A human Id-like helix–loop–helix protein expressed during early development

(transcription regulation/nervous system development)

JOSEPH BIGGS, ELIZABETH V. MURPHY, AND MARK A. ISRAEL

Preuss Laboratory for Molecular Neuro-oncology, Brain Tumor Research Center, Departments of Neurological Surgery and Pediatrics, School of Medicine, University of California, Box 0520, San Francisco, CA 94143

Communicated by Marc Kirschner, November 4, 1991 (received for review September 26, 1991)

ABSTRACT The interaction of helix–loop–helix (HLH) proteins is known to regulate the differentiation of several different tissues, including mammalian muscle and the insect peripheral nervous system. In myoblasts, the products of myogenic HLH genes such as MyoD and ubiquitous HLH proteins such as E12 are present at constant levels throughout development. An E12 monomer and a MyoD monomer form a DNA binding heterodimer that activates muscle-specific genes. These two proteins are unable to dimerize in proliferating myoblasts because a negative regulator HLH protein, Id, is present. We now report the sequence and structure of a human HLH gene related to Id, which has been designated Id-2. Two prominent Id-2 RNA molecules of 2.5 and 1.3 kilobases were found in a number of different human normal and neoplastic tissues. We believe the larger RNA is a precursor of the 1.3-kilobase mRNA that encodes an Id-2 protein of 134 amino acids. The HLH region of the Id-2 protein is 90% homologous to that of myogenic Id, but the homology is most extensive outside the HLH region. The Id-2 gene is highly expressed during early fetal development in several tissues, including those of the central nervous system, but is not expressed in the corresponding mature tissues. Id-2 expression is modulated in association with retinoic acid-induced ganglionic differentiation of the neuroblastoma cell line SMS-KCN. These findings suggest that Id-2 is an inhibitor of tissue-specific gene expression, although its distinctive pattern of expression during development suggests a role different from that of Id.

Transcription factors containing the helix–loop–helix (HLH) motif regulate the expression of tissue-specific genes in a number of mammalian and insect tissues, including the insect peripheral nervous system (PNS) (1). During development of the PNS in Drosophila, HLH genes such as daughterless and the achete/scute genes are required for neurogenesis (2). The activity of these genes is apparently opposed by another HLH gene, extramacrochaetae (emc), which may function by forming heterodimers with the achete and scute genes (3). Biochemical evidence for the inhibition of gene expression by such a mechanism is best defined in studies of mammalian myogenesis. During the proliferation and differentiation of myoblasts, the products of myogenic HLH genes such as MyoD (4) and ubiquitous HLH proteins such as E12 (5) are present at constant levels. During differentiation, an E12 monomer and a MyoD monomer form a DNA binding heterodimer that activates muscle-specific genes such as muscle creatine kinase (6).

E12 and MyoD are unable to dimerize in proliferating myoblasts because of the presence of the HLH protein Id, a suppressor of myogenic differentiation (7). Id dimerizes with the E12 protein but the Id protein lacks a region rich in basic amino acids that is necessary to form the DNA binding domain of the HLH heterodimer. By dimerizing with the E12 HLH domain, Id blocks the interaction of E12 and MyoD, inhibiting expression of muscle-specific genes. Another HLH protein isolated from mouse 3T3 cells, HLH 462, also lacks a basic region required for DNA binding, and it too blocks DNA binding of known HLH transcriptional regulators (8). Similarly, the emc gene of Drosophila lacks such a basic DNA binding domain, providing yet additional evidence that the model proposed for regulation of muscle-specific gene expression will be paralleled in other tissues (4). We have sought Id-like proteins that would be candidates for regulating the expression of lineage-specific genes in mammalian tissues other than muscle. Because mammalian homologues of the achete complex are known to be HLH proteins that may function as positive regulators of gene expression in neural crest precursors of the rat PNS, we were particularly interested in examining nervous system tissues for the presence of negative regulators (9). We now report the DNA sequence and structure of a human HLH gene closely related to Id, designated Id-2. Id-2 contains a HLH domain that is 90% homologous to the HLH domain of Id.

MATERIALS AND METHODS

Cell Culture. Neuroblastoma and glioma cell lines were maintained as described (10, 11). SMS-KCN and SMS-KAN were obtained from Patrick Reynolds (Navel Medical Research Institute, Bethesda, MD), SK-N-SH and LAN-1-19n were obtained from June Biedler (Memorial Sloan-Kettering Cancer Center), and SK-N-SH(F) was obtained from Neil Sidell (University of California, Los Angeles). Other cell lines were obtained from sources indicated in the citations that reference them.

Isolation and Characterization of Id-2. PCR and RACE-PCR (where RACE indicates rapid amplification of cDNA ends) was carried out as described (12). Id-2 cDNAs generated by PCR were cloned into pBluescript SK II (Stratagene). The RACE-PCR cDNA was used to screen a human leukocyte genomic DNA library in EMBL-3 (Clontech). Two phage containing largely overlapping inserts of ~16 kilobases (kb) of genomic sequence each were isolated. One of these phage was digested with EcoRI and Sal I, and these fragments were subcloned into pGEM-3Z (Promega). The genomic subclones were analyzed by Southern blot hybridization with the Id-2 cDNA to determine the location of Id-2-transcribed sequences. One such subclone, pGld-9, which contained a 5.6-kb genomic insert, was digested with EcoRI and Pst I. Two EcoRI/Pst I fragments of 0.45 kb and 1.2 kb from pGld-9 were subcloned into pBluescript SK II (Fig. 1B). DNA sequence was obtained from these three subclones (pGld9, pGld0.45, and pGld1.2) as described above using both

Abbreviations: HLH, helix–loop–helix; PNS, peripheral nervous system; RACE, rapid amplification of cDNA ends; MASH, mammalian achete/scute homolog.

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

1512
plasmid primers (40 primer and reverse primer; Stratagene) and numerous internal primers. A total of 1488 bases of genomic nucleotide sequence was obtained: this includes all genomic DNA indicated by the solid line in Fig. 1A except for ~800 bases within the second intron of the Id-2 gene (Fig. 1B and unpublished data). cDNAs containing the complete Id-2 coding sequences were generated by PCR reactions with primer 202 (Fig. 1) and either primer 208 (Fig. 1) or a primer made from the first six codons of the Id-2 gene (primer 201, not shown). cDNAs were sequenced as double-stranded plasmids by the dideoxynucleotide chain-termination method using a protocol and Sequenase from United States Biochemical.

RNA Blot Analysis. Gels (1.2% agarose/formaldehyde) were run as described (13), except that the agarose mixture contained 1 µg of ethidium bromide per ml to stain rRNA. Twenty to 40 µg of total RNA was loaded in each lane as indicated in the figure legends. Total RNA was prepared from cultured cells and human tissues using the RNAzol protocol (Tel-Test, Friendswood, TX). For quantitation, a photograph of ethidium-stained samples was taken and densitometry was performed using an LKB Ultrascan XL.

Treatment of Neuroblastoma Cell Lines with Retinoic Acid. Neuroblastoma cell line SMS-KCN was plated in 150-mm tissue culture flasks in RPMI 1640 medium plus penicillin and streptomycin with 15% fetal bovine serum. Forty eight hours later SMS-KCN cells were treated with 10 µM retinoic acid [stock solution: 5 mM all-trans-retinoic acid (Sigma) in ethanol] or ethanol alone. Cells were refed with 10 µM retinoic acid or ethanol every 48 hr. Medium was replaced every 4 days.

RESULTS

Structure of the Id-2 Gene. To identify Id-related HLH genes expressed during maturation of nervous system tissues, RACE-PCR (12) was used to determine if mRNAs from such genes were present in total RNA extracted from cultured neuroblastoma cells. Neuroblastoma is a PNS tumor of childhood that gives rise to cell lines retaining many features of normal neural precursors (14). To amplify cDNAs corresponding to mRNAs with homology to Id, the DNA sequence coding for amino acids 89–94 of the Id protein was used as a primer for RACE-PCR. After first strand cDNA was synthesized from neuroblastoma cell RNA using a poly(dT) primer with an attached polylinker, a PCR was performed using the Id sequence primer and a primer homologous to the polylinker region of the poly(dT) primer. This reaction amplified several cDNAs when RNA extracted from the neuroblastoma cell line SK-N-SH or its cloned derivative, SY5Y (15), was used as a template. DNA sequence analysis of subcloned PCR products revealed three different cDNAs encoded by a gene with substantial homology to the Id gene: this Id-like gene was designated Id-2.

Using DNA of an Id-2 cDNA clone as a probe, a human genomic DNA library from normal leukocytes was screened, and two phage containing Id-2 genomic DNA sequence were isolated (see Materials and Methods for details of subcloning and sequencing). DNA sequence from the genomic clones yielded a complete open reading frame that has strong homology to the Id gene and another murine Id-like gene, HLH 462 (8), as shown in Figs. 1 and 2. In addition to strong homology within the HLH domain, there are regions of homology outside this domain (Fig. 2, Id-2 amino acids 1–32 and 100–113). These domains contain potential protein phosphorylation sites. We designated nucleotides 367–369 the Id-2 initiation codon based upon its location at the 5′ end of the open reading frame with strong HLH region homology to Id and HLH462 (Fig. 2), the presence of stop codons 5′ to this position, and the presence of nucleotide sequences conforming

![Fig. 1](A) Diagram of the Id-2 gene. The solid upper line represents genomic DNA. E, EcoRI; P, Pst I. (A) Indicates the location of stretches of poly(A) present in the genomic sequence and incorporated into the primary Id-2 transcript. The bars below the solid line represent exons present in Id-2 cDNAs. Solid areas indicate coding sequence; open areas indicate 5′ or 3′ untranslated sequence. The striped bar indicates 3′ untranslated sequence present in the Id-2 mRNA but not in cDNAs. (B) Nucleotide sequence of the Id-2 gene. The sequence of the coding region and 3′ untranslated region comes from genomic and cDNA clones. The 5′ promoter region is from genomic clones. Intron sequences are not shown, but intron positions are indicated by solid triangles. Primers used in the PCRs discussed in the text are shown by arrows above the nucleotide sequence: RACE primer, nucleotides 562–580; primer 208, nucleotides 274–291; primer 209, nucleotides 245–262; primer 202, nucleotides 763–780. TATA-like and CAAT-like elements are underlined. Potential AP2 binding sites are shown by solid lines above the nucleotide sequence. The predicted amino acid sequence of the Id-2 gene is shown below the nucleotide sequence.
Fig. 2. Comparison of the predicted amino acid sequence of Id-2 with sequences of Id and HLH 462. Protein sequences are shown in single-letter code. Amino acid homology is indicated by dots; the HLH domain is underlined. Numbering refers to Id-2 amino acids. In addition to strong homology within the HLH domain, there are striking regions of homology outside this domain (Id-2 amino acids 1–32 and 100–113).

Although Id-2 has extensive HLH region homology to murine myogenic Id, Id-2 is not simply the human homolog of this gene. We have cloned a partial cDNA from a gene encoding what is clearly the human homolog of Id; outside the HLH region, the predicted amino acid sequence of this clone is quite different from that of Id-2 but is 95% homologous to murine myogenic Id (J. Dahmen and J.B., unpublished). This human myogenic Id clone hybridizes to a 1.1-kb mRNA that is also detected by murine myogenic Id probes and not to the 1.3-kb mRNA detected by Id-2 probes (J.B., unpublished data). Finally, the expression pattern of the myogenic Id gene in tissues and cell lines is different from that observed for Id-2 (J.B., unpublished data).

To find the approximate location of the transcription start site of the Id-2 mRNA, genomic sequence 5′ to the start codon at nucleotide 367 was used to synthesize primers for PCRs on cDNA templates. A primer consisting of genomic DNA sequence 76–93 nucleotides upstream from the start codon (primer 208, Fig. 1) and a primer from the 3′ end of the gene (primer 202, Fig. 1) yielded a PCR product of the predicted size from cDNA made with total RNA from the cell line SK-N-SH. However, a primer consisting of genomic sequence 105–122 nucleotides upstream from the start codon (primer 209, Fig. 1) did not yield a PCR product from the same cDNA when used in a parallel reaction with primer 202. Both sets of primers produced PCR products of the predicted size when used with genomic DNA template (unpublished data). These results place the transcription start site of the Id-2 gene at 108 bases 5′ to the start codon. The transcription start site of eukaryotic genes is commonly an adenine flanked by pyrimidines. Such an adenine is located at 108 bases 5′ to the Id-2 gene start codon at nucleotide 367. A sequence similar to the consensus TATA box is located 25 bases 5′ to this adenine, whereas a CAAT box sequence is located 72–79 bases upstream (Fig. 1). This spacing is common to many eukaryotic genes. A search of the putative promoter sequence revealed two potential binding sites for the transcription factor AP2 (Fig. 1), which mediates transcriptional activation in response to diacylglycerol-activated protein kinase C and CAMP-dependent protein kinase A (17).

Comparison of Id-2 cDNA sequence to the genomic DNA sequence indicated that a 329-base intron is spliced out of the primary Id-2 transcript between codons 116 and 117 and that a second 0.9-kb intron is spliced out of the 3′ untranslated sequence 7 nucleotides after the stop codon (nucleotide 778, Fig. 1). One of the cDNAs generated by RACE-PCR did not contain the 3′ exon but originated in the first intron at a stretch of poly(A) (Fig. 1A and unpublished data). This cDNA could be reproducibly amplified from total RNA in PCR reactions, and control reactions showed that it was not derived from contaminating genomic DNA. We believe this cDNA arises from an unspliced 2.5-kb precursor RNA, which can be detected in some cell lines by Northern blot analysis (unpublished data) and not from an alternately spliced mRNA, since no evidence of alternative mature transcripts is seen on Northern blots. The 3′ untranslated sequence of Id-2 shown in Fig. 1 contains no poly(A) signal. We were able to characterize 640 nucleotides of transcribed Id-2 DNA from the transcription start site to the site at which our poly(dT) primer hybridized, while the major transcript detected by Id-2 probes on a Northern blot is 1.3 kb (for example, see Fig. 3). Several different Id-2 probes hybridize to this 1.3-kb mRNA, including probes that contain no sequences from the HLH coding domain, indicating that the 1.3-kb mRNA is transcribed from the Id-2 gene (unpublished data). Examination of genomic DNA sequence from the region around the 3′ end of the cDNA indicated that this discrepancy resulted from hybridization of the poly(dT) primer used to synthesize the first strand of the RACE cDNAs to a stretch of poly(A) within the 3′ untranslated sequence of the Id-2 mRNA (unpublished data). We have therefore concluded that the Id-2 mRNA contains additional 3′ untranslated sequence not incorporated into the cDNAs we cloned (Fig. 1A and unpublished data). We have been unable to detect any RACE-PCR cDNA generated by hybridization of the poly(dT) primer to a poly(A) tail at the 3′ end of the Id-2 mRNA, suggesting that the Id-2 mRNA may not be polyadenylated. This possibility is supported by our observation that Id-2 mRNA is not enriched in preparations of poly(A)+ RNA (unpublished data).

Expression of Id-2 mRNA in Normal Human Tissue and Neural Tumors. In a survey of human tissues, we found that Id-2 was differentially expressed over the course of development. In brain, heart, and kidney, expression was much higher in fetal tissues than in the corresponding adult tissues.
We were especially interested in the expression of the Id-2 gene in the central nervous system and the PNS. To examine expression of Id-2 in nervous system tissue, Northern blot analysis was utilized to examine RNA from fetal brain, adult brain, and cell lines of nervous system tumors. As shown in Fig. 4, lanes 1–4, Id-2 expression was readily detectable in fetal brain at 80 days of gestation. The level of Id-2 mRNA was substantially lower in earlier and later fetal brain and in adult brain. Neuroblastomas and neuroepitheliomas are tumors of the PNS that occur in childhood, and cells isolated from these tumors are thought to represent undifferentiated neural crest cells that would normally have differentiated into neuroendocrine and neuronal cells of the PNS (11). Id-2 expression was detected in 12 of 15 neuroblastoma lines and 2 of 2 neuroepithelioma lines examined. Examples of Id-2 expression in these nervous system tumors are shown in Fig. 4, lanes 5–7 and 10. The neuroblastoma lines in which Id-2 expression was detected were SK-N-SH, SHEP, SY5Y, SK-N-SH(F), LAN-1, LAN-1-6S, LAN-1-19N, SMS-KAN, SK-N-AS, 12 IMR-32, NGP, and SMS-KCNR (18). Neuroblastoma lines in which expression could not be detected by Northern analysis were SK-N-DZ, GECAN, and SK-N-FI (18). The neuroepithelial cell lines examined were SK-N-MC and TC-32. The level of Id-2 expression varies among the Id-2 cell lines we examined. The expression observed in cell lines SK-N-SH and SMS-KAN illustrates the range of Id-2 expression, which we estimate by normalized densitometry to be 2- to 3-fold (Fig. 4, lanes 5 and 6). The Id-2 gene is also expressed in cell lines derived from gliomas, the most common of all malignant brain tumors arising in the glial cells of the central nervous system (10, 19). Glioma cell lines have been divided into three categories based on cell morphology and patterns of oncogene expression: bipolar, epithelial, and pleomorphic-glial (10). The Id-2 gene was expressed in 5 of 5 bipolar and 2 of 2 pleomorphic-glial cell lines examined but in only 1 of 4 epithelial lines examined; two examples are shown in Fig. 4, lanes 8 and 9. The bipolar glioma cell lines examined were U343 MG, U118 MG, U138 MG, SF-295, and SF-216. The epithelial glioma cell line expressing Id-2 was U87 mg; the three that did not express Id-2 by Northern analysis were SF-210, U343 MGA C12.6, and T-98G. The two pleomorphic glial cell lines examined were U251 mg and SF268 (10).

Modulation of Id-2 mRNA Expression by Retinoic Acid. Some neuroblastoma cell lines have been used as models of PNS development (14, 20). Several of these cell lines will differentiate to neurons when treated with retinoic acid. One of the best characterized of such lines is SMS-KCNR, a cell line in which the morphologic changes that occur during differentiation have been correlated with the expression of a number of genes that mark neuronal differentiation (21). We monitored the retinoic acid-induced differentiation of SMS-KCNR cells by examining their morphology at various times after treatment. As shown in Fig. 5B, evidence of SMS-KCNR neuronal differentiation is obvious 48 hr after retinoic acid treatment when neurites extending from the treated cells are readily observed. After 7 days of treatment, these neuritic processes have formed bundles that resemble neurons. When Id-2 mRNA levels in retinoic acid-treated SMS-KCNR cells were analyzed by Northern blot hybridization, it was observed that the level of Id-2 mRNA falls as the cells differentiate to neurons (Fig. 5A). By 96 hr of retinoic acid treatment the level of Id-2 expression is substantially lower when compared to controls.

DISCUSSION

Expression of the Id-2 gene in human fetal nervous system and nervous system tumors at a level much higher than that observed in adult brain suggests that Id-2 may function in a
is retinoic acid, leads to morphological, physiological, and cell biological evidence of ganglionic differentiation.

In preliminary experiments we found that in vitro induction of Schwannian and neuroblastomictive differentiation was associated with decreased expression of Id-2. We therefore evaluated the expression of Id-2 during retinoic acid-induced neuronal differentiation of SMS-KCNCR. We detected decreased Id-2 mRNA levels in retinoic acid-treated SMS-KCNCR cells. The expression of Id-2 at 80 days of gestation corresponds well with neuronal proliferation in the human central nervous system (24). This finding is compatible with the possibility that differentiation of nervous system tissues must be inhibited while proliferation is occurring. The finding of elevated Id-2 expression in neuroblasts lends additional credence to this possibility, since this particular tumor is widely believed to arise in association with the arrested differentiation of neural crest precursor cells. Our findings suggest that the down-regulation of Id-2 may be required for neuronal and glial differentiation to proceed in much the same manner as decreased Id allows myogenic differentiation.

We thank Jane Johnson and David Anderson for MASH cDNAs, Steven Reeves and Benny Usog for help with figures and photographs, and Lucy de la Calzada for help in preparing the manuscript. Fetal tissues were obtained from Dr. T. H. Shepard (Central Laboratory of Embryology, University of Washington, Seattle). We thank the Preuss Foundation and the family of Aaron Price for their generous support of this research. J.B. is supported by a Cancer Research Coordinating Committee grant. E.V.M. is supported by the National Neurofibromatosis Foundation's Young Investigator's Award.