A novel initiator regulates expression of the nontissue-specific helix–loop–helix gene ME1

Daniel H. Shain*, Toomas Neuman¹ and Mauricio X. Zuber†

Department of Biochemistry and Molecular Biology and ¹Department of Anatomy and Neurobiology, Colorado State University, Fort Collins, CO 80523, USA

Received February 13, 1995; Revised and Accepted March 27, 1995

GenBank accession no. U19863

ABSTRACT

The mouse ME1 gene (HEB, REB and GE1, homologues in human, rat and chick, respectively) is a member of the nontissue-specific helix–loop–helix (HLH) gene family that includes E2A, E2-2 and *Drosophila daughterless*. We have examined the factors that control ME1 gene expression. ME1 is a single copy gene that spans ≥ 150 kb of DNA and contains >10 exons. Transcription was directed by an unusual initiator element that contained a 13 bp poly d(A) tract flanked by palindromic and inverted repeat sequences. Both RNase protection and primer extension analyses mapped the ME1 transcriptional start site to the center of the 13 bp poly d(A) tract. The ME1 initiator and its proximal sequences were required for promoter activity, supported basal levels of transcription, and contributed to cell type-specific gene expression. Other *cis*-elements utilized by the TATA-less ME1 promoter included a cluster of Sp1 response elements, E-boxes and a strong repressor. Collectively, our results suggest that the ME1 initiator and other *cis*-elements in the proximal promoter play an important role in regulating ME1 gene expression.

INTRODUCTION

Helix–loop–helix (HLH) transcription factors are pivotal regulators of cell fate in a number of developmental pathways (for reviews, see 1,2). At the molecular level, HLH proteins direct gene expression by recognizing the DNA consensus sequence CANNTG, known as an E-box (3–5). Based upon their tissue distribution or amino acid sequence, HLH proteins are classified into several different groups. Tissue-specific HLH proteins comprise the largest HLH family and include the extensively characterized myogenic proteins which regulate muscle differentiation (6–9), and the *achaete–scute* proteins which are involved in *Drosophila* neurogenesis and sex determination (10–12).

Nontissue-specific HLH proteins comprise a much smaller family (the E-protein family) that includes E2A (13), E2-2 (14), ME1 [ME1, HEB, REB and GE1, homologues in mouse (15), human (16), rat (17) and chick (15), respectively] and *Drosophila*

daughterless (18). E-proteins are required heterodimer partners of the tissue-specific HLH proteins (5,19), and also form functional homodimers (16,20,21). Although E-proteins are expressed in a nontissue-specific manner, they are not expressed ubiquitously; rather, they have dynamic expression patterns that overlap to varying degrees (15,22–24). Due to their functional redundancy, the role of individual E-proteins has been difficult to determine (25,26). Recently, however, E2A was implicated in the initial stages of immunoglobulin gene rearrangements and in B cell formation (28,29). Surprisingly, a mouse E2A heterozygous mutant contained about half as many B cells as normal mice suggesting that E2A levels are translated into B cell numbers (29). Similar HLH mediated dosage effects were observed in muscle differentiation (8,30,31) and *Drosophila* sex determination (11,32), suggesting that levels of HLH gene expression are tightly controlled during cell fate decisions. It is therefore intriguing to ask how these ‘master’ regulators of cell fate are themselves regulated?

Analysis of tissue-specific HLH genes in the MyoD family and the *achaete–scute* complex indicate that their expression is directed by unique response elements (33,34), and that they are cross-regulated by other HLH proteins (35–37). By comparison, little is known about the regulation of nontissue-specific HLH genes. To gain an understanding of how these genes are regulated, we investigated the factors that control ME1 gene expression. Previously, we and others have shown that ME1 mRNA is enriched in morphogenetically active regions of the cerebellum, neural tube and in T cells (15–17,21). Importantly, ME1 homodimers and heterodimers recognize distinct E-boxes indicating that ME1 may direct lineage-specific gene expression (16,17,32). Here we demonstrate the complexity of the ME1 gene and its regulation at the transcriptional level. ME1 is driven by a TATA-less promoter that utilizes a novel initiator in concert with other *cis*-elements to regulate transcriptional activity.

MATERIALS AND METHODS

Plasmids

The pBSSK-/CAT parent vector containing a *Hind*III–*Bam*HI fragment from the chloramphenicol acetyltransferase (CAT) gene was a gift from Dr Norman Curthoys. The *Bam*HI and *Sac*I sites downstream of the CAT gene were destroyed by blunting with T4 DNA polymerase and re-ligating (38). To make the ME1

* To whom correspondence should be addressed

† Deceased

promoter constructs, a *KpnI*–*SmaI* (blunted) ME1 genomic fragment was first subcloned into the *SmaI* – *EcoRV* sites of pBSSK⁺. The –3000 ME1 promoter construct was made by ligating an *XhoI*–*SmaI* ME1 genomic fragment into the *XhoI*–*HindIII* sites of the pBSSK⁺/CAT vector. To make the –1004 promoter construct, a *HindII*–*HindIII* ME1 genomic fragment was subcloned into the *NaeI*–*HindIII* sites of the pBSSK⁺/CAT vector. The remaining deletion constructs and block mutants were derived from the –1004 construct using the restriction endonuclease sites shown in Figure 2. Incompatible ends were blunted with T4 DNA polymerase and ligated according to standard procedures (38). Full length ME1 cDNA was cloned into the pRcCMV vector (Invitrogen) using *HindIII* linkers.

Southern blot analysis

Mouse genomic DNA (30 µg) was digested with the appropriate restriction endonucleases and electrophoresed at 1 V/cm on a 0.8% agarose gel. DNA was transferred to a GeneScreen⁺ (DuPont) membrane using a capillary procedure under alkaline conditions. ME1 cDNA probes were synthesized by random-primed labelling (Amersham) in the presence of [α -³²P]dCTP (3000 Ci/mmol; Amersham). Hybridizations were performed overnight in 50% formamide, 5 × SSPE, 5% SDS at 42°C. Washes were performed in 0.1 × SSC, 0.5% SDS at 68°C until no radioactivity could be detected in the wash solution. Autoradiography was performed using Hyperfilm-MP (Amersham).

Sub-genomic library construction

Mouse genomic DNA (100–200 µg) was digested with *EcoRI* and electrophoresed on a 1% low melt agarose gel (FMC genetic technology grade) at 1 V/cm to achieve adequate separation. The region containing 4 kb fragments was excised and the DNA was electroeluted as described previously (39). Southern analysis was performed on 1/10 of the sample to confirm that the correct DNA fragment was present. The remainder of the sample was electrophoresed as described above and the smallest size range of DNA fragments was excised (based on Southern analysis of the sub-band). The 4 kb *EcoRI* fragments were ligated into λ Zap II (Stratagene) arms at a molar ratio of 2:1 (insert:vector). λ DNA was packaged using Gigapack Gold II packaging extracts (Stratagene) and plated on the host strain XL1Blue MRF'. Plaque hybridizations were performed as described (38) using a cDNA probe specific to the first exon of the ME1 gene.

DNA sequencing

Sequencing was performed on both strands using the T3 and T7 primers on subcloned genomic fragments. Sequence reactions were performed using the Sequenase Version 2.0 kit (US Biochemical) according to the manufacturers specifications. Restriction endonuclease and transcription factor binding sites were identified by MacVector version 3.5 software (IBI).

RNA analyses

Total RNA from tissues or cell lines was purified as described previously (40). For RNase protection analyses, template DNA was linearized with the appropriate restriction endonuclease and antisense RNA was synthesized *in vitro* using T7 RNA polymerase (Stratagene) in the presence of [α -³²P]UTP as described (38).

Approximately 20 µg of total RNA and 10⁴–10⁵ c.p.m. of riboprobe were co-precipitated and resuspended in a solution containing 80% formamide, 100 mM sodium citrate (pH 6.4), 300 mM sodium acetate (pH 6.4) and 1 mM EDTA. RNA was denatured at 90°C for 3–4 min and allowed to anneal overnight at 45°C. RNase digestion was performed with an RPA II kit (Ambion) using RNase A and RNase T1 for 30 min at 37°C. For primer extensions, 20 µg of total RNA and 10⁴–10⁵ c.p.m. of the [γ -³²P]ATP end-labeled, CAT specific oligonucleotide 5'-CAACGGTCGTATATCCAGTGA-3' were incubated at 65°C for 2 min. Annealing buffer was then added, bringing the solution to 200 mM NaCl, 40 mM Tris (pH 7.5) and 5 mM EDTA (pH 8.0) in a final volume of 20 µl. Annealing was performed for 3 h at 50°C upon which the RNA was precipitated and resuspended in 15 µl of a solution containing 10 U AMV reverse transcriptase, 1 × buffer (supplied by CloneTech), 15 U RNase inhibitor and 1 mM dNTPs. After a 60 min incubation at 42°C the reaction was quenched on ice, precipitated with 0.5 µl 10 M ammonium acetate and 50 µl EtOH, and resuspended in loading buffer (50% formamide containing 0.2% xylene cyanol and 0.2% bromophenol blue). Protected RNA fragments and primer extension products were separated on a 6% polyacrylamide–40% urea gel and subjected to autoradiography using Hyperfilm-MP (Amersham).

Cell culture and transfection assays

PCC7, Swiss 3T3 and Cos-7 cell lines were maintained in Dulbecco's Modified Eagles Medium (Gibco) supplemented with 10% fetal calf serum, 2 mM L-glutamine, 100 U penicillin-G/ml and 100 mg streptomycin/ml. Cells were grown at 37°C in a humidified incubator with 5% CO₂. Transient transfections were performed by calcium phosphate co-precipitation (41). Cells were plated at 20–40% confluency on a 100 mm plate (Nunc) and transfected with 10 µg of test plasmid and 2 µg of a β -gal vector driven by the RSV promoter. Medium was changed 12–24 h after transfection and cells were harvested after 36–48 h. Cell extracts were prepared by three freeze–thaw cycles in 200 µl of a 10 mM Tris (pH 8.0), 1 mM EDTA solution and protein concentrations were determined using a BioRad protein assay kit. Equal amounts of protein (150 µg) were used in each CAT assay which was incubated at 37°C for 30–90 min. Transfection efficiencies were determined by a β -gal assay (38) but did not significantly affect the values presented. The percentage conversion of ¹⁴C chloramphenicol (Amersham) was quantitated on a PhosphorImager (Molecular Dynamics). Enzyme activity above the pBSSK⁺/CAT vector was normalized to the –240 construct. For each construct, at least two independent clones were tested and values are the results from at least three independent experiments.

Micro-extraction procedure for the preparation of whole cell extracts and electrophoretic mobility shift assay (EMSA)

These methods were performed as described previously (42). Briefly, extraction buffer contained 20 mM HEPES (pH 7.8), 450 mM NaCl, 0.2 mM EDTA, 0.5 mM DTT, 25% glycerol and a mixture of protease inhibitors: PMSF (0.5 mM), leupeptin (0.5 mg/ml), pepstatin (0.7 mg/ml), aprotinin (1 mg/ml) and bestatin (40 mg/ml). After sonication, extracts were cleared by micro-centrifugation for 5 min at 4°C. Binding conditions for EMSA were 10 mM HEPES (pH 7.8), 1 mM spermidine, 5 mM MgCl₂,

Figure 1. Complexity of the ME1 gene. Mouse genomic DNA (30 μ g) was subjected to Southern blot analysis using the restriction endonucleases *Bgl*II, *Eco*RI, and *Pst*I. (A) Hybridization with a full length ME1 cDNA probe. Summation of bands in each lane suggests that the ME1 gene spans >100 kb of DNA. Based on the number of bands in the *Bgl*II lane it can be concluded that the ME1 gene contains more than 10 exons. (B) Hybridization with a probe specific to the first ME1 exon. Bands in each lane are as follows: *Bgl*II, 6.5 kb; *Eco*RI, 4 kb and *Pst*I, 23 kb. The presence of a single band in each lane suggests that ME1 is a single copy gene. A sub-genomic library was constructed from the 4 kb *Eco*RI band to clone the ME1 promoter.

50 mM KCl, 0.5 mM DTT, 9% glycerol, 0.8 mg poly d(I-C), and 10 000–50 000 c.p.m. of labelled oligonucleotide. After incubation for 15 min at 37°C, binding reactions were loaded on a 5% polyacrylamide gel, electrophoresed to separate DNA-protein complexes, and analyzed by autoradiography. EMSA analyses were performed with several batches of nuclear extracts and no differences were observed.

RESULTS

Complexity of the ME1 gene

To gain an understanding of the size and the intron/exon structure of the ME1 gene, a genomic Southern blot analysis was performed using a full length ME1 cDNA probe (Fig. 1A). Summation of bands in each restriction endonuclease lane indicated that the ME1 gene spans $\geq$ 100 kb of DNA. Since intron sequences are not fully represented by this analysis, the actual length of ME1 could be considerably longer than 100 kb. Supporting this idea, 30 independent ME1 genomic clones with an average insert size of 15 kb were isolated from an EMBL3 mouse genomic library (CloneTech). Of these, a non-contiguous set of 12 clones was identified that spanned $\sim$ 150 kb of DNA (not shown). Chromosomal walking failed to bridge gaps in this set of clones indicating that introns in the ME1 gene are relatively long (>15 kb). In addition, it can be concluded that there are more than 10 exons in the ME1 gene based upon the number of bands in the *Bgl*II restriction endonuclease lane (Fig. 1A).

Previous studies indicated that full length ME1 cDNA is $\sim$ 5 kb in length (15,21). A genomic Southern blot using a 5' fragment from a near full length ME1 cDNA revealed a single, major band in *Bgl*II, *Eco*RI and *Pst*I restriction endonuclease lanes, suggesting that ME1 is a single copy gene (Fig. 1B). To clone the ME1 promoter, a sub-genomic library was constructed from the 4 kb *Eco*RI band which contained the first ME1 exon (we were unable to isolate this region from the mouse genomic library described

Figure 2. DNA sequence of the ME1 promoter and first exon. The DNA sequence of the first ME1 exon and $\sim$ 1350 bp of 5' flanking sequences are shown. The ME1 transcriptional start site occurs at the center of a 13 bp poly d(A) tract and is indicated +1. A 16 bp palindrome and a 9 bp inverted repeat flanking the transcriptional start site are represented by dashed underlines. Collectively, this region ($\sim$ 24 to +17 bp) is referred to as the ME1 initiator. The longest ME1 cDNA clone obtained from a mouse random-primed cDNA library (Stratagene) is shown in bold at +18. Restriction endonuclease sites utilized in the construction of ME1 promoter mutants are underlined. Sp1 response elements and E-boxes are double underlined. The asterisk at +236 denotes the 3' splice junction of the first exon. The ME1 translation start site occurs downstream of the first exon and therefore is not shown.

above). The complete DNA sequence of the first exon and 1350 bp of 5' flanking sequences are shown in Figure 2. The intron/exon junction of the first ME1 exon (+236) was determined by the point at which the ME1 cDNA and ME1 genomic sequences diverged. The translation start site of ME1 occurs downstream of exon 1 and is therefore not indicated in Figure 2.

Transcriptional initiation in the ME1 gene occurs within a poly d(A) tract

The longest ME1 cDNA obtained from a mouse random-primed cDNA library (Stratagene) terminated 18 bp downstream of a 13 bp poly d(A) tract (Fig. 2). Since most cDNA clones terminated at similar 5' locations, we utilized an RNase protection assay to determine whether the ME1 transcriptional start site was located in this region. Three different antisense riboprobes were tested, all of which generated identical results. A representative RNase protection assay using an antisense riboprobe synthesized from the *Eco*RI09 (+121 bp) to the *Bam*HI (–45 bp) site (Fig. 2) is shown in Figure 3. As a positive control, a sense RNA strand was synthesized from the *Bam*HI to the *Eco*RI09 site using the same plasmid DNA. It is important to note that the antisense riboprobe and the sense RNA strand generate a protected band of the expected size (166 bp) and also a smaller, hypersensitive site that maps to the 13 bp poly d(A) tract. This hypersensitive region is likely to result from 'breathing' of the poly d(A) tract since hydrogen bonding is relatively weak in A-T rich stretches.

In new-born mouse brain, kidney, liver and a teratocarcinoma PCC7 cell line (43), protected ME1 mRNA fragments formed a

Figure 3. Determination of the ME1 transcriptional start site by RNase protection analysis. Total RNA (20 µg) from each new-born tissue or cell line was analyzed with an ME1 antisense riboprobe synthesized from the *Eco*O109 (+121 bp) to the *Bam*HI site (−45 bp). The ME1 riboprobe is 187 bp in length; protected fragments in brain, PCC7 cells, kidney and liver RNA are from 113 to 121 bp. The most 5' protected fragment (121 bp) mapped to the center of a 13 bp poly d(A) tract (Fig. 2). Kidney RNA has additional protected fragments of 70 and 90–100 bp. No protected fragments were observed in yeast RNA. A sense RNA strand synthesized from the *Bam*HI (−45 bp) to the *Eco*OR109 (+121 bp) site generated a protected fragment of the expected size (166 bp) and also a smaller, hypersensitive site that mapped to the 13 bp poly d(A) tract. Bands located between the ME1 probe and the ladder are from an unrelated riboprobe.

ladder that mapped to the 13 bp poly d(A) tract (Fig. 3). One explanation for this banding pattern is that transcription initiates at the center of the poly d(A) tract (i.e. the longest protected fragment), and smaller bands result from 'breathing' of the poly d(A) tract as observed with the sense RNA strand. We cannot rule out the possibility that transcription initiates from multiple locations within the poly d(A) tract or slightly upstream of the poly d(A) tract. In the latter case, a protected fragment may not be observed due to the hypersensitivity of this region. It is clear, however, that the start site(s) does not extend as far as the *Bam*HI site (−45 bp) since this fragment is protected by the sense RNA strand but not by endogenous RNA samples. By the same argument, it is unlikely that RNA secondary structure is interfering with the analysis (i.e. the sense RNA template generates a fully protected RNA fragment of 166 bp). It is also unlikely that this region represents an intron/exon boundary since most ME1 cDNA clones terminated immediately downstream of the poly d(A) tract (44). Moreover, there are no consensus splice sites in this region. Thus the major transcriptional start site of the ME1 gene is likely to occur within, or slightly proximal to the 13 bp poly d(A) tract. Interestingly, the poly d(A) tract is flanked by the 16 bp palindrome 5'-GCTGAGGCGCCTCAGC-3' from −23 to −9 bp and the 9 bp inverted repeat 5'-GTCCGCCTG-3' (centered around a guanosine nucleotide) from +9 to +17 bp (Fig. 2). This unusual combination of sequences is referred to as the ME1 initiator element. Based upon the sequence upstream of the ME1 initiator, the ME1 gene appears to be driven by a TATA-less promoter (Fig. 2). DNA consensus elements in the 5' flanking region include a cluster of Sp1 sites (−154 to −134 bp) and several E-boxes (−396, −330, −251 and −240 bp).

Several additional transcriptional start sites downstream of the poly d(A) tract appear to be utilized in new-born kidney. It has not been determined whether these represent real transcriptional start sites or whether they result from RNA degradation. It is interesting, however, that new-born kidney expressed the highest levels of ME1 mRNA among the tissues that were tested (44).

Identification of response elements in the ME1 promoter

Upon identification of the major ME1 transcriptional start site we next examined this region for promoter activity. A series of ME1 promoter mutants were fused with the reporter gene chloramphenicol acetyltransferase (CAT) and tested by transient transfection analysis in PCC7, Swiss 3T3 (mouse fibroblast) and Cos-7 (monkey kidney) cells. PCC7 cells express considerably higher levels of endogenous ME1 mRNA than either Swiss 3T3 or Cos-7 cells based upon Northern blot analysis (44). We were unable to identify a cell line that did not express endogenous ME1 mRNA; this was not too surprising considering the widespread expression of the ME1 gene (15–17,21).

Results from transient transfection assays suggest that the major *cis*-elements of the ME1 promoter are located within −570 bp since larger constructs, including a −3000 bp construct (not shown), do not differ significantly. This does not preclude the possibility, however, that important regulatory elements exist beyond the region that was tested. The ME1 core promoter can be divided into three response regions; a region proximal to the transcriptional start site (−44 to +58 bp) which contains the ME1 initiator, an enhancer region (−240 to −28 bp), and a repressor region (−570 to −314 bp) (Fig. 4). Deletion of the ME1 initiator in the Δ−44/+58 mutant abrogated promoter activity in all three cell types suggesting that this region plays a central role in regulating promoter activity. Removal of the enhancer region, which contains a cluster of three Sp1 consensus elements, abrogated promoter activity in the Δ−285/−28 mutant (all cell types), and attenuated activity 5- to 6-fold in the Δ−826/−47 block mutant (PCC7 and Cos-7 cells; activity was abrogated in Swiss 3T3 cells). The difference observed between the Δ−285/−28 and Δ−826/−47 constructs was likely to result from the repressor region which is present in the Δ−285/−28 construct but absent in the Δ−826/−47 construct. Deletion of the repressor region in constructs −314 and −240 resulted in a 3- to 4-fold activation of promoter activity.

Swiss 3T3 cells appear to recognize a cell type-specific repressor in a region between −1004 and −570 bp (compare constructs −1004 and −570 in Fig. 4). Also, significant levels of CAT activity were observed in PCC7 and Cos-7 cells when the ME1 initiator and its proximal sequences were tested independently (mutant Δ−826/−47), implying that this region could support basal levels of transcription in these cells but not in Swiss 3T3 cells. This raises the possibility that the ME1 initiator is involved in cell type-specific gene expression. It should be noted that mutant Δ−285/−28, which lacks the enhancer, did not generate basal levels of promoter activity; this was likely to result from the repressor region (−570 to −314 bp) which is contained in this construct.

Primer extension analyses were performed to confirm that transfected cells utilized the poly d(A) tract as their major transcriptional start site in ME1 promoter constructs. PCC7 cells transfected with the constitutively active −240 ME1 promoter

Figure 4. ME1 promoter deletion analysis. A series of deletion and block mutants in the ME1 promoter were cloned into the expression vector pBSSK+/CAT. Restriction endonuclease sites indicated in Figure 2 were utilized for all constructs. The large, open box represents the first 58 bp in ME1 exon #1; the arrow indicates the ME1 transcriptional start site; solid ovals represent Sp1 consensus elements; and small, open boxes represent E-boxes (in all figures). Quantitation of CAT activity in PCC7, Swiss 3T3 and Cos-7 cells suggests that the ME1 promoter can be divided into three response regions; a region containing the ME1 initiator and proximal sequences (−44 to +58 bp) which is essential for promoter activity in all cell types, an enhancer region (−240 to −28 bp), and a repressor region (−570 to −314 bp). A Swiss 3T3 specific repressor element may be located between −1004 and −570 bp. Basal levels of promoter activity were generated from mutant Δ−826/−47 in PCC7 and Cos-7 cells, but not in Swiss 3T3 cells. CAT activities were normalized to the −240 construct and are the average of a least three independent experiments. The experimental mean and standard deviation from each construct are shown.

construct were used in the analysis so that extension products could be readily detected. Extension from a CAT specific oligonucleotide mapped the ME1 transcriptional start site to the center of the 13 bp poly d(A) tract (Fig. 5). This implies that the poly d(A) tract, and not a cryptic initiator, was utilized as the major transcriptional start site in ME1 promoter constructs. This also provides direct evidence that transcription of the ME1 gene initiates at the center of the poly d(A) tract, which is consistent with RNase protection results (Fig. 3). Importantly, both techniques mapped the transcriptional start site to the same nucleotide within the 13 bp poly d(A) tract.

Role of the ME1 initiator element in cell type-specific gene expression

To further explore the role of the ME1 initiator, we determined whether it was required in the constitutively active −240 ME1 promoter construct. A region proximal to the transcriptional start site (−32 to +58 bp) was deleted from the −240 construct and the resulting double mutant (Δ−32/+58) was expressed in cultured cells. Surprisingly, levels of promoter activity were attenuated ~5-fold in PCC7 cells but were not significantly affected in Swiss 3T3 and Cos-7 cells (Fig. 6). This is in apparent contradiction with previous results which demonstrated a requirement for this region in all three cell types (Fig. 4, mutant Δ−44/+58). This may be due to a cluster of unchecked Sp1 response elements in the truncated −240 construct (i.e. the repressor upstream of the Sp1

Figure 5. Primer extension analysis of the ME1 transcriptional start site in transfected PCC7 cells. The CAT specific oligonucleotide 5'-CAACGGTCG-TATATCCAGTGA-3' was used to determine the transcriptional start site utilized by the constitutively active −240 ME1 promoter construct. PCC7 cells transfected with the −240 construct generated a major extension product that mapped to the center of the 13 poly d(A) tract (compare with ME1 ladder); untransfected PCC7 cells lack any detectable bands. In separate experiments, reverse transcriptase read through the poly d(A) tract using a sense mRNA or a DNA template (not shown); this provides strong evidence that the extension product observed resulted from initiation at the center of the poly d(A) tract.

sites is deleted). Sp1 proteins are potent activators of transcription that are thought to position basal transcriptional machinery in TATA-less promoters (45–48). It is possible, therefore, that Sp1 proteins directed transcriptional activity in Swiss 3T3 and Cos-7 cells by utilizing a cryptic initiator in the Δ−32/+58 construct. In the context of a longer ME1 promoter, upstream regulators may place more stringent constraints on the selection of a start site, perhaps by tempering the activity of Sp1 proteins.

Whatever the case may be, it is significant that promoter activity was attenuated in PCC7 cells but not in Swiss 3T3 or Cos-7 cells. This implies that cell types differ in their ability to recognize the ME1 initiator and its proximal sequences. To test this idea, a synthetic DNA fragment containing the ME1 initiator (−32 to +22 bp) was utilized in an electrophoretic mobility shift assay (EMSA) with protein extracts from PCC7 and Swiss 3T3 cells (Cos-7 cells were excluded from this analysis since meaningful comparisons could not be made between the mouse (PCC7 and Swiss 3T3) and primate (Cos-7) cell lines). EMSA analysis indicated that a different complement of proteins recognized the ME1 initiator in PCC7 and Swiss 3T3 cell extracts (Fig. 7). The ME1 initiator formed four major protein complexes with Swiss 3T3 cell extracts while at least seven complexes were observed with PCC7 cell extracts. Although complexes in Swiss

Figure 6. Deletion of the ME1 initiator in a constitutive ME1 promoter construct. A region proximal to the ME1 transcriptional start site (–32 to +58 bp) was deleted in the constitutively active –240 ME1 promoter construct. CAT activity was attenuated ~5-fold in PCC7 cells, but was not significantly affected in Cos-7 or Swiss 3T3 cells. CAT activities were normalized to the –240 construct and are the average of at least three experiments. The experimental mean and standard deviation from each construct are shown.

Figure 7. Cell type-specificity of ME1 initiator binding complexes. EMSA analysis of PCC7 (lanes 1, 2 and 3) and Swiss 3T3 (lanes 4, 5 and 6) cell extracts (10 μ g of protein) using a double-stranded oligonucleotide (–32 to +22 bp) containing the ME1 initiator. Arrows in the left and right margins indicate PCC7 and Swiss 3T3 protein complexes, respectively. Unlabeled ME1 initiator oligonucleotide (lanes 1 and 6) and a non-specific competitor oligonucleotide 5'-GGGCAGCTGGCAGGAAGCAGGTCATGTGGCAAGGCTATTTGG-3' corresponding to a 42 bp region of the immunoglobulin heavy chain enhancer containing μ E2 and μ E3 motifs (lanes 2 and 5), were present at 400-fold excess to demonstrate the specificity of binding complexes.

3T3 cells co-migrated with those from PCC7 cells, this does not mean they are composed of the same proteins (although this remains a possibility). Shifted complexes in both cell types bound the ME1 initiator specifically since they were not competed with non-specific competitor DNA.

ME1 regulates its own promoter

To determine the role of E-boxes in the ME1 promoter, several ME1 promoter constructs were co-transfected with a pRcCMV/ME1 construct that expressed a full length ME1 protein (Fig. 8). Experiments were performed in Cos-7 cells (derived from monkey kidney) to reduce the probability of ME1 interacting with endogenous HLH proteins. In the presence of the four proximal

Figure 8. Autoregulation of the ME1 promoter. ME1 promoter constructs containing four E-boxes (–570) or two E-boxes (–314) were co-transfected in Cos-7 cells with the pRc/CMV vector alone, or with a pRc/CMV vector that expressed full length ME1 protein. The –570 construct, containing four E-boxes, was enhanced ~3-fold in the presence of ME1 protein; the –314 construct, containing the two proximal E-boxes, was repressed ~3-fold in the presence of ME1 protein. These results are representative of three independent experiments.

E-boxes (construct –570), ME1 enhanced promoter activity between 2- and 3-fold. In a construct containing the two proximal E-boxes (construct –314), promoter activity was attenuated between 2- and 3-fold. Thus, the ME1 protein enhanced or repressed its gene promoter depending upon where the ME1 promoter was truncated.

DISCUSSION

We have investigated the regulatory properties of the nontissue-specific HLH gene ME1. Our results indicate that ME1 is a single copy gene that spans ≥ 150 kb of DNA and contains a conservative estimate of 10 exons. It is also known that the ME1 gene is differentially spliced into two isoforms (15,17,21). When compared to other HLH genes, the structure of the ME1 gene is unexpectedly complex. The MyoD and myogenin genes, for example, span ~3 kb of DNA and have only three exons (35,49), while the E2A and SCL genes are contained on ~20 kb of DNA (50,51). Importantly, HLH genes are susceptible to chromosomal rearrangements that have been linked with a number of leukemias (52–57). Of particular interest is the nontissue-specific E2A gene, a close relative of ME1, that undergoes at least two separate chromosomal translocations associated with B cell leukemia (53,54,56). In both cases, the amino terminus of E2A fused with the DNA binding domain of another transcription factor (53,54,56). ME1 has been implicated in regulating the Akv murine leukemia virus (21), raising the possibility that an ME1 chromosomal translocation could influence the progression of this disease. Whether the length or complexity of the ME1 gene has made it particularly susceptible to chromosomal translocations, and whether an ME1 translocation would have an oncogenic effect in the cell, remain important questions.

Functional elements in the ME1 promoter

ME1 lacks the common TATA-box that is present in most eukaryotic promoters, including all the known tissue-specific HLH genes (35,49,50). Consistent with other TATA-less promoters, ME1 contains several Sp1 response elements proximal to its transcriptional start site. Sp1 proteins are strong transcriptional enhancers that play a central role in directing the activity of

TATA-less promoters (45–48). The observation that a region containing three Sp1 sites was required for ME1 promoter activity suggests that Sp1 proteins also play an important role in regulating ME1 gene expression.

The most striking feature of the ME1 promoter is the initiator element which contains a 13 bp poly d(A) tract flanked by a 16 bp palindrome and a 9 bp inverted repeat. The observation that transcription initiated predominantly from the center of the poly d(A) tract suggests that this region is a strong initiator. This is based upon the TATA-less mouse thymidylate synthase promoter which appears to utilize multiple start sites due to the lack of a strong initiator element (58).

It is intriguing that putative DNA binding elements flank the poly d(A) tract in the ME1 initiator. Recent evidence indicates that the TATA-less terminal deoxynucleotidyl transferase (TdT) initiator element can independently direct the assembly of a stable pre-initiation complex by recruiting one or several TATA box-binding (TBP)-associated factors (TAFs) (59). The ME1 initiator and its proximal sequences supported basal levels of transcription in PCC7 and Cos-7 cells (Fig. 4, mutant Δ -826/-47), suggesting that this region could act in a similar manner. Consistent with this idea, the ME1 initiator was recognized by numerous protein complexes in PCC7 and Swiss 3T3 cell extracts. Whether these proteins are part of the general transcriptional machinery or represent novel DNA binding proteins is currently under investigation.

Our results also support a role for the ME1 initiator in cell type-specific gene expression. This is based on three lines of evidence. First, the ME1 initiator and its proximal sequences supported basal levels of transcription in PCC7 cells, but not in Swiss 3T3 cells (Fig. 4, mutant Δ -826/-47); secondly, deletion of the initiator element in a constitutive ME1 promoter construct attenuated promoter activity only in PCC7 cells (Fig. 6, mutant Δ -32/+58); and thirdly, a synthetic ME1 initiator fragment (-32 bp to +22 bp) bound protein complexes in PCC7 cell extracts that were absent in Swiss 3T3 extracts (Fig. 7). Collectively, our results support a model in which a cell type-specific protein(s) (i.e. TAF) enhances transcription in PCC7 cells by acting through the ME1 initiator. This is also consistent with the relatively high levels of endogenous ME1 mRNA observed in PCC7 cells (although at this point we cannot rule out the possibility that other factors also contribute to this observation).

The ME1 initiator is unusual not only because its sequence is unrelated to other initiators (47,60), but also because it alludes to an unconventional mechanism of transcriptional initiation. While this mechanism remains unclear, the sequence of the ME1 initiator element provides some subtle, yet informative, clues. Poly d(A) tracts are well known for their ability to induce static bends in DNA (61–64) that facilitate DNA–protein and protein–protein interactions in DNA regulatory domains (65–67). Importantly, poly d(A) induced bends modulate transcriptional activity in a number of gene promoters (68–70). While we are unaware of a poly d(A) tract that has been associated with transcriptional initiation, DNA bending appears to be an integral part of the initiation process. This was demonstrated most dramatically by Rees *et al.*, who provided direct evidence for DNA bending in an open complex and also during elongation using scanning force microscopy (71). In addition, van der Vliet and Verrijzer propose that the energy stored in bent DNA can be used to form an open complex and to initiate elongation (66). Considering the nature of

the ME1 initiator, it is plausible that bending of the 13 bp poly d(A) tract (either static or protein-induced) may be associated with either or both of these processes. Clearly, it will be important to determine the overall structure of the ME1 initiator and its mechanism of initiating transcription.

Other elements in the ME1 promoter included a strong repressor region and several E-boxes. The repressor was localized to a 256 bp region (-570 to -314 bp) that attenuated promoter activity between 4- and 5-fold in all cell lines tested. The two E-boxes located within this region (-396 and -330 bp) may contribute to repression since E-boxes have been shown to support repressor activity in some cases (72–74). However, activity of the -570 ME1 promoter construct, which contains both of these E-boxes, was enhanced upon over-expression of the ME1 protein suggesting that repression occurs through a different *cis*-element. On the other hand, we cannot rule out the possibility that ME1 homodimers out-competed a repressor that normally acts through one or both of these E-boxes.

The cluster of four E-boxes in the ME1 proximal promoter (-396, -330, -251 and -240 bp) appear to be functional since the ME1 protein modulated the activity of its promoter. However, it is unclear why the ME1 protein both enhanced and repressed promoter activity depending upon where the promoter was truncated. While we can only speculate on this observation, it is likely that differences in the complement of proteins bound to each construct contributed to this anomaly. Also, over-expression of the ME1 protein may have influenced promoter activity by saturating the E-boxes or sequestering endogenous transcription factors. Although tissue-specific HLH genes are cross-regulated by other HLH proteins (35–37), this is the first indication that nontissue-specific HLH genes are regulated by other HLH proteins. Recent evidence suggests that E2A is not autoregulated or cross-regulated by other E-proteins (E2-2 or ME1) during B cell differentiation (29). Whether this represents an important difference between E2A and ME1 gene regulation remains to be determined.

At this point, we do not know whether the *cis*-elements identified in this study are sufficient to properly direct ME1 gene expression. Expression of the MyoD gene is influenced by tissue-specific elements 5 and 20 kb upstream of the transcriptional start site (33,34), and an enhancer was identified >50 kb downstream of the initiator in the TATA-less Wilms' tumor gene (75). Additional factors such as mRNA stability or differential splicing could also play a role in defining ME1 gene expression (76). Thus, the role of regulatory elements identified in this study will require further analysis *in vivo* (i.e., transgenic mice).

ACKNOWLEDGEMENTS

We express our deepest sympathy to the family and friends of Dr Mauricio X. Zuber, who lost his life in a rock climbing accident June 21, 1994. Mauricio was an inspiration to all who knew him and his guidance, support and enthusiasm were essential to the work described here. We thank Drs Jim Bamberg and Collin Clay for Swiss 3T3 and Cos-7 cells, respectively, and Drs Jim Bamberg, Collin Clay and Jakyoun Yoo for helpful discussions and reading of the manuscript. This work was supported by March Dimes Basil O'Connor Starter Grant FY94-0835, NSF grant BNS-9158411, and NIH grant NS33804.

REFERENCES

- 1 Jan, N.Y. and Jan, L.Y. (1993) *Cell*, **75**, 827–830.
- 2 Weintraub, H. (1993) *Cell*, **75**, 1241–1244.
- 3 Church, G.M., Ephrussi, A., Gilbert, W. and Tonegawa, S. (1985) *Nature*, **313**, 798–801.
- 4 Ephrussi, A., Church, G.M., Tonegawa, S. and Gilbert, W. (1985) *Science*, **227**, 134–140.
- 5 Murre, C., McCaw, P.S., Vaessin, H., Caudy, M., Jan, L.Y., Jan, Y.N., Cabrera, C.V., Buskin, J.N., Hauschka, S.D., Lassar, A.B., Weintraub, H. and Baltimore, D. (1989) *Cell*, **58**, 537–544.
- 6 Davis, R.L., Weintraub, H. and Lassar, A.B. (1987) *Cell*, **51**, 987–1000.
- 7 Pinney, D.F., Pearson-White, S.H., Konieczny, S.F., Latham, K.E. and Emerson, C.P. (1988) *Cell*, **53**, 781–793.
- 8 Rudnicki, M.A., Schnegelsberg, P.N.J., Stead, R.H., Braun, T., Arnold, H.H. and Jaenisch, R. (1993) *Cell*, **75**, 1351–1359.
- 9 Tapscott, S.J., Davis, R.L., Thayer, M.J., Cheng, P., Weintraub, H. and Lassar, A.B. (1988) *Science*, **242**, 405–411.
- 10 Brand, M., Jarman, A.P., Jan, L.Y. and Jan, Y.N. (1993) *Development*, **119**, 1–17.
- 11 Parkhurst, S.M., Bopp, D. and Ish-Horowicz, D. (1990) *Cell*, **63**, 1179–1191.
- 12 Rodriguez, I., Hernandez, R., Modolell, J. and Ruiz-Gomez, M. (1990) *EMBO J.*, **9**, 3583–3592.
- 13 Murre, C., McCaw, P.S. and Baltimore, D. (1989) *Cell*, **56**, 777–783.
- 14 Henthorn, P., Kiledjian, M. and Kadesch, T. (1990) *Science*, **247**, 467–470.
- 15 Neuman, T., Keen, A., Knapik, E., Shain, D., Ross, M., Nornes, H.O. and Zuber, M.X. (1993) *Eur. J. Neurosci.*, **5**, 311–318.
- 16 Hu, J., Olson, E. and Kingston, R.E. (1992) *Mol. Cell. Biol.*, **12**, 1031–1042.
- 17 Klein, E.S., Simmons, D.M., Swanson, L.W. and Rosenfeld, M.G. (1993) *Genes Dev.*, **7**, 55–71.
- 18 Caudy, M., Vassin, H., Brand, M., Tuma, R., Jan, L.Y. and Jan, Y.N. (1988) *Cell*, **55**, 1061–1067.
- 19 Lassar, A.B., Davis, R.L., Wright, W.E., Kadesch, T., Murre, C., Voronova, A., Baltimore, D. and Weintraub, H. (1991) *Cell*, **66**, 305–315.
- 20 Bain, G., Gruenwald, S. and Murre, C. (1993) *Mol. Cell. Biol.*, **13**, 3522–3529.
- 21 Nielsen, A.L., Pallisgaard, N., Pedersen, F.S. and Jorgensen, P. (1992) *Mol. Cell. Biol.*, **12**, 3449–3459.
- 22 Aronheim, A., Shiran, R., Rosen, A. and Walker, M.D. (1993) *Nucleic Acids Res.*, **21**, 1601–1606.
- 23 Roberts, V.J., Steenbergen, R. and Murre, C. (1993) *Proc. Natl. Acad. Sci. USA*, **90**, 7583–7587.
- 24 Soosar, A., Chiaramello, A., Zuber, M.X. and Neuman, T. (1994) *Mol. Brain Res.*, **25**, 176–180.
- 25 Suda, K., Nornes, H.O. and Neuman, T. (1994) *Neurosci. Lett.*, **177**, 87–90.
- 26 Zhuang, Y., Kim, C., Bartelmez, S., Cheng, P.-F., Groudine, M. and Weintraub, H. (1992) *Proc. Natl. Acad. Sci. USA*, **89**, 12132–12136.
- 27 Zhuang, Y., Kim, C., Bartelmez, S., Cheng, P.-F., Groudine, M. and Weintraub, H. (1992) *Proc. Natl. Acad. Sci. USA*, **89**, 12132–12136.
- 28 Bain, G., Robanus Maandag, E.C., Izon, D.J., Amsen, D., Kruisbeek, A.M., Weintraub, B., Krop, I., Schlissel, M.S., Feeney, A.J., van Roon, M., van der Valk, M., Riele, H.P.J., Berns, A. and Murre, C. (1994) *Cell*, **79**, 885–892.
- 29 Zhuang, Y., Soriano, P. and Weintraub, H. (1994) *Cell*, **79**, 875–884.
- 30 Benezra, R., Davis, D.L., Lockshon, D., Turner, D.L. and Weintraub, H. (1990) *Cell*, **61**, 49–59.
- 31 Lassar, A.B., Thayer, M.J., Overell, R.W. and Weintraub, H. (1989) *Cell*, **58**, 659–667.
- 32 Cline, T.W. (1993) *Trends Genet.*, **9**, 385–390.
- 33 Goldhamer, D.J., Faerman, A., Shani, M. and Emerson, C.P. (1992) *Science*, **256**, 538–542.
- 34 Tapscott, S.J., Lassar, A.B. and Weintraub, H. (1992) *Mol. Cell. Biol.*, **12**, 4994–5003.
- 35 Edmondson, D.G., Cheng, T., Cserjesi, P., Chakraborty, T. and Olson, E.N. (1992) *Mol. Cell. Biol.*, **12**, 3665–3677.
- 36 Martinez, C. and Modolell, J. (1991) *Science*, **251**, 1485–1487.
- 37 Martinez, C., Modolell, J. and Garrell, J. (1993) *Mol. Cell. Biol.*, **13**, 3514–3521.
- 38 Sambrook, J., Fritsch, E.F. and Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
- 39 Shain, D.H., Yoo, J., Slaughter, R.G., Hayes, S.E. and Ji, T.H. (1992) *Anal. Biochem.*, **200**, 47–51.
- 40 Chomczynski, P. and Sacchi, N. (1987) *Anal. Biochem.*, **162**, 156–159.
- 41 Graham, F.L. and van der Eb, A.J. (1973) *Virology*, **52**, 456–467.
- 42 Neuman, T., Metsis, M., Persson, H. and Gruss, P. (1993) *Mol. Brain Res.*, **20**, 199–208.
- 43 Pfeiffer, S.E., Jakob, H., Mikoshiba, K., Dubois, P., Guenet, J.L., Nicolas, J.-F., Gaillard, J., Chevance, G. and Jakob, F. (1991) *J. Cell Biol.*, **88**, 57–66.
- 44 Shain, D.H. and Zuber, M.X. (unpublished).
- 45 Araki, E., Murakami, T., Shirotani, T., Kanai, F., Shinohara, Y., Shimada, F., Mori, M., Shichiri, M. and Ebina, Y. (1991) *J. Biol. Chem.*, **266**, 3944–3948.
- 46 Boisclair, Y.R., Brown, A.L., Casola, S. and Rechler, M.M. (1993) *J. Biol. Chem.*, **268**, 24892–24901.
- 47 Kollmar, R., Sukow, K.A., Sponagle, S.K. and Farnham, P.J. (1994) *J. Biol. Chem.*, **269**, 2252–2257.
- 48 Pugh, B.F. and Tjian, R. (1991) *Genes Dev.*, **5**, 1935–1945.
- 49 Zingg, J., Alva, P. and Jost, J. (1991) *Nucleic Acids Res.*, **19**, 6433–6439.
- 50 Begley, C.G., Robb, L., Rockman, S., Visvader, J., Bockamp, E.O., Chan, Y.S. and Green, A.R. (1994) *Gene*, **138**, 93–99.
- 51 Sun, X. and Baltimore, D. (1991) *Cell*, **64**, 459–470.
- 52 Green, A.R. and Begley, C.G. (1991) *J. Cell Cloning*, **10**, 269–276.
- 53 Inaba, T., Roberts, W.N., Shapiro, L.H., Jolly, K.W., Raimondi, S.C., Smith, S.D. and Look, A.T. (1992) *Science*, **257**, 531–534.
- 54 Kamps, M.P., Murre, C., Sun, X. and Baltimore, D. (1990) *Cell*, **60**, 547–555.
- 55 Mellentin, J.D., Smith, S.D. and Cleary, M.L. (1989) *Cell*, **58**, 77–83.
- 56 Nourse, J., Mellentin, J.D., Galili, N., Wilkinson, J., Stanbridge, E., Smith, S.D. and Cleary, M.L. (1990) *Cell*, **60**, 535–545.
- 57 Xia, Y., Brown, L., Yang, C., Tsan, J.T., Siciliano, M.J., Espinosa, R., Le Beau, M.M. and Baer, R.J. (1991) *Proc. Natl. Acad. Sci. USA*, **88**, 11416–11420.
- 58 Geng, Y. and Johnson, L.F. (1993) *Mol. Cell. Biol.*, **13**, 4894–4903.
- 59 Martinez, E., Chiang, C.M., Ge, H. and Roeder, R.G. (1994) *EMBO J.*, **13**, 3115–3126.
- 60 Smale, S.T. and Baltimore, D. (1989) *Cell*, **57**, 103–113.
- 61 Koo, H., Drak, J., Rice, J.A. and Crothers, D.M. (1990) *Biochemistry*, **29**, 4227–4234.
- 62 Levene, S.D., Wu, H. and Crothers, D.M. (1986) *Biochemistry*, **25**, 3988–3995.
- 63 Nelson, H.C.M., Finch, J.T., Luisi, B.F. and Klug, A. (1987) *Nature*, **330**, 221–226.
- 64 Zinkel, S.S. and Crothers, D.M. (1987) *Nature*, **328**, 178–181.
- 65 Ryder, K., Silver, S., DeLucia, A.L., Fanning, E. and Tegtmeyer, P. (1986) *Cell*, **44**, 719–725.
- 66 van der Vliet, P.C. and Verrijzer, C.P. (1993) *Bioessays*, **15**, 25–32.
- 67 Zahn, K. and Blattner, F.R. (1987) *Science*, **236**, 416–422.
- 68 Ertl, A.P. and Gill, G.N. (1993) *Gene*, **136**, 361–364.
- 69 Highlander, S.K. and Weinstock, G.M. (1994) *DNA Cell Biol.*, **13**, 171–181.
- 70 Toyonaga, T., Matsushita, O., Katayama, S., Minami, J. and Okabe, A. (1992) *Microbiol. Immunol.*, **36**, 603–613.
- 71 Rees, W.A., Keller, R.W., Vesenka, J.P., Yang, G. and Bustamante, C. (1993) *Science*, **260**, 1646–1649.
- 72 Giacca, M., Gutierrez, M.I., Menzo, S., D'Adda Di Fagnana, F. and Falaschi, A. (1992) *Virology*, **186**, 133–147.
- 73 Leshkowitz, D., Aronheim, A. and Walker, M.D. (1992) *DNA Cell Biology*, **11**, 549–558.
- 74 Ruezinsky, D., Beckman, H. and Kadesch, T. (1991) *Genes Dev.*, **5**, 29–37.
- 75 Frazier, G.C., Wu, Y.J., Hewitt, S.M., Maity, T., Ton, C.C., Huff, V. and Saunders, G.F. (1994) *J. Biol. Chem.*, **269**, 8892–8900.
- 76 Darnell, J.E., Jr. (1982) *Nature*, **297**, 365–371.