphorylated exclusively at serine residues and forms dimeric bHLH complexes either by self-association or by heterologous interaction with the *E2A* gene products (E12 or E47). The resultant HEN1/HEN1 homodimers and HEN1/E2A heterodimers bind DNA in a sequence-specific manner. Moreover, a binding site selection procedure revealed that HEN1-HEN1 homodimers preferentially recognize E-box motifs represented by an 18-bp consensus sequence (GGGNCG CAGCTGCGNCCC). The E-box half-site recognized by HEN1 polypeptides (GGGNCGCAG) is distinct from those of other known bHLH proteins, suggesting that HEN1 binds, and regulates the transcription of, a unique subset of target genes during neural development.**

Many transcription factors possess the basic helix-loop-helix (bHLH) motif, a conserved domain of 50 to 60 amino acids that mediates protein dimerization and DNA recognition (for reviews, see references 35, 40, and 41). The bHLH motif has the potential to form two amphipathic α helices separated by an intervening loop. Structural and mutational studies show that both helices are involved in protein dimerization, whereas DNA binding is largely mediated by basic residues at the amino-terminal segment of the first helix (17, 20, 49). The bHLH proteins fall into at least three classes (41): the ubiquitously expressed class A proteins, the tissue-specific class B proteins, and a third class of proteins that have the bHLH domain arrayed in tandem with a leucine zipper (the bHLH-Zip motif). Some tissue-specific class B proteins can be further assigned to specific subgroups on the basis of exceptional amino acid homology, similar expression patterns, and common functional properties. For example, the myogenic subgroup consists of four muscle-specific bHLH proteins (MyoD, myogenin, Myf5 and Myf6 [herculin and MRF4]) which share 75% amino acid identity within the bHLH domain; notably, each of these has the ability to induce myogenic conversion of primitive mesodermal cells (reviewed in references 24, 42, and 50). Likewise, the *achaete-scute* complex encodes four highly related bHLH proteins that facilitate cell type determination during neurogenesis in *Drosophila melanogaster* (2, 48). Distinct subgroups of bHLH proteins have also been defined on the basis of common pathogenic properties. For example, the *TAL1*, *TAL2*, and *LYL1* genes are activated by tumor-specific chromosome rearrangements in patients with T-cell acute lymphoblastic leukemia (4, 7, 13, 22, 39, 52). Since the proteins encoded by these genes share a striking degree of amino acid identity (>85%) within their bHLH domains, they are likely to promote T-cell leukemogenesis by a common mechanism.

The bHLH proteins do not bind DNA as isolated polypeptides, but instead they associate to form dimers which in turn recognize DNA in a sequence-specific manner (35, 40, 41). Although certain class A proteins form bHLH homodimers by self-association, the known class B proteins do not homodimerize effectively; instead, they interact with class A proteins to form stable bHLH heterodimers. Consequently, the DNA-binding and functional properties of tissue-specific class B proteins are dependent on interaction with class A proteins. For example, the induction of muscle differentiation by bHLH myogenic factors (e.g., MyoD1) is mediated by heterodimers (e.g., MyoD1/E12) involving class A bHLH polypeptides such as E12 or E47 (37). The bHLH dimers recognize the canonical E-box sequence (CANNTG), a *cis*-acting control element found in many eukaryotic transcription enhancers (15, 19, 40). However, certain bHLH dimers bind with higher efficiency to more defined E-box sequences that have specific nucleotides at the internal and flanking residues of the CANNTG core (9). Thus, the subordinate genes subject to regulation by a given bHLH protein are likely to be determined, at least in part, by its dimerization properties and DNA-binding preferences.

We recently identified a unique subgroup of bHLH genes, *HEN1* and *HEN2* (also known as *NSCL1* and *NSCL2*), by cross-hybridization with *TAL1* (11). *HEN1* and *HEN2* encode polypeptides of 133 and 135 amino acids, respectively, that are exclusively expressed in the embryonic nervous system and in certain tumor cell lines derived from neuroendocrine cells (5, 11, 26). The expression pattern of these genes suggests that they function as regulators of neural differentiation. Given the properties of other tissue-specific class B bHLH proteins, we anticipated that HEN1 and HEN2 would associate with class A bHLH proteins, such as E12 and E47, and that the ensuing heterodimers would recognize E-box enhancer motifs. We now report that HEN1 is a 20-kDa phosphoprotein that can indeed form heterodimers with E12 and E47. Unexpectedly, however, we

* Corresponding author. Mailing address: Department of Microbiology, UT Southwestern Medical Center, 5323 Harry Hines Blvd., Dallas, TX 75235.

found that HEN1 polypeptides also self-associate to form homodimers that recognize an extended E-box motif.

MATERIALS AND METHODS

RNase protection assays for rat *Hen1* transcripts. A 98-bp fragment of rat *Hen1* sequence was obtained by using oligonucleotides HMT3.2 (5′-GCCTTCGCCGAGCTGCGC A-3′) and HMT3.3 (5′-TAGGAGATATAGCAGATGGC-3′) as primers for PCR amplification of rat genomic DNA; these primers were designed from conserved sequences encoding the bHLH domains of mouse and human *Hen1* (11). The 98-bp PCR fragment (GCCTTCGCCGAGCTGCGCAAGC TGCTGCCCACTCTGCCCCCGGACAAGAAACTCTCC AAGATTGAGATCCTGCGCCTGGCCATCTGCTATAT CTCCTA) was cloned into the *Sma*I site of pGEM4. The resultant pGEM4/RatHEN1 plasmid was then linearized with *Eco*RI and transcribed with T7 RNA polymerase in the presence of [γ-^{32}P]ATP to generate a radiolabeled RNA fragment of 166 nucleotides. This fragment was used as a probe in RNase protection assays with the Ambion RNase protection assay kit (Austin, Tex.).

PC-12 differentiation. PC-12 cells were grown in Dulbecco modified Eagle medium (supplemented with 10% fetal calf serum, 100 U of penicillin per ml, 100 μg of streptomycin per ml, 2 mM glutamine, and 0.1 mM nonessential amino acids) at 37°C in 7.5% CO_2. The cells were plated out at 7×10^6 cells per 150-mm-diameter plate. After 3 days, the media were replaced with 10 ml of fresh supplemented Dulbecco modified Eagle medium, and the cells were incubated overnight. The following day, the cells were treated with 0.1 μg of 2.5 S nerve growth factor (NGF) (UBI, Lake Placid, N.Y.) per ml and incubated for 48 h. The media were then replaced with fresh, supplemented Dulbecco modified Eagle medium containing 0.1 μg of NGF per ml, and the cells were incubated for an additional 16 h. RNA was harvested from treated and untreated cells by using the Stratagene RNA isolation kit.

Production of antisera. The pGEX3X/HEN1(1-59) plasmid was used for the bacterial expression of GST-HEN1(1-59), a glutathione *S*-transferase (GST) fusion protein that contains the 59 amino-terminal residues of human HEN1. The plasmid was constructed by cloning a filled-in *Bam*HI-*Pst*I fragment of genomic *HEN1* DNA into the *Sma*I site of pGEX-3X (45). GST-HEN1(1-59) polypeptides were then expressed in *Escherichia coli*, purified by affinity chromatography on glutathione-agarose beads, and used to produce antisera from two rabbits as described previously (14). Rabbit antisera specific for the TAL1 and E2A polypeptides have been reported (14, 31).

Immunoprecipitations and phosphoamino acid analysis. An expression plasmid encoding the entire open reading frame of human *HEN1* was generated by cloning a *Bam*HI-*Sma*I genomic HEN1 DNA fragment into the pCMV4 vector (3). Cos1 cells were transiently transfected with the expression plasmid and radiolabeled with [^{35}S]methionine or $^{32}P_i$ as described previously (14). Immunoprecipitations of the radiolabeled cell lysates, either in the presence or the absence of polypeptide competitors, was performed as described by Cheng et al. (14). Likewise, phosphoamino acid analysis of immunoprecipitated polypeptides was also conducted as previously described (14), by the method of Boyle et al. (10).

Protein dimerization assay. GST fusion proteins were prepared as described above except that the fusion proteins were not eluted from the glutathione-agarose beads; instead, the loaded beads were used directly in the binding assays as a 50% slurry in buffer C (described by Smith and Johnson [45]). The GST-E12(217-645) polypeptide contains the 438 carboxy-terminal residues of human E12, including the bHLH domain. GST-E47(524-609) harbors an 86-amino-acid stretch of human E47 that encompasses the bHLH domain. GST-TAL1(166-331) contains the 166 carboxy-terminal residues of human TAL1, including the bHLH domain (30). The GST-HEN1(1-133) fusion protein includes the entire amino acid sequence of human HEN1. For production of in vitro-translated HEN1, a *Bam*HI-*Bgl*II fragment containing the entire open reading frame of human HEN1 was cloned into the *Bam*HI site of pBS-SK (Stratagene) to create the HEN1-BB2/pBS plasmid. This plasmid was linearized and transcribed by using the Megascript T_7 transcription kit as described by the manufacturer (Ambion). In vitro translation of RNA in rabbit reticulocyte lysates was performed in the presence of [^{35}S]methionine as instructed by the manufacturer (Promega). Protein dimerization assays were performed by adding 10 μl of the HEN1-programmed reticulocyte lysate and 50 μl of glutathione-agarose beads (loaded with GST fusion protein) to 1 ml of ELB buffer (250 mM NaCl, 50 mM *N*-2-hydroxyethylpiperazine-*N*′-2-ethanesulfonic acid [HEPES, pH 7.4], 5 mM EDTA, 1 mM phenylmethylsulfonyl fluoride, 0.5 mM dithiothreitol, 0.1% Nonidet P-40) containing 10 mg of bovine serum albumin (BSA) per ml. Following incubation for 1 to 2 h at 4°C, the beads were washed five times with 1 ml of ELB buffer and boiled for 5 min in 30 μl of Laemmli sample buffer (36) with 5% β-mercaptoethanol. The eluants were fractionated by electrophoresis on a sodium dodecyl sulfate (SDS)–15% polyacrylamide gel.

CASTing. Cyclic amplification and selection of targets (CASTing) was performed as described by Wright et al. (51). Briefly, Tag-HEN1 protein was produced from HEN1/pBS-HA, a pBluescript-SK plasmid encoding the influenza virus hemagglutinin epitope (MAYPYDVPDYAGGPM) (21) fused to residues 8 to 133 of human HEN1. Linearized HEN1/pBS-HA was transcribed by using the Megascript T_3 transcription kit (Ambion) and then subjected to in vitro translation in rabbit reticulocyte lysates (Promega). The target oligonucleotide for CASTing contains PCR primer sites that flank a central core of 35 random nucleotides (25). Five micrograms of the oligonucleotide was converted to double-stranded DNA by annealing an excess of 3′ primer and performing one primer extension reaction with *Taq* DNA polymerase (25). For the first round of CASTing, a 20-μl binding reaction mixture containing the double-stranded oligonucleotide (10 μl), 4 μl of rabbit reticulocyte lysate, and 2 μl of binding buffer (20 mM HEPES [pH 7.6], 100 mM KCl, 20% glycerol, 0.2 mM EDTA, 1 mM dithiothreitol, 10 mg of BSA per ml) was incubated at room temperature for 20 min. Alternatively, for CASTing experiments with whole-cell extracts from transfected Cos1 cells (e.g., see Fig. 6), each binding reaction mixture included 4 μl of the Cos1 cell extract in lieu of the reticulocyte lysate; these extracts were prepared as described by Lassar et al. (37). In either case, 2 μl of magnetic beads coated with the 12CA5 monoclonal antibody was added, and the mixture was incubated for an additional 30 min at 4°C with gentle agitation. The beads were then resuspended in 500 μl of washing buffer (1× phosphate-buffered saline with 0.1% BSA and 0.1% Nonidet P-40), retrieved with a 12-lb (1 lb = 453.59237 g) magnet, washed four times with 500 μl of washing buffer, and finally resuspended in 100 μl of PCR solution (*Taq*I buffer containing 200 μM deoxynucleoside triphosphate and 1 μM [each] the 3′ and 5′ primers). As described previously (51) overam-

plification of the DNA was avoided by monitoring aliquots taken every five PCR cycles by gel electrophoresis; amplified DNA from the appropriate sample was precipitated with ethanol, resuspended in 20 μl of binding buffer, and subjected to a new round of CASTing. After six rounds, the selected DNA was PCR amplified and cloned into M13mp18 phage vector. Individual clones were then subjected to dideoxynucleotide sequence analysis (44).

EMSA. Reading 5′ to 3′, sequences of the upper strand of each complementary pair of oligonucleotides are as follows; the consensus probe, TCGAAGGACGCAGCTGATCCCC; the mutant probe, TCGAAGGACGCCGCTTATCCCC; and the hybrid probe AATTAACAGCTGCGTCCC. ^{32}P-labeled, double-stranded oligonucleotide probes were prepared and used for electrophoretic mobility shift assays (EMSAs) as described previously (30). Each EMSA reaction mixture contained either 10 μl of rabbit reticulocyte lysate or 10 μl of nuclear extract from transfected Cos1 cells (18). For supershift experiments, the EMSA reaction mixtures were supplemented with 1 μl of the appropriate rabbit antiserum or 5 μl of the 12CA5 monoclonal antibody. The in vitro-translated HEN1 polypeptides were generated as described above. The pE12R plasmid was used to produce E12 polypeptides by in vitro transcription and translation (41).

RESULTS

***Hen1* transcription is activated upon neuronal differentiation of PC-12 cells.** The rat adrenal pheochromocytoma cell line PC-12 responds to treatment with NGF by developing a neuronal phenotype that includes the production of neurotransmitters and the development of neurites (27). A change in the expression of *Hen1* in NGF-induced PC-12 cells would support the potential role of this gene in neuronal differentiation. In order to utilize this system, a rat *Hen1* probe was required; hence, the mouse and human *HEN1* sequences were compared, and primers were designed from conserved nucleotide sequences encoding helix I and helix II of the bHLH domain (see Materials and Methods). These primers were used to PCR amplify a 98-bp *Hen1* sequence from rat genomic DNA. The rat PCR product was cloned into pGEM4 and subjected to nucleotide sequence analysis. The resultant pGEM4/RatHEN1 plasmid was transcribed with T7 RNA polymerase in the presence of [γ-^{32}P]ATP to generate a 166-nucleotide RNA fragment. This fragment, which contains 98 residues of antisense rat *Hen1* mRNA sequence, was then used as a probe in RNase protection assays.

PC-12 cells were grown for 3 days in the presence or absence of NGF. Neurite outgrowth was clearly visible in the treated cells. RNA was harvested from the cells and annealed with the radiolabeled rat RNase protection assay probe. The samples were treated with RNase, and the products of each digest were analyzed by denaturing polyacrylamide gel electrophoresis and autoradiography. As illustrated in Fig. 1, NGF-treated PC-12 cells were stimulated to produce *Hen1* transcripts, while none was present in the untreated cells. Hence, *Hen1* transcription is activated upon NGF-induced neuronal differentiation of PC-12 cells.

***HEN1* encodes a 20-kDa polypeptide.** *HEN1* and *HEN2* potentially encode proteins of 133 and 135 amino acids, respectively (11). The carboxy-terminal 74 residues of these proteins, including the bHLH domains, are highly homologous (72 amino acids identical), suggesting that they may dimerize and bind DNA in a similar, if not identical, manner. To identify protein products of the *HEN1* gene, polyclonal

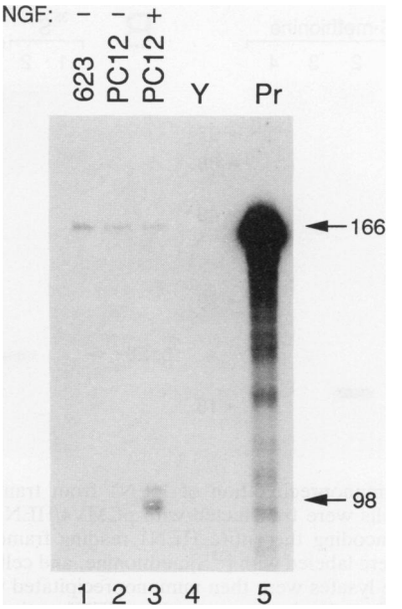

FIG. 1. Activation of *HEN1* gene transcription upon NGF treatment of PC-12 cells. A ^{32}P-labeled RNA probe of 166 nucleotides was generated by in vitro transcription of a plasmid containing 98 residues of antisense rat *Hen1* sequence. The radiolabeled probe was annealed with 10 μg of either yeast (Y) tRNA (lane 4) or total RNA from 623 rat medullary thyroid carcinoma cells (lane 1), untreated PC12 cells (lane 2), or NGF-treated PC12 cells (lane 3). After RNase treatment and denaturing polyacrylamide gel electrophoresis, radiolabeled RNA fragments were detected by autoradiography. An aliquot of the untreated RNA probe (Pr) was also fractionated (lane 5).

antisera were raised by immunizing rabbits with GST-HEN1(1-59), a fusion protein that contains the amino-terminal 59 residues of human HEN1. Since this region bears little homology to HEN2, these antisera should recognize HEN1 polypeptides specifically. Cos1 cells were then transfected with pCMV4/HEN1, a mammalian expression vector containing the entire open reading frame of *HEN1*. Approximately 40 h later, the transfected cells were briefly starved in a methionine-free medium and then cultured for 2 h in the presence of [^{35}S]methionine. As shown in Fig. 2A, an anti-HEN1 serum (lane 2), but not the corresponding preimmune serum (lane 1), immunoprecipitated a 20-kDa species (pp20^{HEN1}) from radiolabeled lysates of transfected Cos1 cells. Immunoprecipitation of pp20^{HEN1} was specific in the sense that it could be blocked by competition with unlabeled GST-HEN1(1-59), the HEN1 fusion protein used as an immunogen (Fig. 2A, lane 3), but not with an equivalent quantity of unlabeled GST polypeptide (lane 4). Moreover, pp20^{HEN1} was not observed in immunoprecipitates from Cos1 cells that were transfected with the parental pCMV4 expression plasmid (data not shown). The electrophoretic mobility of pp20^{HEN1} is slower than expected in light of its predicted molecular mass (14.5 kDa).

HEN1 is phosphorylated at serine residues. The functional properties of transcription factors are often regulated by protein phosphorylation (32). To determine whether HEN1 is also modified in this manner, parallel cultures of pCMV4/HEN1-transfected Cos1 cells were labeled with either [^{35}S]methionine or $^{32}P_i$. As illustrated in Fig. 2B, pp20^{HEN1} was readily immunoprecipitated from lysates of both ^{35}S-labeled and ^{32}P-labeled Cos1 cells. Hence, pp20^{HEN1} is

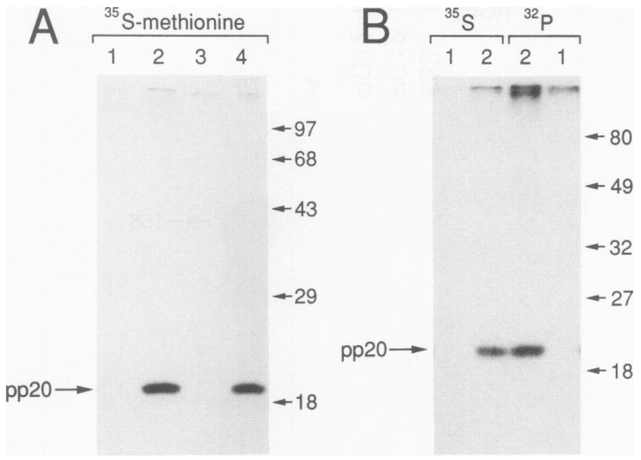

FIG. 2. Immunoprecipitation of HEN1 from transfected Cos1 cells. Cos1 cells were transfected with pCMV4/HEN1, an expression vector encoding the entire HEN1 reading frame. (A) Transfected cells were labeled with [^{35}S]methionine, and cell lysates were prepared. The lysates were then immunoprecipitated with a preimmune serum (lane 1); the corresponding HEN1 antiserum (lane 2); HEN1 antiserum in the presence of a specific competitor, GST-HEN1(1-59) (lane 3); or HEN1 antiserum in the presence of a nonspecific competitor, GST (lane 4). (B) Transfected cells were labeled with either [^{35}S]methionine or $^{32}P_i$. Cell lysates were then precipitated with the preimmune serum (lanes 1) or the HEN1 antiserum (lanes 2). The mobilities of molecular mass standards (in kilodaltons) are indicated to the right of each autoradiogram. The immunoprecipitated HEN1 polypeptides are indicated as pp20.

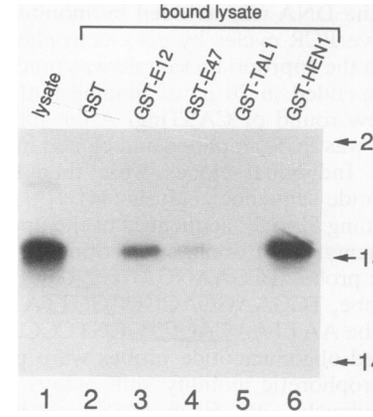

FIG. 4. Protein-protein interactions involving HEN1. Radiolabeled HEN1 polypeptides were generated by in vitro translation in rabbit reticulocyte lysates. An aliquot of the lysate (1 μl) was fractionated by electrophoresis on an SDS–15% polyacrylamide gel (lane 1). Additional aliquots (10 μl each) were incubated with purified GST fusion proteins and glutathione-agarose beads. After extensive washing, the beads were treated with 15 mM glutathione, and the eluants were fractionated by electrophoresis. The binding reactions were conducted with parental GST (lane 2), GST-E12(217-654) (lane 3), GST-E47(524-609) (lane 4), GST-TAL1(166-331) (lane 5), or GST-HEN1(1-133) (lane 6). Equivalent levels of the GST fusion proteins were used for each binding reaction mixture, as confirmed by staining the gel with Coomassie blue prior to autoradiography. The mobilities of molecular mass standards (in kilodaltons) are indicated to the right of the autoradiogram.

modified posttranslationally in Cos1 cells by protein phosphorylation. To determine its phosphoamino acid content, ^{32}P-labeled pp20^{HEN1} was immunoprecipitated from transfected Cos1 cells, fractionated by SDS-polyacrylamide gel electrophoresis, and recovered by electroelution. Acid hydrolysates of ^{32}P-labeled pp20^{HEN1} were then separated by two-dimensional electrophoresis on thin-layer cellulose. As shown in Fig. 3, phosphoserine, but not phosphothreonine or phosphotyrosine, was detected in hydrolysates of pp20^{HEN1} from transfected Cos1 cells.

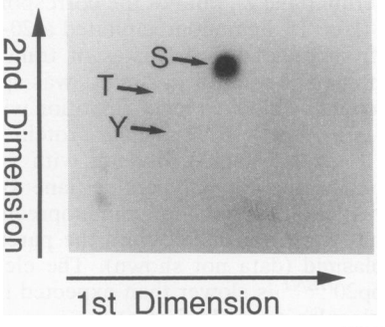

FIG. 3. Phosphoamino acid analysis of pp20^{HEN1} from transfected Cos1 cells. Cos1 cells were transfected with pCMV4/HEN1 and labeled with $^{32}P_i$. ^{32}P-labeled pp20^{HEN1} was then purified from cell lysates by immunoprecipitation and SDS-polyacrylamide gel electrophoresis. Acid hydrolysates of pp20^{HEN1} were prepared and fractionated by electrophoresis on thin-layer cellulose at pH 1.9 (1st dimension) and pH 3.5 (2nd dimension). The positions of unlabeled standards, detected with ninhydrin, are shown: Y, phosphotyrosine; T, phosphothreonine; S, phosphoserine.

HEN1 forms dimers with itself and with the E2A proteins. Tissue-specific class B proteins form bHLH dimers by heteromeric interaction with ubiquitous class A proteins (41). Heterodimer formation appears to be obligatory for both the DNA-binding potential and functional activity of the known class B bHLH proteins. Therefore, a protein-binding assay was used to determine whether the bHLH domain of HEN1 also interacts stably with those of E12 and E47, class A proteins encoded by the *E2A* locus (29, 40). Fusion polypeptides comprising GST linked to the bHLH domains of various proteins, including E12, E47, TAL1, and HEN1, were prepared and loaded onto glutathione-agarose beads. ^{35}S-labeled pp20^{HEN1} was generated by in vitro translation in rabbit reticulocyte lysates, and equivalent aliquots of the lysate were mixed with glutathione-agarose beads loaded with different GST fusion polypeptides. The beads were washed extensively so that retention of radiolabeled HEN1 on the beads would represent specific interaction with the GST fusion polypeptides. The absorbed proteins were then fractionated by electrophoresis on a denaturing polyacrylamide gel, and the presence of radiolabeled HEN1 was determined by autoradiography. As shown in Fig. 4, radiolabeled pp20^{HEN1} is readily absorbed to glutathione-agarose beads in the presence of either the purified GST-E12 or GST-E47 polypeptides (Fig. 4, lanes 3 and 4). Absorption of pp20^{HEN1} is specifically mediated by the E2A moieties of these fusion proteins, since pp20^{HEN1} is not retained in the presence of GST-TAL1 (Fig. 4, lane 5) or the parental GST polypeptide (Fig. 4, lane 2). Thus, HEN1 resembles other class B proteins in its ability to form bHLH complexes by association with class A proteins; this result was confirmed by reciprocal experiments in which in vitro-translated E2A proteins were evaluated for binding to purified GST-HEN1 polypeptides (data not shown). Surpris-

ingly, however, radiolabeled HEN1 polypeptides are also absorbed to the glutathione-agarose beads in the presence of purified GST-HEN1 (Fig. 3, lane 6). Thus, unlike other known class B proteins, HEN1 has the potential to form bHLH homodimers.

Sequence preferences for DNA recognition by HEN1. To evaluate the DNA-binding activity of HEN1, EMSAs were attempted with an oligonucleotide containing μE5 (GAAC CAGAACACCTGCAGCA), a well-characterized E-box element from the immunoglobulin heavy-chain gene enhancer. However, these experiments failed to demonstrate specific binding to μE5 by either HEN1/HEN1 homodimers or HEN1/E2A heterodimers (data not shown). This implied that DNA recognition by HEN1 is restricted to a particular subset of E-box elements. Therefore, we sought to identify HEN1-binding sequences by using the CASTing procedure to screen a pool of degenerate oligonucleotides for high-affinity HEN1-binding sites (51). A HEN1 polypeptide (residues 8 to 133) with an epitope tag (MAYPYDVPDYAGGP) appended to its amino terminus was produced by in vitro translation in reticulocyte lysates. Lysates containing this polypeptide (tag-HEN1) were then incubated with a molar excess of double-stranded oligonucleotides. These oligonucleotides were 75 bp in length, and they shared common priming sites for the PCR that flanked a 35-bp core of complete sequence degeneracy (25). DNA-protein complexes were purified from the mixture by absorption to magnetic beads coated with 12CA5, a monoclonal antibody that specifically recognizes a sequence within the epitope tag (YPYDVPDYA). Bound DNA was released from the beads by heat denaturation, amplified by PCR, and subjected to a further round of CASTing. After six CASTing cycles, the final PCR product was ligated into an M13 cloning vector, and the DNA sequences of 54 individual PCR fragments were determined.

Of the 54 DNA fragments generated by CASTing with tag-HEN1, 50 contained a single E-box sequence (CANNTG) within the 35-bp core; each of the other 4 fragments contained two E-box elements (Fig. 5A). Notably, 49 of the 58 E-box motifs shared a common central sequence (CAGCTG), implying that tag-HEN1 preferentially binds a particular subset of possible E-box elements. Nucleotide sequences encompassing the 58 E-box elements were then aligned, and a consensus sequence for DNA recognition by HEN1 polypeptides was calculated. As illustrated in Fig. 5B, the 18-bp consensus from residues −9 to +9 is almost perfectly palindromic (GGGNCGCAGCTGCGNCCC). This suggests that HEN1 binds DNA as a symmetrical homodimer in which each polypeptide recognizes one half-site of the consensus E box (GGGNCGCAG).

We then examined whether the DNA recognition properties of in vivo-generated HEN1 resemble those of HEN1 polypeptides produced in vitro. Therefore, Cos1 cells were transfected with mammalian expression vectors that did (pCMV4/tag-HEN1) or did not (pCMV4) encode the tag-HEN1 polypeptide described above. The cells were harvested at 48 h after transfection, and whole-cell lysates were prepared. CASTing experiments were then performed in parallel with lysates from either the pCMV4-transfected (i.e., control) or the pCMV4/tag-HEN1-transfected cells. After six cycles of CASTing, the final PCR products were ligated into an M13 cloning vector, and the sequences of individual PCR fragments were determined. Thirteen DNA fragments selected with the control lysate were analyzed; none contained an E-box sequence (data not shown). In contrast, each of the 50 fragments selected with the tag-HEN1 lysate contained at least one E-box element (Fig. 6A). As shown in Fig. 6B, alignment of the 51 E-box elements yields a symmetrical consensus sequence (NGGNCNCA GCTGCGNCCC) very similar to that selected by the in vitro-generated HEN1 homodimers (GGGNCGCAGCTGC GNCCC [Fig. 5B]).

HEN1 binding to the selected consensus sequence requires an intact E-box motif. The DNA-binding potential of HEN1 was confirmed by EMSA experiments. For this purpose, an oligonucleotide based on the consensus binding sequence derived in the CASTing experiments was designed; since the consensus is highly palindromic (GGGNCGCAGCTGCGNC CC), a modified sequence was chosen to prevent intramolecular annealing of the single-stranded oligonucleotide. The chosen sequence for the consensus oligonucleotide (AGGA CGCAGCTGATCCCC) should be suitable, since similar sequences had been selected by CASTing with HEN1 (Fig. 5A). A corresponding mutant oligonucleotide (AGGACG CGGCTTATCCCC) bearing two nucleotide substitutions in the E-box core was also synthesized. Reticulocyte lysates containing the tag-HEN1 polypeptide were then evaluated in an EMSA with a radiolabeled double-stranded probe derived from the consensus oligonucleotide. As shown in Fig. 7, a protein-DNA complex that presumably represents tag-HEN1 homodimers is readily observed upon addition of the tag-HEN1 lysate (lane 2) but not of an unprogrammed lysate (lane 1). Moreover, formation of the radiolabeled complex is inhibited by the presence of increasing quantities of unlabeled consensus probe (lanes 3 to 7). In contrast, the unlabeled mutant probe failed to inhibit formation of the radiolabeled complex (lanes 8 to 12). Therefore, HEN1 binds DNA in a sequence-specific manner that is dependent on the E-box core.

DNA recognition by native HEN1 polypeptides. The CASTing and EMSA experiments described above were performed with a modified polypeptide (tag-HEN1) in which the seven amino-terminal residues of HEN1 were replaced with a synthetic sequence of 14 residues that includes the 12CA5 epitope. Thus, it was necessary to demonstrate that the consensus sequence derived by CASTing is also recognized by wild-type HEN1 polypeptides. Therefore, reticulocyte lysates programmed to produce either tag-HEN1 or wild-type HEN1 were evaluated by EMSA with the radiolabeled consensus probe. As illustrated in Fig. 8, DNA-protein complexes were generated by both the tag-HEN1 (lane 6) and the wild-type HEN1 (lane 2) lysates. As expected, the tag-HEN1 complexes migrate slightly more slowly than the equivalent HEN1 complexes, reflecting the additional mass of the epitope tag. To confirm the protein composition of these complexes, EMSAs were also carried out in the presence of polyclonal rabbit antisera. The mobility of the wild-type HEN1 complexes was greatly retarded (i.e., supershifted) by addition of a HEN1-specific antiserum (Fig. 8, lane 4, SSh) but not the corresponding preimmune serum (lane 3) or an irrelevant antiserum (lane 5). Likewise, the tag-HEN1 complexes were specifically supershifted by either the anti-HEN1 serum (lane 8) or the epitope-specific 12CA5 monoclonal antibody (lane 9). Thus, the protein-DNA complexes illustrated in Fig. 8 represent specific recognition of the consensus HEN1 binding sequence by both tag-HEN1 and wild-type HEN1 polypeptides.

To confirm the DNA-binding properties of in vivo-generated HEN1 polypeptides, parallel EMSAs were conducted with both in vitro-translated HEN1 and nuclear lysates from transfected Cos1 cells. Thus, the Cos1 cells were transfected with either an expression plasmid encoding wild-type HEN1

A

1. GCCAAGGGGACGCAGCTGGGTCCCGCGGGAGGGCT
2. TGACGCAGCTGATCCCACACCTATATATTCCGTGC
3. GAACTGAGATTGGCGCCGGGCACCAGCTGCGCCGC
4. TGACGATCGCTTATCTATTAAAGGCAGCAGCTGCG
5. GAGTGTGGCCGCAGCTGCTAACGCGAATCGTATGT
6. CCACGTGTCAAAAAAGCCGCCAGCTGCTACCTCAG
7. CTCGCGAGCAATACGGGCTCAGCTGCGTCACTGCT
8. GATACTCGGTAGCAGCTGCGTCCCTGATTTAAGTC
9. CATCCTCACATAAGAGGTAACAGCTGCGGCTCCGT
10. ATAATGAACGCAAGGTGGGTCGCAGCTGGACCCCC
11. TCGGATGACTCTGATAATGTGCGCAGCTGGGACCC
12. CGTGTGCGGCATGGCCGCAGCTGCTCCCCAACTTC
13. GCTATTGTGAGGGTAGCAGCTGCGTCTTCTCAGCG
14. ATGGAGCGACAGATTGGGGGCTCAGCTGCGTTCCC
15. GCTCTAATAGGACGCAGCTGGTCCCCATCGGGACA
16. AAGGCTACATGGCACAGGACGCAGCTGCGGCCTCC
17. TTTATGGTGTGCGCATTTGGTCTCAGCTGCGTCCC
18. ATGGCGCAGCTGCGTCACGTCTCCAAGCATCAGGG
19. GTTAAATGGCGCAGCTGCGTCCCCGGCCACAAGGG
20. GTCAATCGTAATGGGTGGATCAGCTGTGGCCCCAT
21. TAAACGTAACAGGGACGCAGCTGCTCCATACCACC
22. TACGGCCCCAGGTTGGTCGCAGCTGCGTCATTTTG
23. CCAGGGGTTGGTAGCAGCTGCGTCCTAGAGTTCTC
24. ATCTTACGGGTGGGAGAGCAGCTGCGGCCCGCCCT
25. ATATGGCGCAGCTGCGTCTTTTGCTGGAAGTTCGA
26. TCGGGGATGGTGTGGGGCGCAGCTGATACCCGTCA
27. ACGTGAAGAGCCGCAGCTGCGCCCTCAAAAAGGAC
28. CTCTGGTTGTTGGGCTCAGCTGCTCCTACTCCGGT
29. GTACCAAGTGTCGAATAAGGACGCAGCTGTGGTCC
30. TTCGCTATGGTCCCAGCTGTGCCCTTCCAAGGTCT
31. ACTTAAGCAGTCCCAGAGGGACGCAGCTGGGTCCG
32. AGTTGTTATTTCCTAGAGGACGCAGATGTGGCCCC
33. CGGGTAGCAGCTGGGTCCGAATCGATACGCTTGCT
34. GCACGCGGCACGGGGCCGCAGCTGCAGGGATTCTT
35. GATAAATGGGCCACAGCTGCGCCCGCAGGTAGAAC
36. CGGAACCACGGGCATCAGCTGCGGTTATTGGGCCC
37. AATTGGGGCCGCAGCTGCTCCCCCATCCTTTGCTC
38. ATTGAATTATGAACTATGGAGACGCAGCTGCGGCC
39. AGGGGCTCAGGTGAGCCCCAACTGAAGGCGGCTCC
40. GATCGCGGGAGCAGCTGCTACCCAGAATATGGCGC
41. TGCAGGGACGCAGGTGCTCCTCTTTCTCTTTCGAC
42. TACACGTTGGTCGCAGCTGCTGGGATTTCGACGGC
43. CGAGTTAGTTGGGCACAGCTGCGCCCCTGTGTTAG
44. CAGGGGGCGCAGCTGTGCCACCCTTTTTGAAGGCA
45. CTAAAATCGATTCGGGGCTCAGCTGCGTCCCTGCC
46. TGCGTACCCCAGCTGCGTCCCCGACTATGACGGGG
47. CACTGCTAGGGACGCAGCTGGTCTCCATTATTGTA
48. ATCCGATCCAGCCATCCGTGGGATCAGCTGCTGCC
49. GTGGCGCACCTGCGTCACCTTGGGCCATTCTGCTC
50. AAGTATGGGGGCCGCAGGTGCGTCCCCTAACTCGG
51. TCACGCGCATGAAAAGGGGCCGCAGCTGCGCTCAC
52. GTGACGCAGCTGCGTCCTAGCCAAGAGTAATAGTC
53. GCCTGGGAAGGTTTGACACCAGCTGCGTCACTTGC
54. CGCAGCTGTGTCCCTAGAAACCTTGTCCCCTTACG

FIG. 5. Preferred sequences for DNA binding by HEN1. (A) The nucleotide sequences of 54 individual DNA fragments selected by CASTing with in vitro-translated tag-HEN1 polypeptides. Potential E-box elements are underlined. (B) Calculation of a consensus sequence for optimal DNA recognition by HEN1. The 58 E-box sequences obtained by CASTing were aligned from positions −11 to +11. The consensus sequence was determined by calculating the percent frequency of each nucleotide at each position. A frequency of 50% or greater was arbitrarily chosen as a meaningful bias towards a given nucleotide at a particular position.

B

E-box core

position	-11	-10	-9	-8	-7	-6	-5	-4	-3	-2	-1	1	2	3	4	5	6	7	8	9	10	11
%C	16	12	7	0	2	29	72	14	100	0	5	88	0	0	67	3	30	80	68	57	41	16
%A	33	29	19	5	5	21	25	5	0	100	3	2	0	0	7	5	11	5	16	13	22	27
%G	27	31	52	78	91	30	4	64	0	0	91	9	0	100	12	67	23	5	5	9	10	24
%T	24	27	22	16	2	20	0	17	0	0	0	2	100	0	14	25	36	9	11	20	26	33
consensus			G	G	G	N	C	G	C	A	G	C	T	G	C	G	N	C	C	C		

(pCMV4/HEN1) or a corresponding control plasmid (pCMV4); 48 h after transfection, nuclear lysates were prepared for use in EMSAs with the radiolabeled consensus oligonucleotide. As illustrated in Fig. 9, protein-DNA complexes were generated upon addition of nuclear lysates from pCMV4/HEN1-transfected cells (lane 6) but not from pCMV4-transfected cells (lane 9). The electrophoretic mobility of these complexes is similar to those formed with in vitro-translated HEN1 (lane 3). The radiolabeled complexes generated by in vivo-produced HEN1 (lane 6) are inhibited by a 400-fold excess of the unlabeled oligonucleotide probe (lane 7), and they are supershifted by the HEN1-specific antiserum (lane 5). Thus, the DNA-binding properties of in vivo-produced HEN1 polypeptides (lanes 5 to 7) resemble those of in vitro-translated HEN1 (lanes 2 to 4).

DNA recognition by HEN1/E2A heterodimers. Protein binding assays indicate that HEN1 is capable of both homodimer formation by self-association and heterodimer formation by interaction with either E12 or E47 (Fig. 4). Previous studies had shown that the DNA recognition sequence for bHLH dimers (i.e., the E-box motif) is composed of two half-sites, each of which is bound predominantly by one of the two dimerized bHLH polypeptides (9, 20). CASTing with in vitro-translated HEN1 selects for a symmetric E-box consensus sequence that represents the preferred binding site for HEN1 homodimers (Fig. 5 to 8); as discussed below, this sequence is composed of two identical HEN1 half-sites (GGGNCGCAG). To determine whether HEN1 heterodimers also have DNA-binding potential, additional EMSAs were performed with the consensus oligonucleotide probe described above, as well as a hybrid oligonucleotide probe (GGGACGCAGCTGTTAATT) composed of one HEN1 half-site paired with a half-site preferentially bound by E2A polypeptides (AACAG) (9, 31, 46). As illustrated in Fig. 10, the HEN1 consensus probe is recognized by HEN1 homodimers and HEN1/E12 heterodimers but not by E12 homodimers (lanes 1 to 3). Conversely, the hybrid probe is bound by E12 homodimers and HEN1/E12 heterodimers but not by HEN1 homodimers (lanes 4 to 6). It is somewhat surprising that the hybrid probe is readily recognized by

A

1. CAGTTTGCTTCGGTCTCAGCTGCGCTCCCGCCATA
2. TAGCAGCTGCGACTCCAAGACCAACAATCTCCTTG
3. CTGACTAACCTAAGCATGACGCAGGTGCGCCCCAA
4. ACAAAGAGGTCCCATATGTGGCCTTGATGTGCCTT
5. TCCCAACTCTCCAGTGGACGCAGATGATGCTCGGT
6. AGCATTAGGTGTGGGTAGCAGGTGGTGCCCCGGCT
7. GTAACCGATACACCCATGGGATCAGCTGTGGCTAT
8. TTATATGGGGTATTGCGGGCCCATATGCGTCCCTA
9. CTGTGGGCCCCAGGTGCGGCCCCTCCCAGGACACC
10. TCCCAAATGCGCAGTGGGCCACAGCTGGTGCCCGC
11. AGTCCCCGATTTCCTTGGTCTCAGCTGTGCTCGCG
12. ATTGGAACGCTAATTTGGCCCCATATGCGTCCCGT
13. CTCATTGGGCCTCGGTTAGGGATCAGCTGGTGCAC
14. CGCCAGATACTTTGCAGGTCGCAGCTGCTCCCATC
15. GTCACAGCTGTGCCCTGCAATAGGAAACCGTCTTA
16. GGAGGGGAGCTGTGGAGCAGGTGCGTCACTTCGGA
17. CTTTTTTGACGGCCGCAGATGCGACCCACGATGCG
18. TATTTGGTGCTCAGCTGTGACTCATTGCGGTACCT
19. TGGCGCAGCTGGGACTCCGATTAGTGGAAGCACAC
20. TTGCACCAGCTGCGTCCTGCCATTATGTACGTGTT
21. GGCACATCTGCGTCCGTTTCTTCGTAAGTGATCCA
22. TCAGCTGCTCCCATCCAACCGGCAAGGGATCATCA
23. TCGCTCAGGTCCCACCTGCGTCCCTTTGTCAGCGT
24. GAGGATACTGAGGGGACGCAGCTGGTGTCTGACCC
25. TGATATTCGTGGCGAAATGGGGCTCAGCTGCGCCC
26. GATCGCAGTTCCCTGGGGCCCAGCTGCGTCCCTAT
27. AAAGGACGCACATGGTCGAGCGTCCGCATCCTCCG
28. CAACATTTTGAAAGTACGCAGCTGCGTCCCTCGTT
29. GCTCAGCTGGTGTCACGTCTTGAAAAAACAAGCCG
30. GCGCAGCTGCCACACTAGTGATACGCTTTTGCGCC
31. TTGTGGTCTCAGCTGGGGTCACCAGCGTAGTCAAT
32. TGTTATACCATGGCTCTCCGGGCTCAGCTGATCTC
33. GTCCGCAGCTGCGTCACCTCATTACACTGCCATTG
34. TGACGATATGGGGCTCAGCTGGTACCATCGTCATT
35. TTCGATTCGCGAGTTAGACCCCAGGTGCGTCCCTA
36. TCTCTGAGGGGTCCCACCTGAGACCCTTGTGGTTG
37. GGTAGGGGCTCAGGTGCGGCTAGTATTAAATGAGC
38. CGCAGCTGAGACGTGGACGGGGTAGGCAATTTTTT
39. CCACATGGGACCCAGCTGAGATCCCTGCTCCAGAT
40. AGTATCGTAGCACCAGCTGCTACCCATTGCGTACC
41. GATGACGGGACCCAGCTGCGGCTCAAACCAACTCC
42. ACTAAACCGATGGGGCACAGCTGCGTCACTCGTAG
43. GTGAAGAGTGCTCGCCCGGGCACAGCTGCTCCCCA
44. TCTAATGGCTTGACGACGGGGCACCAGCTGAGGTC
45. TCCGCGCGCCACGTACGAAAGTAGCAGCTGCGCTG
46. CGCACCTGCTCCCTCTTGTNACTCTTGCCGTTCTC
47. AAGAGTTACCCTCAAGATCAGCTGCGCCCCCCCTG
48. TCCACTCAATAGTACAGTATCAGCTGTGACCATCT
49. GGTTACCAGCTGCGTCCCCCAAATCAACATTTTGC
50. GTGGGACGCAGCTGCGTGCAAAAGTCTCACTTCCC

FIG. 6. CASTing with in vivo-produced HEN1 polypeptides. (A) The sequences of 51 DNA fragments selected by CASTing with lysates from Cos1 cells transfected with the pCMV4/tag-HEN1 expression plasmid. Potential E-box elements are underlined. (B). The 52 E-box sequences obtained by CASTing were aligned from positions −11 to +11. A consensus sequence for preferential recognition by in vivo-generated tag-HEN1 was calculated as described in the legend to Fig. 5.

B

E-box core

position	-11	-10	-9	-8	-7	-6	-5	-4	-3	-2	-1	1	2	3	4	5	6	7	8	9	10	11
%C	21	18	17	0	0	20	78	29	100	0	8	75	0	0	58	2	29	77	71	58	30	30
%A	31	18	22	7	7	18	21	11	0	100	2	13	0	0	12	0	21	2	12	19	15	20
%G	18	21	44	84	82	35	0	33	0	0	83	12	0	100	17	71	25	4	4	8	20	23
%T	31	44	17	9	11	27	0	27	0	0	8	0	100	0	13	27	25	17	14	15	35	27
consensus			N	G	G	N	C	N	C	A	G	C	T	G	C	G	N	C	C	C		

E12/E12 homodimers; perhaps the presence of a single E2A half-site is sufficient for DNA binding by E2A homodimers. In any case, it is clear that HEN1 polypeptides, like the *E2A* gene products, have the potential to bind DNA as either homodimers or as heterodimers involving other bHLH proteins. Moreover, these homomeric and heteromeric bHLH complexes exhibit distinct, but overlapping, DNA-binding specificities which reflect the half-site preferences of their constituent polypeptides.

DISCUSSION

HEN1 and HEN2 constitute a subgroup of bHLH proteins that share exceptional amino acid homology over a 74-residue stretch that includes the bHLH domain (5, 11, 26). Transcripts encoding these proteins are found primarily in the subependymal layer of the neuroepithelium in day 9.5 to 14.5 embryonic mice (5). Thus, expression of HEN1 and HEN2 is highly restricted with respect to both tissue distribution and temporal pattern, suggesting a potential role for these proteins in the early development of mammalian nervous tissue. This is consistent with the observation that *HEN1* gene transcription is activated upon NGF-induced neuronal differentiation of PC-12 cells.

Similar functions have been suggested for MASH1 and MASH2, a distinct subgroup of neuron-specific bHLH proteins that represent mammalian homologs of the *Drosophila achaete-scute* gene products (33). MASH1 is expressed in embryonic, but not adult, rat brain tissue. In particular, its expression is restricted to neuroepithelial precursors in the proliferative region of the hindbrain ventricular zone (38). MASH1 is also expressed in a positionally restricted manner in the forebrain and in postmigratory neural crest cells of the peripheral nervous system. Hence, the spatial expression of MASH1 partially overlaps HEN1 expression at an equivalent time point in neural development.

Several tissue-specific bHLH proteins have been shown to regulate decisions regarding cell fate within their respective developmental lineages. Most notably, the mammalian myogenic factors (e.g., MyoD) and the *Drosophila achaete-scute* gene products are required for cell type determination during myogenesis and neurogenesis, respectively (2, 24, 42, 48, 50). The functions of these proteins are expressed upon obligate interaction with ubiquitous class A bHLH polypep-

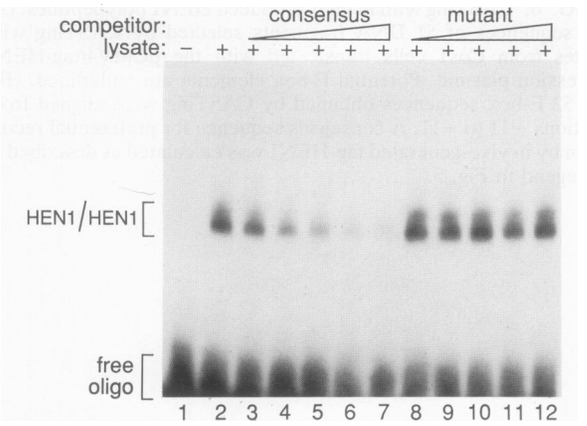

FIG. 7. The sequence specificity of DNA binding by in vitro-translated HEN1 polypeptides. A double-stranded ^{32}P-labeled probe was prepared from oligonucleotides representing the consensus sequence for HEN1 recognition (top strand: AGGACGCAGCT GATCCCC). The probe was incubated with either an unprogrammed reticulocyte lysate (lane 1) or lysates containing the tag-HEN1 polypeptide (lanes 2 to 12). After electrophoresis on a native 5% polyacrylamide gel, the radiolabeled probe was detected by autoradiography. Free oligonucleotides migrated rapidly toward the anode. Protein-oligonucleotide complexes involving HEN1/HEN1 homodimers are designated with a bracket. For some binding reactions, the tag-HEN1 lysates were preincubated with a 10-fold (lanes 3 and 8), 50-fold (lanes 4 and 9), 100-fold (lanes 5 and 10), 200-fold (lanes 6 and 11), or 400-fold (lanes 7 and 12) excess of unlabeled oligonucleotides; the competitor oligonucleotides had either the consensus sequence for HEN1 recognition (lanes 3 to 7) or a corresponding mutant sequence (lanes 8 to 12).

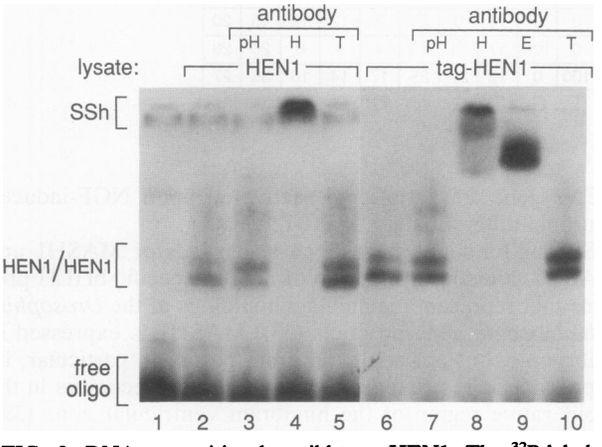

FIG. 8. DNA recognition by wild-type HEN1. The ^{32}P-labeled consensus probe was incubated with an unprogrammed reticulocyte lysate (lane 1), lysates containing wild-type HEN1 polypeptides (lanes 2 to 5), or lysates containing tag-HEN1 polypeptides (lanes 6 to 10). Some binding reaction mixtures were supplemented with a HEN1-specific rabbit antiserum (lanes 4 and 8), the corresponding preimmune serum (lanes 3 and 7), an irrelevant antiserum (anti-TAL1 serum [lanes 5 and 10]) or the epitope-specific 12CA5 monoclonal antibody (lane 9). The binding reactions were fractionated by electrophoresis on a native 5% polyacrylamide gel, and the radiolabeled probe was detected by autoradiography. Protein-oligonucleotide complexes involving HEN1/HEN1 homodimers are indicated with a bracket. Supershifted complexes generated by antibody recognition are denoted as SSh.

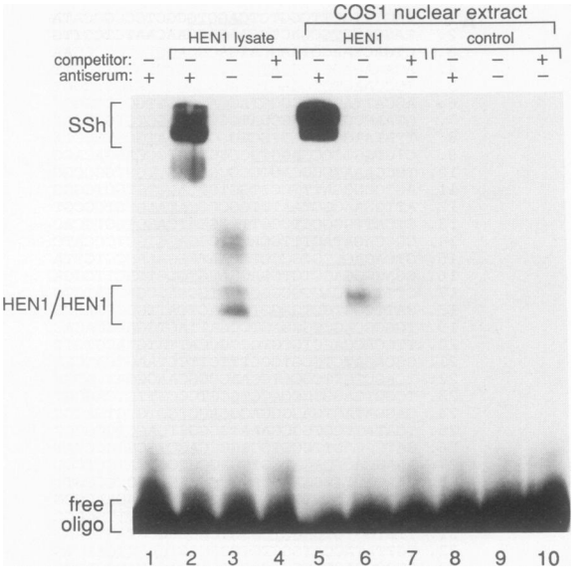

FIG. 9. DNA binding by in vivo-generated HEN1 polypeptides. The ^{32}P-labeled consensus probe was incubated with an unprogrammed reticulocyte lysate (lane 1), lysates containing wild-type HEN1 polypeptides (lanes 2 to 4), or nuclear extracts from Cos1 cells transfected with either pCMV4/HEN1 (lanes 5 to 7) or the control pCMV4 plasmid (lanes 8 to 10). Some binding reaction mixtures were supplemented with a 400-fold excess of the unlabeled consensus probe (lanes 4, 7, and 10) or the HEN1-specific antiserum (lanes 1, 2, 5, and 8). The binding reactions were then evaluated by EMSA. Protein-oligonucleotide complexes involving HEN1/HEN1 homodimers are indicated with brackets, and the supershifted complexes are denoted as SSh.

tides such as E12 and E47 (37). The resultant bHLH heterodimers (e.g., MyoD/E12) presumably influence cell fate by regulating the transcription of particular subsets of subordinate genes. The participation of class A proteins is required at least in part because the known class B polypeptides do not self-associate to form homodimers under physiological conditions (35, 41). Thus, the DNA-binding potential of the myogenic bHLH factors, the *achaete-scute* gene products, and their mammalian homologs, MASH1 and MASH2, is expressed upon heterodimer formation with class A bHLH proteins (12, 34, 41, 47). Similarly, DNA recognition by the leukemic TAL1 polypeptides is also contingent upon heterodimer formation with class A proteins (30, 31). Thus, we anticipated that the neuron-specific HEN polypeptides would likewise require interaction with class A proteins for full expression of their DNA-binding potential. Our results show that HEN1 does in fact associate in vitro with either E12 or E47 to form heterodimeric complexes with DNA-binding activity. However, under both in vitro and in vivo conditions, HEN1 polypeptides also form homodimers that recognize DNA in a sequence-specific manner. In this respect HEN1 is clearly distinct from other known class B proteins. It remains to be established whether HEN1 homodimers and HEN1 heterodimers regulate the transcription of distinct subsets of subordinate target genes and thereby potentially mediate unique functions during neural development. It is conceivable that either homomeric or heteromeric HEN1 complexes predominate in different cellular settings. This could be achieved by mechanisms that regulate the availability of class A proteins, including differential expression of the Id polypeptides—a class of regula-

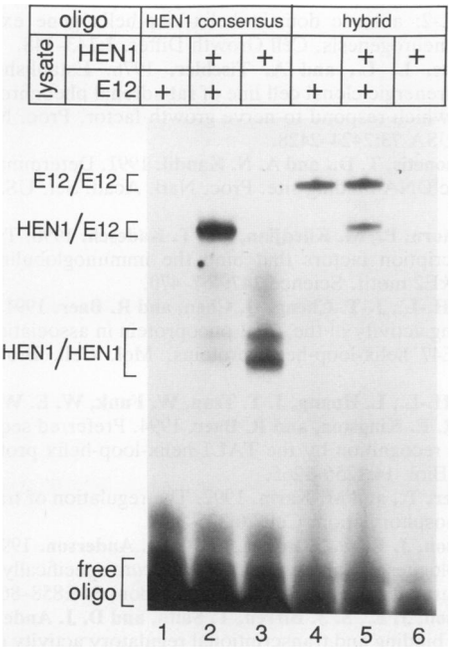

FIG. 10. DNA recognition by HEN1/E12 heterodimers. EMSAs were conducted with either the ^{32}P-labeled consensus probe (lanes 1 to 3) or the hybrid probe (lanes 4 to 6). These probes were incubated with reticulocyte lysates containing E12 polypeptides alone (lanes 1 and 4), HEN1 polypeptides alone (lanes 3 and 6), or both E12 and HEN1 (lanes 2 and 5). Protein-oligonucleotide complexes involving E12/E12 homodimers, HEN1/E12 heterodimers, and HEN1/HEN1 homodimers are indicated with brackets; the protein composition of each complex was confirmed in separate EMSA experiments by supershift analysis with E2A-specific and HEN1-specific rabbit antisera (data not shown).

tory helix-loop-helix factors that suppress the activity of class A proteins by sequestration (6).

The CASTing experiments yielded a consensus sequence for DNA recognition by HEN1 (GGGNCGCAGCTGCG NCCC) that includes the core E-box element CAGCTG. The palindromic nature of this consensus is consistent with the notion that DNA binding was mediated by homodimeric HEN1 protein complexes. Dang et al. (16) have identified an amino acid position within the basic region of bHLH proteins that is a major determinant of sequence recognition at the inner residues of the E-box core. Thus, bHLH polypeptides with arginine at this position preferentially bind CACGTG (e.g., c-MYC, USF, CBF1, MAX, TFE3, and PHO4), whereas those with a corresponding aliphatic residue (methionine, leucine, or valine) bind optimally to CAGCTG. In accord with this scheme, HEN1 polypeptides, which bear valine at the appropriate position (HEN1 residue 88), preferentially recognize E-box sequences with the CAGCTG core. Other lines of evidence suggest that the specificity of DNA binding by different bHLH proteins is also influenced by flanking nucleotides outside the E-box core (23, 28, 43). It is noteworthy, therefore, that clear preferences for HEN1 binding, as represented by the derived consensus sequence, were observed throughout six flanking residues on either side of the CAGCTG core. The extended length of this consensus should impart HEN1 homodimers with sequence specificities distinct from those of most other bHLH complexes.

Blackwell and Weintraub (9) have proposed that the E-box motif can be viewed as a combination of two half-sites, each of which interacts with one of the two polypeptides that compose the bHLH dimer. This notion, supported by a recent crystallographic study of protein-DNA complexes (20), implies that each bHLH polypeptide will preferentially bind particular half-site sequences. Several bHLH proteins have now been analyzed by random binding site selection methods analogous to the CASTing procedure. The E-box half-site recognized optimally by HEN1 (GGGNCGCAG) is distinct from those determined for the other bHLH proteins, including the myogenic factors MyoD and myogenin (AANAACAG) (9, 51), the MYC oncoproteins (CAC) (1, 8), the leukemic TAL1 polypeptide (ACCAT) (31), and class A proteins encoded by the *E2A* locus (AACAC or AACAG) (9, 31, 46). Therefore, if HEN1 does indeed function in vivo as a transcription factor, then the bHLH dimers involving HEN1 are likely to bind, and possibly regulate the expression of, a unique subset of subordinate target genes. These target genes presumably mediate functional properties of HEN1 during neural development, and their identification should be facilitated by the preferred HEN1 binding sequences defined in this study. Preliminary experiments with reporter constructs bearing multiple copies of the preferred HEN1 binding sequence have not revealed transcriptional activation by the HEN1 polypeptide. Thus, the regulatory function of HEN1 may require the cooperation of other, as yet unknown, *trans*-acting factors.

ACKNOWLEDGMENTS

We are grateful to Melanie Cobb for providing untreated and NGF-treated PC-12 cells.

This work was supported by a grant from the National Cancer Institute (CA46593). R.B. is a member of the Simmons Arthritis Research Center and the recipient of a Faculty Research Award from the American Cancer Society (FRA-421).

ADDENDUM IN PROOF

The HUGO Nomenclature Committee has adopted the designations *NHLH1* and *NHLH2* as gene symbols to replace *HEN1/NSCL1* and *HEN2/NSCL2*, respectively.

REFERENCES

1. **Alex, R., O. Sözeri, S. Meyer, and R. Dildrop.** 1992. Determination of the DNA sequence recognized by the bHLH-zip domain of N-Myc protein. Nucleic Acids Res. **20**:2257–2263.
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. **Andersson, S., D. L. Davis, H. Dahlback, H. Jornvall, and D. W. Russell.** 1989. Cloning, structure and expression of the mitochondrial cytochrome P-450 sterol 26-hydroxylase, a bile acid biosynthetic enzyme. J. Biol. Chem. **264**:8222–8229.
4. **Begley, C. G., P. D. Aplan, S. M. Denning, B. F. Haynes, T. A. Waldman, and I. R. Kirsch.** 1989. The gene SCL is expressed during early hematopoiesis and encodes a differentiation-related DNA-binding motif. Proc. Natl. Acad. Sci. USA **86**:10128–10132.
5. **Begley, C. G., S. Lipkowitz, V. Gobel, K. A. Mahon, V. Bertness, A. R. Green, N. M. Gough, and I. R. Kirsch.** 1992. Molecular characterization of *NSCL*, a gene encoding a helix-loop-helix protein expressed in the developing nervous system. Proc. Natl. Acad. Sci. USA **89**:38–42.
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. **Bernard, O., N. Lecointe, P. Jonveaux, M. Souyri, M. Mauchauffe, R. Berger, C. J. Larsen, and D. Mathieu-Mahul.** 1991. Two site-specific deletions and the t(1;14) translocation restricted to human T-cell acute leukemias disrupt the 5′ part of the *tal-1* gene. Oncogene **6**:1477–1488.
8. **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.
9. **Blackwell, T. K., and H. Weintraub.** 1990. Differences and similarities in DNA-binding preferences of MyoD and E2A protein complexes revealed by binding site selection. Science **250**:1104–1110.
10. **Boyle, W. J., P. van der Geer, and T. Hunter.** 1991. Phosphopeptide mapping and phosphoamino acid analysis by two-dimensional separation on thin-layer cellulose plates. Methods Enzymol. **210**:110–149.
11. **Brown, L., R. Epsinosa III, M. M. Le Beau, M. J. Siciliano, and R. Baer.** 1992. *HEN1* and *HEN2*: a subgroup of basic helix-loop-helix genes that are coexpressed in a human neuroblastoma. Proc. Natl. Acad. Sci. USA **89**:8492–8496.
12. **Cabrera, C. V., and M. C. Alonso.** 1991. Transcriptional activation by heterodimers of the *achaete-scute* and *daughterless* gene products of *Drosophila*. EMBO J. **10**:2965–2973.
13. **Chen, Q., J.-T. Cheng, L.-H. Tsai, N. Schneider, G. Buchanan, A. Carroll, W. Crist, B. Ozanne, M. Siciliano, and R. Baer.** 1990. The *tal* gene undergoes chromosome translocation in a T-cell leukemia and potentially encodes a helix-loop-helix protein. EMBO J. **9**:415–424.
14. **Cheng, J.-T., H.-L. Hsu, L.-Y. Hwang, and R. Baer.** 1993. Products of the *TAL1* oncogene: basic helix-loop-helix proteins phosphorylated at serine residues. Oncogene **8**:677–683.
15. **Church, G. M., A. Ephrussi, W. Gilbert, and S. Tonegawa.** 1985. Cell type specific contacts to immunoglobulin enhancers in nuclei. Nature (London) **313**:798–801.
16. **Dang, C. V., C. Dolde, M. L. Gillison, and G. J. Kato.** 1992. Discrimination between related DNA sites by a single amino acid residue of Myc-related basic-helix-loop-helix proteins. Proc. Natl. Acad. Sci. USA **89**:599–602.
17. **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.
18. **Dignam, J. D., R. M. Lebovitz, and R. G. Roeder.** 1983. Accurate transcription initiation by RNA polymerase II in a soluble extract from isolated mammalian nuclei. Nucleic Acids Res. **11**:1475–1489.
19. **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.
20. **Ferré-D'Amaré, A. R., G. C. Prendergast, E. B. Ziff, and S. K. Burley.** 1993. Recognition by Max of its cognate DNA through a dimeric b/HLH/Z domain. Nature (London) **363**:38–45.
21. **Field, J., J.-I. Nikawa, D. Broek, B. MacDonald, L. Rodgers, I. A. Wilson, R. A. Lerner, and M. Wigler.** 1988. Purification of a *RAS*-responsive adenyl cyclase complex from *Saccharomyces cerevisiae* by use of an epitope addition method. Mol. Cell. Biol. **8**:2159–2165.
22. **Finger, L. R., J. Kagan, G. Christopher, J. Kurtzberg, M. S. Hershfield, P. C. Nowell, and C. M. Croce.** 1989. Involvement of the *TCL5* gene on human chromosome 1 in T-cell leukemia and melanoma. Proc. Natl. Acad. Sci. USA **86**:5039–5043.
23. **Fisher, F., and C. R. Goding.** 1992. Single amino acid substitutions alter helix-loop-helix protein specificity for bases flanking the core CANNTG motif. EMBO J. **11**:4103–4109.
24. **Funk, W. D., M. Ouellette, and W. E. Wright.** 1991. Molecular biology of myogenic regulatory factors. Mol. Cell. Biol. **8**:185–195.
25. **Funk, W. D., and W. E. Wright.** 1992. Cyclic amplification and selection of targets for multicomponent complexes: myogenin interacts with factors recognizing binding sites for basic helix-loop-helix, nuclear factor 1, myocyte-specific enhancer-binding factor 2 and COMP1 factor. Proc. Natl. Acad. Sci. USA **89**:9484–9488.
26. **Göbel, V., S. Lipkowitz, C. A. Kozak, and I. R. Kirsch.** 1992. *NSCL-2*: a basic domain helix-loop-helix gene expressed in early neurogenesis. Cell Growth Differ. **3**:143–148.
27. **Greene, L. L., and A. Tischler.** 1976. Establishment of a noradrenergic clonal cell line of rat adrenal pheochromocytoma cells which respond to nerve growth factor. Proc. Natl. Acad. Sci. USA **73**:2424–2428.
28. **Halazonetis, T. D., and A. N. Kandil.** 1991. Determination of the c-Myc DNA-binding site. Proc. Natl. Acad. Sci. USA **88**:6162–6166.
29. **Henthorn, P., M. Kiledjian, and T. Kadesch.** 1990. Two distinct transcription factors that bind the immunoglobulin enhancer *m*E5*lk*E2 motif. Science **247**:467–470.
30. **Hsu, H.-L., J.-T. Cheng, Q. Chen, and R. Baer.** 1991. Enhancer binding activity of the *Tal-1* oncoprotein in association with the E12/E47 helix-loop-helix proteins. Mol. Cell. Biol. **11**:3037–3042.
31. **Hsu, H.-L., L. Huang, J. T. Tsan, W. Funk, W. E. Wright, J.-S. Hu, R. E. Kingston, and R. Baer.** 1994. Preferred sequences for DNA recognition by the TAL1 helix-loop-helix proteins. Mol. Cell. Biol. **14**:1256–1265.
32. **Hunter, T., and M. Karin.** 1992. The regulation of transcription by phosphorylation. Cell **70**:375–387.
33. **Johnson, J. E., S. J. Birren, and D. J. Anderson.** 1990. Two rat homologues of *Drosophila achaete-scute* specifically expressed in neuronal precursors. Nature (London) **346**:858–861.
34. **Johnson, J. E., S. J. Birren, T. Saito, and D. J. Anderson.** 1992. DNA binding and transcriptional regulatory activity of mammalian achaete-scute homologous (MASH) proteins revealed by interaction with a muscle-specific enhancer. Proc. Natl. Acad. Sci. USA **89**:3596–3600.
35. **Kadesch, T.** 1993. Consequences of heteromeric interactions among helix-loop-helix proteins. Cell Growth Differ. **4**:49–55.
36. **Laemmli, U. K.** 1970. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature (London) **227**:680–685.
37. **Lassar, A. B., R. L. Davis, W. E. Wright, T. Kadesch, C. Murre, A. Voronova, D. Baltimore, and H. Weintraub.** 1991. Functional activity of myogenic HLH proteins requires hetero-oligomerization with E12/E47-like proteins *in vivo*. Cell **66**:305–315.
38. **Lo, L.-C., J. E. Johnson, C. W. Wuenschell, T. Saito, and D. J. Anderson.** 1991. Mammalian *achaete-scute* homolog 1 is transiently expressed by spatially restricted subsets of early neuroepithelial and neural crest cells. Genes Dev. **5**:1524–1537.
39. **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.
40. **Murre, C., P. Schonleber 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.
41. **Murre, C., P. Schonleber 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.
42. **Olson, E. N.** 1990. MyoD family: a paradigm for development? Genes Dev. **4**:1454–1461.
43. **Prendergast, G. C., and E. B. Ziff.** 1991. Methylation-sensitive sequence-specific DNA binding by the c-Myc basic region. Science **251**:186–189.
44. **Sanger, F., A. R. Coulson, B. G. Barrell, A. J. H. Smith, and B. A. Roe.** 1980. Cloning in single-stranded bacteriophage as an aid to rapid DNA sequencing. J. Mol. Biol. **143**:161–178.
45. **Smith, D. B., and K. S. Johnson.** 1988. Single-step purification of polypeptides expressed in *Escherichia coli* as fusion proteins with glutathione *S*-transferase. Gene **67**:31–40.
46. **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.
47. **Van Doren, M., H. M. Ellis, and J. W. Posakony.** 1991. The *Drosophila extramacrochaetae* protein antagonizes sequence-

specific DNA binding by *daughterless/achaete-scute* protein complexes. Development **113**:245–255.

48. **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.
49. **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.
50. **Weintraub, H., R. Davis, S. Tapscott, M. Thayer, M. Krause, R. Benezra, T. K. Blackwell, D. Turner, R. Rupp, S. Hollenburg, Y. Zhuang, and A. Lassar.** 1991. The MyoD gene family: nodal point during specification of the muscle cell lineage. Science **251**:761–766.
51. **Wright, W. E., M. Binder, and W. Funk.** 1991. Cyclic amplification and selection of targets (CASTing) for the myogenin consensus binding site. Mol. Cell. Biol. **11**:4104–4110.
52. **Xia, Y., L. Brown, C. Y.-C. Yang, J. T. Tsan, M. J. Siciliano, R. Espinosa III, M. M. Le Beau, and R. J. Baer.** 1991. *TAL2*, a helix-loop-helix gene activated by the (7;9)(q34;q32) translocation in human T-cell leukemia. Proc. Natl. Acad. Sci. USA **88**:11416–11420.