FEBS Letters 440 (1998) 332–336 FEBS 21170

Selective conservation of an E-protein gene promoter during vertebrate evolution

Daniel H. Shain[a,*], Mauricio X. Zuber[1,a], Jeff Norris[a], Jakyoung Yoo[2,c], Toomas Neuman[3,b]

[a] *Department of Biochemistry and Molecular Biology, Colorado State University, Fort Collins, CO 80523, USA*
[b] *Department of Anatomy and Neurobiology, Colorado State University, Fort Collins, CO 80523, USA*
[c] *U.S. Department of Agriculture, Agricultural Research Service, Roman L. Hruska U.S. Meat Animal Research Center (MARC), P.O. Box 166, Clay Center, NE 68933-0166, USA*

Received 8 October 1998

Abstract The murine E-protein gene ME1 encodes a non-tissue-specific, helix-loop-helix transcription factor that is associated with morphological development. ME1 gene expression is regulated by a TATA-less promoter that contains multiple Sp1 consensus elements, E-boxes, and a novel transcription initiation site. In this study, we compared DNA homologous to the ME1 promoter from vertebrate species ranging from frog to human. A region of striking sequence similarity was identified in a region corresponding to the ME1 transcription initiation site (ME1 Inr). Within this region, a poly d(A) tract and a 9-bp inverted repeat (5′-GTCCGCCTG) were highly conserved in all species that were examined. Protein complexes that recognized these DNA elements were present among distant vertebrates (frog, chick, monkey and human), and were able to bend the ME1 Inr to a similar extent (∼60°) as the previously described murine MBPα and MBPβ proteins. Collectively, these results suggest that an ME1 Inr-like element and its associated proteins functioned in an ancestral vertebrate more than 350 million years ago.

Key words: E-protein; Promoter; Vertebrate evolution; Gene expression; Mouse; DNA bending

1. Introduction

E-proteins comprise a subfamily of helix-loop-helix transcription factors that play a fundamental role in cell differentiation and development (for review, see [1,2]). Three classes of E-protein genes (E2A [3], E2-2 [4] and HEB [5]) have been identified among vertebrates, and appear to be descendants of the ancestral *Drosophila daughterless* gene [6,7]. Although there is considerable redundancy within the E-protein family with respect to expression and function, vertebrate E-proteins genes display clearly different expression patterns during development [8–10] and appear to have unique functional roles. For example, products of the E2A gene are required for B cell formation and immunoglobulin rearrangements [11,12].

Expression of the murine ME1 gene (counterpart of human HEB) is enriched in morphogenetically active regions during development [8,10], and in the hippocampus and olfactory epithelium of the adult mouse [10]. Analysis of the ME1 proximal promoter has revealed several response elements including a novel initiator, four E-box consensus sites, and a strong repressor [13]. In an effort to identify regions of the ME1 promoter that have been functionally conserved over the course of vertebrate evolution, we have isolated counterpart genomic DNA from frog, rat and human and compared this to the mouse sequence. Our results indicate that a region proximal to the ME1 transcription initiation site is strikingly conserved, and this region recruits a similar complement of proteins which bends the ME1 Inr ∼60° in all the vertebrate species examined.

2. Materials and methods

2.1. Cloning genomic DNA

Genomic clones from human were obtained by screening a human genomic library (Stratagene) at low stringency with a DNA fragment containing the ME1 Inr (−45 to +121 bp in the ME1 promoter [13]). Conditions of hybridization were: 25% formamide, 5× SSPE, 5% SDS at 42°C; washes were in 0.5× SSC at 55°C. The REB genomic fragment was amplified from rat genomic DNA (generous gift from Don Marsh) using the primers 5′-GGCGCGGAGGGATCCGGA and 5′-CGGTCCCCCCGAATAGAAC, which were conserved between the HEB and ME1 sequences. Cycles of PCR were: 92°C, 1 min; 56°C, 1 min; 72°C, 30 s for 30 cycles. *Xenopus* genomic clones were isolated from an *X. laevis* genomic library (gift from Angie Ribera) using the 5′-end of XE1 cDNA [7] as a probe, which was obtained by RACE-PCR (CloneTech). Conditions of hybridization were: 50% formamide, 5× SSPE, 5% SDS at 42°C; washes were in 0.2× SSC at 65°C.

2.2. RNase protection analysis

Total RNA was purified as described previously [14]. HEB template DNA was linearized with either *Bam*HI (antisense) or *Sac*I (sense) and transcribed in vitro using T7 and T3 RNA polymerase, respectively, in the presence of [α-^{32}P]UTP as described [15]. Approximately 20 μg of total RNA and 10^4–10^5 cpm 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. Protected RNA fragments were separated on a 6% polyacrylamide-40% urea gel visualized by autoradiography using Hyperfilm-MP (Amersham).

2.3. Electrophoretic mobility shift assays (EMSA)

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

*Corresponding author. Present address: 385 LSA Department of Molecular and Cell Biology, University of California, Berkeley, CA 94720, USA. Fax: (1) (510) 643-6791.
E-mail: shain@uclink4.berkeley.edu

[1] Deceased.

[2] Present address: 491 LSA, Department of Molecular and Cell Biology, University of California, Berkeley, CA 94720, USA.

[3] Present address: Department of Surgery, Cedars-Sinai Medical Center, Davis Building, Room 4018, 8700 Beverly Boulevard, Los Angeles, CA 90048, USA.

PII: S0014-5793(98)01417-3

of protease inhibitors: PMSF (0.5 mM), leupeptin (0.5 μg/ml), pepstatin (0.7 μg/ml), aprotinin (1 μg/ml) and bestatin (40 μg/ml). All steps were performed at 4°C. After sonication, extracts were cleared by microcentrifugation for 5 min. Binding conditions for EMSA were: 10 mM HEPES, pH 7.8, 1 mM spermidine, 5 mM $MgCl_2$, 50 mM KCl, 0.5 mM DTT, 9% glycerol, 0.8 mg poly d(I-C), 100 000 cpm of labeled oligonucleotide (cold competitor DNA was used at a 200-fold molar excess), and 10 μg of nuclear extract. Following incubation for 15 min at 37°C, reactions were applied to a polyacrylamide gel, electrophoresed to separate DNA-protein complexes and analyzed by autoradiography.

2.4. Gel permutation analysis

The ME1/pBend2 vector containing the ME1 Inr insert was digested with the appropriate restriction endonucleases (shown in Fig. 3). DNA fragments were then filled in with [^{32}P]-αdCTP as described [16], electrophoresed on a 5% polyacrylamide gel, and visualized by autoradiography. Relative mobilities were measured from a point originating at the loading well.

2.5. Cell culture

PCC7, Cos-7, U373 and Colo320 cells 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. Transient transfections were performed by calcium phosphate co-precipitation [17]; transfection efficiencies were determined by β-gal assays, as described [15].

3. Results and discussion

3.1. Cloning of ME1 counterpart DNA from diverse vertebrates

Different strategies were used to obtain human (HEB), rat (REB) and frog (XE1.1 and XE1.2) counterpart DNA. HEB was isolated from a human genomic library (Stratagene) using the 5′-region of ME1 as a probe. A fragment of the REB promoter was cloned by PCR using sequences on either side of the poly d(A) tract that were conserved between ME1 and HEB. Although this primer set was used successfully in all mammals that were tested, we were unable to amplify ME1 homologues from more distant vertebrates including chicken and frog (*Xenopus laevis*). Attempts to clone these genomic fragments by low stringency hybridization were also unsuccessful. We then focused on *X. laevis* since it represented the most distant species from the mammalian sequences (ME1, REB and HEB). To clone *X. laevis* DNA, the 5′-end of an XE1 cDNA [7] was obtained by RACE-PCR (rapid amplification of cDNA ends by PCR) and was used to screen a *Xenopus* genomic library. Fortuitously, two independent clones (XE1.1 and XE1.2) were identified by this analysis. XE1.1 and XE1.2 appear to be pseudoalleles that arose

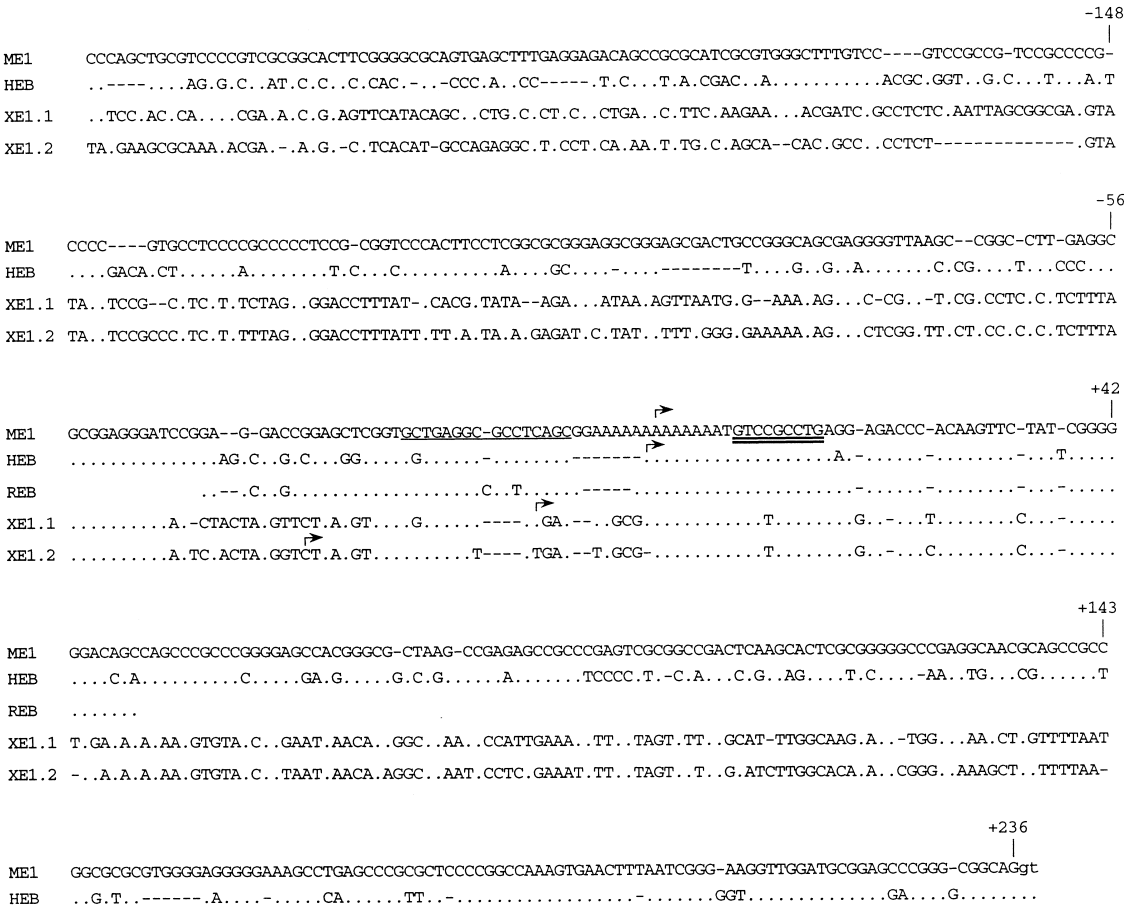

Fig. 1. Nucleotide comparison of the ME1 promoter and homologous DNA from other vertebrates. Partial sequences of ME1 (mouse), HEB (human), REB (rat), XE1.1 (*Xenopus*) and XE1.2 (*Xenopus*) are shown. The ME1 sequence is shown in its entirety while differences in other sequences are indicated. Dots represent sequence identity and dashes indicate gaps. Numbering begins at the ME1 transcription start site, which occurs at the center of the 13-bp poly d(A) tract. A 16-bp palindrome and a 9-bp inverted repeat flanking the ME1 transcription start site are underlined and double underlined, respectively. Arrows indicate the putative transcription start site in each sequence (see text). Lower case letters represent intron sequences.

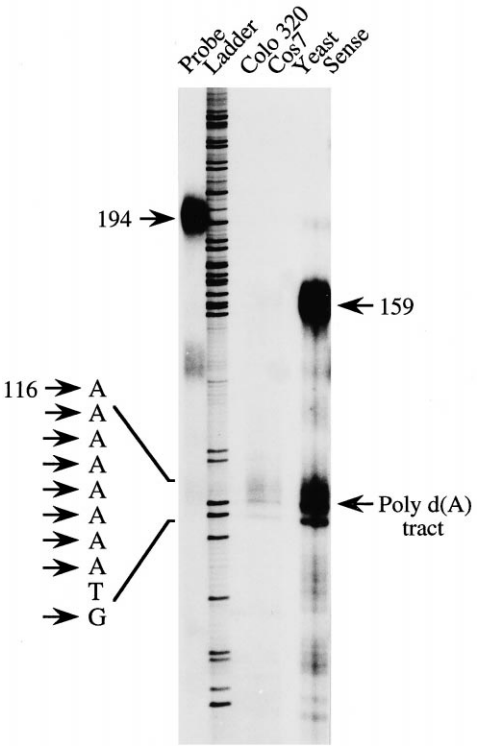

Fig. 2. Identification of the HEB transcription start site by RNase protection analysis. Total RNA (10 μg) was annealed with a 194-bp HEB antisense RNA. The largest protected fragment in Colo 320 (human) and Cos-7 (monkey) cells mapped to the 5′ end of an 8-bp poly d(A) tract. No protected bands were observed in yeast RNA. A sense strand generated a protected fragment of the expected size (159 bp), and also a smaller, hypersensitive site that mapped to the 8-bp poly d(A) tract.

from *Xenopus* tetraploidization approximately 30 million years ago [7,18].

The nucleotide sequences of the ME1 promoter [13], and the putative HEB, REB, XE1.1 and XE1.2 promoters were aligned with Clustal V software [19] and adjusted by eye (Fig. 1). A region of striking sequence similarity was observed in a 96-bp region from −54 bp to +42 bp of the ME1 promoter. Within this region, the ME1 gene was 82% identical to HEB, 88% identical to REB (within a region between −41 and +30), 69% identical to XE1.1 and 67% identical to XE1.2. This should be compared to the similarity of the basic, helix-loop-helix (bHLH) region which encodes the highly conserved DNA binding and dimerization motifs [7]. Within the bHLH region, ME1 sequence identity ranges from 81% in XE1.2 to 95% in REB. Clearly, stringent evolutionary pressures have acted upon both the regulatory and coding sequences of these genes.

A poly d(A) tract was identified in all genes and ranged in size from 7 bp in XE1.2 to 13 bp in ME1. Transcription appeared to initiate within or proximal to the poly d(A) tract in all genes examined (indicated by arrows). The ME1 transcription start site has been described previously [13], HEB transcription initiation was determined by RNase protection analysis (see Fig. 2 below), XE1.2 by RACE-PCR (described above), and the putative start site of XE1.1 represents the most 5′ cDNA clone from an *X. laevis* random-primed cDNA library (Stratagene). It should be noted that the most 5′ XE1.2 cDNA clone from the same library was within 5 bp of the longest XE1.2 RACE-PCR product. Based upon the location of the putative transcription start sites, the region downstream of the poly d(A) tracts comprises the 5′-untranslated region in all species.

The 9-bp inverted repeat 5′-GTCCGCCTG-3′ immediately downstream of the poly d(A) tract was absolutely conserved in ME1, HEB and REB, and contained a single mismatch in both XE1.1 and XE1.2 (T→C conversion at the fourth position). This region constitutes the MBPβ binding site in the ME1 Inr which is susceptible to DNA bending ([20] see below). Sequences upstream of poly d(A) tract were less conserved but displayed regions of apparent homology. This was evident in a 10-bp sequence 5′-GCGGAGGGAT-3′ (from −54 bp to −44 bp in ME1) which was present in all species. The 16-bp palindrome in ME1 was conserved among mammals (ME1, HEB and REB), but only partially conserved in the *Xenopus* genes.

Sequences flanking the region from −54 bp to +42 bp displayed reduced sequence identity. In the *Xenopus* genes, there was no apparent homology to the mammalian sequences. It is noteworthy, however, that both XE1.1 and XE1.2 contained Sp1 consensus sites and E-boxes within their proximal promoters (similar to ME1). The HEB promoter was clearly re-

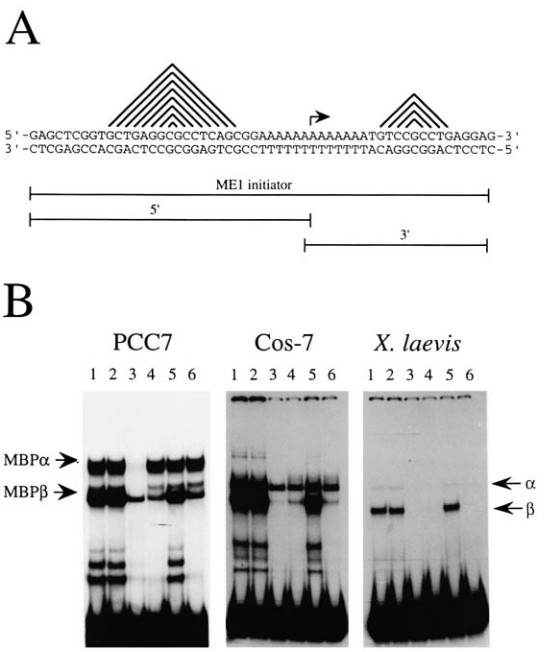

Fig. 3. ME1 Inr binding proteins from distant vertebrates. A: A synthetic oligonucleotide comprising the ME1 Inr and flanking restriction endonuclease sites is shown [20]. The 16-bp palindrome upstream from the poly d(A) tract and the downstream 9-bp inverted repeat are indicated. The arrow designates the transcription start site [20]. Oligonucleotides used in this experiment are indicated below. B: EMSA analysis of the ME1 Inr with nuclear extracts from PCC7 (mouse), Cos-7 (monkey) and *X. laevis* (stage 28 embryos). The ME1 Inr oligonucleotide was [α-^{32}P]dCTP labeled and was present in each lane. Cold competitor DNAs in each panel were: lane 1, none; lane 2, random 26mer; lane 3, ME1 Inr; lane 4, 3′ oligo; lane 5, 5′ oligo; lane 6, 5′+3′ oligos. Competitor DNA was used at a 200-fold molar excess. The predominant protein complexes that bound the ME1 Inr in PCC7, designated MBPα and MBPβ, are indicated [20]. The putative homologues of these complexes in Cos-7 cells and *X. laevis* are labeled α and β, respectively.

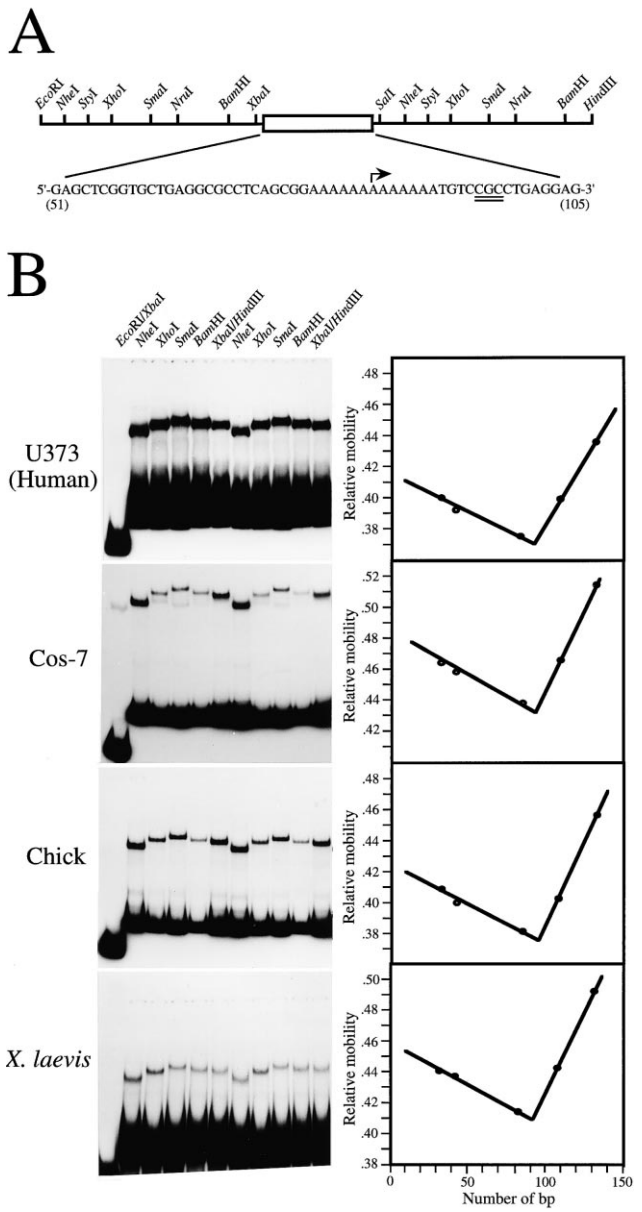

Fig. 4. Protein-induced bending of the ME1 Inr by protein complexes from distant vertebrates. A: Vector map of the ME1 Inr oligonucleotide cloned into the *Xba*I and *Sal*I sites of pBend2 [20]. Restriction endonuclease sites used for the bending analysis are indicated; numbering initiates at the *Eco*RI site. B: Circular permutation analysis of the ME1 Inr in the presence of nuclear extracts from U373 (human), Cos-7 (monkey), chick (embryonic day 12) and *X. laevis* (stage 28) cells. The left panel shows restriction fragments from A, electrophoresed on a native 5% polyacrylamide gel. The *Eco*RI-*Xba*I fragment was included as a negative control since it lacks the ME1 Inr sequence. The right panel plots the relative mobility against the distance from the *Eco*RI site. Putative α and β protein complexes, which were not resolved by this analysis, bend the ME1 Inr ~60° according to the equation $\mu M/\mu E = \cos\alpha/2$, where μM is the distance of bound probe, μE is the distance of unbound probe and α is the angle of bending [21]. The bend mapped to the center of the inverted repeat 5′-GTCCGCCTG-3′ (shown in A as a double underlined sequence).

lated to ME1, but lacked several response elements that were present in ME1 [13]. For example, HEB did not contain proximal E-boxes and shared little sequence similarity to a strong repressor region in the ME1 promoter (sequence not shown). The 5′-untranslated regions of ME1 and HEB were closely related and both appear to utilize the same splice site at the first intron-exon boundary [13].

3.2. HEB transcription initiation

The putative transcriptional start site of the HEB gene was determined by RNase protection analysis (Fig. 2). An HEB antisense riboprobe was synthesized from the *Sac*I to the *Bam*HI restriction endonuclease site of the HEB promoter. As a positive control, a sense RNA strand was synthesized in the opposite orientation. RNase protection analysis of these two RNAs yielded a fragment of the expected size (159 bp) and also a smaller set of protected bands which mapped to an 8-bp poly d(A) tract in the HEB gene. This hypersensitive site probably resulted from 'breathing' of the poly d(A) tract since hydrogen bonding is relatively weak within A-T rich stretches.

In both human (Colo 320) and monkey cells (Cos-7), protected bands mapped to the HEB 8-bp poly d(A) tract. The pattern observed is essentially identical to that described for the ME1 gene, which initiates transcription within a 13-bp poly d(A) tract [13]. While there are several models that could explain this banding pattern, we propose that transcription initiation in the HEB gene occurs at a site proximal to the 5′-end of the poly d(A) tract. Since this is based partially upon the similarity between the HEB and ME1 promoters (in both sequence and RNase protection results), we cannot rule out other possibilities. For example, transcription could initiate at multiple locations within the HEB poly d(A) tract or slightly upstream from the poly d(A) tract, and a similar pattern of protected bands may result. It seems clear, however, that transcription does not initiate at a site upstream from the *Bam*HI site (−43 bp in the HEB promoter) since no protected bands in either Colo 320 or Cos-7 lanes corresponded to the full length probe (compare with the 159-bp band in the sense lane). Thus, the primary transcription start site of HEB is likely to occur in a region proximal to the 8-bp poly d(A) tract, which is similar to that observed in the mouse ME1 promoter [13].

3.3. Identification of protein-binding complexes that recognize the ME1 Inr

There are at least seven protein complexes in mouse PCC7 cells that recognize the ME1 Inr [13,20]. Due to the high degree of similarity observed between the ME1 Inr and its counterpart DNA in other vertebrates, we were interested to know whether protein extracts from other species could also recognize ME1 Inr DNA. Electrophoretic mobility shift assays (EMSA) using extracts from Cos-7 cells generated a pattern of bands that was highly reminiscent of that observed in mouse (Fig. 3; lanes 1, 2). At least seven Cos-7 protein complexes bound an oligonucleotide containing the ME1 Inr, and two of these displayed characteristics similar to the MBPα and MBPβ complexes identified in mouse [20], based upon competition analysis (Fig. 3, lanes 4–6). Specifically, the putative α complex was competed by the ME1 Inr oligo, and the putative β complex was competed by a 3′ oligo containing the 9-bp inverted repeat. The putative α complex in Cos-7 cells was also competed to some extent by the 3′ oligo.

EMSA analysis using *Xenopus laevis* protein extracts (stage 28) generated a pattern quite different from that observed in mouse or Cos-7 cells (Fig. 3). Only two protein complexes were observed in *Xenopus* extracts and these were detected

at significantly reduced levels. Nevertheless, the two bands may be related to the mouse MBPα and MBPβ complexes based upon their behavior when competed with different regions of the ME1 Inr. Specifically, neither of these complexes were competed by the 5′ region of the ME1 Inr while the putative β complex in *Xenopus* was competed by the 3′ oligo. The putative α complex in *Xenopus* was competed by the 3′ oligo which differs from that observed in mouse.

Upon identifying protein complexes from different species that bound the ME1 Inr (Fig. 3), we then tested whether they could bend the ME1 Inr as do protein extracts from mouse cells which introduce a ∼60° bend within the 9-bp inverted repeat [20]. Using circular permutation analysis [21] with fragments of an ME1 Inr/pBend2 construct [20], protein complexes from human, monkey, chicken, and *Xenopus* cells also bent the 9-bp inverted repeat ∼60° (Fig. 4). This suggests that protein complexes related to the mouse MBPα and MBPβ complexes are present throughout the vertebrate lineage and, considering the proximity and conservation of the putative MBPα and MBPβ binding sites (the poly d(A) tract and 9-bp inverted repeat, respectively [20]), these protein complexes may have a related function among the vertebrate species examined.

Collectively, these results demonstrate the selective conservation among the ME1 promoter and its homologous DNA in diverse vertebrates. Although several general features were present in all the promoters examined (e.g. Sp1 consensus sites, absence of a TATA-box), most striking is the apparent maintenance of an ME1 Inr-like sequence among vertebrates as distant as frog and human. It is significant that the ME1 Inr and homologous DNA from other species appears to be coincident with transcription initiation in each species examined. That proteins from various vertebrates can recognize, and bend the ME1 Inr, supports the notion that homologous α and β protein complexes are present throughout the vertebrate lineage. Taken together, these results suggest that an ME1 Inr-like element, in conjunction with its associated protein complexes (MBPα and MBPβ), may have functioned as a transcriptional initiation unit in an ancestral organism that predated terrestrial vertebrates.

Acknowledgements: We thank Nicole Newell for critical reading of the manuscript. This work was supported by March of Dimes Basil O'Conner Starter Grant FY94-0835, NSF Grant BNS-0158411, and NIH grant NS33804 to M.X.Z.

References

[1] Jan, N.Y. and Jan, L.Y. (1993) Cell 75, 827–830.
[2] Weintraub, H. (1993) Cell 75, 1241–1244.
[3] Murre, C., McCaw, P.S. and Baltimore, D. (1989) Cell 56, 777–783.
[4] Henthorn, P., Kiledjian, M. and Kadesch, T. (1990) Science 247, 467–470.
[5] Hu, J., Olson, E. and Kingston, R.E. (1992) Mol. Cell. Biol. 12, 1031–1042.
[6] Caudy, M., Vassin, H., Brand, M., Tuma, R., Jan, L.Y. and Jan, Y.N. (1988) Cell 55, 1061–1067.
[7] Shain, D.H., Neuman, T. and Zuber, M.X. (1997) Genetics 146, 345–353.
[8] Klein, E.S., Simmons, D.M., Swanson, L.W. and Rosenfeld, M.G. (1993) Genes Dev. 7, 55–71.
[9] Roberts, V.J., Steenbergen, R. and Murre, C. (1993) Proc. Natl. Acad. Sci. USA 90, 7583–7587.
[10] Neuman, T., Keen, A., Knapik, E., Shain, D., Ross, M., Nornes, H.O. and Zuber, M.X. (1993) Eur. J. Neurosci. 5, 311–318.
[11] Bain, G., Robanus Maandag, E.C., Izon, D.J., Amsen, D., Kruesbeek, A.M., Weintraub, B., Krop, I., Schlissel, M.S., Feeney, A.J., van Roon, M., van der Valk, M., Riele, H.P.J., Bernes, A. and Murre, C. (1994) Cell 79, 885–892.
[12] Zhuang, Y., Soriano, P. and Weintraub, H. (1994) Cell 79, 875–884.
[13] Shain, D.H., Neuman, T. and Zuber, M.X. (1995) Nucleic Acids Res. 23, 1696–1703.
[14] Chomczynski, P. and Sacchi, N. (1987) Anal. Biochem. 162, 156–159.
[15] Sambrook, J., Fritsch, E.F. and Maniatis, T. (1989) Molecular Cloning: A Laboratory Manual, 2nd Edn., Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.
[16] Ausubel, F.M., Brent, R., Kingston, R.E., Moore, D.D., Seidman, J.G., Smith, J.A. and Struhl, K. (1994) Current Protocols in Molecular Biology, Wiley, New York, NY.
[17] Graham, F.L. and van der Eb, A.J. (1973) Virology 52, 456–467.
[18] Bisbee, C.A., Baker, M.A., Wilson, A.C., Hadji-Azimi, I. and Fischberg, M. (1997) Science 195, 785–787.
[19] Higgins, D.G., Bleasby, A.J. and Fuchs, R. (1992) Comput. Appl. Biosci. 8, 189–191.
[20] Shain, D.H., Zuber, M.X. and Neuman, T. (1998) Nucleic Acids Res. 26, 1019–1025.
[21] Wu, H.M. and Crothers, D.M. (1984) Nature 308, 509–513.