5' flanking sequence and genomic structure of Egr-1, a murine mitogen inducible zinc finger encoding gene

Chon-Hwa Tsai-Morris, Xinmin Cao and Vikas P. Sukhatme

Department of Medicine, Howard Hughes Medical Institute, University of Chicago, Chicago, IL 60637, USA

Received June 9, 1988; Revised and Accepted August 23, 1988 Accession no. X12617

ABSTRACT
Egr-1 is a murine zinc finger encoding cDNA whose expression is modulated by a variety of ligand-receptor interactions and is often coregulated with c-fos (1). This study reports the isolation of a mouse Egr-1 genomic clone, its intron-exon structure, and 935 bp of 5' flanking sequence. The gene spans about 3.8 kb and consists of 2 exons and one 700 bp intron. S1 nuclease protection and primer extension analysis were used to define the transcription initiation site. "TATA" and "CCAAT" sequences were located at nucleotides -26 and -337 respectively. In addition, there exist five elements whose sequence is nearly identical to the inner core 10 nucleotide region (CCATATTAGG) of the c-fos serum response element, four Spl consensus sequences, two API target sequence analogs, and two potential cAMP response elements. These results will ultimately lead to a detailed definition of the intracellular events regulating Egr-1 expression.

INTRODUCTION
Egr-1 is a recently identified (2) early growth response gene that is rapidly induced in diverse cell types (fibroblasts, epithelial cells and lymphocytes) following mitogenic stimulation. Egr-1 mRNA also increases dramatically during cardiac and neural cell differentiation, and following membrane depolarization both in vitro and vivo (1). This expression pattern along with the fact that the Egr-1 cDNA sequence predicts a protein with three zinc fingers suggests that Egr-1 may function in a more general role as a nuclear intermediary in signal transduction, as has also been postulated for c-fos (3-5), c-jun (6) and related genes (7).

Even though Egr-1 and c-fos are structurally unrelated, they are often coregulated in a strikingly similar manner. These data suggest that both Egr-1 and c-fos are likely to share cis-acting 5' regulatory elements. To further explore this relationship and as a first step towards our long term goal of defining in detail
the intracellular events that modulate Egr-1 expression, we have sequenced about 1 kb of 5' flanking region of the mouse Egr-1 gene. In addition, we have defined its intron-exon structure in order to ultimately compare its genomic organization to that of other Egr (early growth factor responsive) genes (manuscript in preparation), whose finger domains have been found to be nearly identical to those of Egr-1 but are otherwise structurally different.

MATERIALS AND METHODS
Genomic Library Construction and Isolation/Analysis of Mouse Egr-1 Genomic Clone

Balb/c 3T3 liver DNA was used for construction of a λFix genomic library by a partial fill-in cloning method (8). 10^5 unamplified clones in E. coli. strain JC7623 (rec B, rec C, sbc B) (9) were screened with a ^32P-labeled mouse Egr-1 plasmid OC 3.1 (1) containing a full length 3.1 kb cDNA insert. Membranes (GeneScreenPlus, NEN) were hybridized for 16 hours at 65°C in 1% SDS, 10% dextran sulfate and 1 M NaCl. The filters were washed to a final stringency of 65°C in 0.2 x SSC. Autoradiographs were prepared by exposing the filters for 18 hours at -70°C with an intensifying screen. A single clone (mgEgr-1.1) was obtained which also hybridized to the extreme 5' 120 bp EcoRI - Apal fragment from plasmid OC 3.1.

Southern Blot Analysis

1 µg of DNA from clone mgEgr-1.1 and 10 µg of genomic Balb/c 3T3 liver DNA were digested with various restriction enzymes, electrophoresed and blotted onto GeneScreenPlus (NEN) paper. These blots were hybridized at 65°C to a ^32P-labeled 3.1 kb Egr-1 insert (derived from OC 3.1) using 50 x 10^6 cpm for the genomic blot and 5 x 10^6 cpm for the clone blot. The pre-hybridization, hybridization and washing conditions used were as described above for the genomic library screening.

DNA Sequencing

A 2.4 kb PvuII - PvuII fragment and a 6.6 kb XbaI - XbaI fragment (see Figure 2) derived from mgEgr-1.1 were subcloned into the SmaI and XbaI sites of pUC13 and pUC18 respectively, and the resulting plasmids were denoted as p2.4 and p6.6. These plasmids were sequenced using the chain-termination method (10)
by a combination of sequential specific oligonucleotide priming and by cloning of various restriction fragments in pUC13.

**Primer Extension Analysis**

A $^{32}$P end-labeled oligonucleotide complementary to nucleotides 7-35 (CTGGCGGCGGCGGCGAATCGCGGCGGCGG) of the Egr-1 cDNA sequence (1) was annealed to 10 μg of total cellular RNA prepared from serum and cycloheximide stimulated mouse fibroblasts and from kidney tissue (which shows minimal expression of Egr-1 mRNA (1)) for 3 minutes at 75-85°C, then placed at 60-65°C for 60 minutes and then allowed to cool slowly to 42°C. Hybridization was performed in 73 μl for twenty minutes at 42°C in 15.4 mM DTT, 40.6 μg/ml Actinomycin D, 0.65 U/μl RNasin, 0.54 U/μl AMV-RT, 0.1 mg/ml BSA and 95.9 μM dNTP. The reaction was then "chased" with 940 μM dNTPs, 2.8 mM DTT and 0.68 U/μl AMV-RT for 15 minutes in a final volume of 90 μl. Hybrids were subsequently precipitated and reaction products analyzed on a 6% polyacrylamide gel containing 8 M urea.

**SI Nuclease Assay**

From plasmid p2.4, a 500 bp Smal - Smal fragment (see Figure 4) was isolated, phosphatased and labeled with 100 μCi $^{32}$P-y-ATP (11). It was then annealed to 10 μg of total RNA (described above in the primer extension section) at 85-95°C for 10 minutes, and hybridized for 18 hours in a volume of 30 μl at 60-62°C in the presence of formamide, 0.4 M NaCl, 0.05 M PIPES and 0.83 mM EDTA. Hybrids were treated with 1 U/μl SI nuclease for 1.5 hours at 37°C in a final volume of 300 μl. The reaction products were then extracted with phenol/chloroform and precipitated prior to analysis by electrophoresis as described in the primer extension section.

**RESULTS**

**Isolation of a Mouse Genomic Clone Containing the Egr-1 Gene**

Initially, we unsuccessfully screened a total of approximately 1.5 x 10⁶ genomic clones from three different libraries: NIH3T3 $^\lambda$FIX library (from Stratagene, La Jolla), mouse cosmid library (gift from Dr. Yun-Fai Chris Lau, University of California School of Medicine, San Francisco) and mouse genomic EMBL3 library (from Clontech, Palo Alto). Therefore, a Balb/c 3T3 genomic library was constructed in $^\lambda$FIX and 100,000
Figure 1. Southern Analysis of Genomic liver and Phage mgEgr-1.1 DNA. For the left panel, 1μg of phage mgEgr-1.1 DNA was digested with Rsal (R), BglII (B), and PvuII (P), electrophoresed through a 1% agarose gel, blotted, and hybridized with ^32P-labeled 5 x 10^6 cpm of a full-length Egr-1 insert OC3.1 (1). Exposure was for 20 minutes with an intensifying screen at -70°C. For the right panel, 10 μg of liver Balb/c 3T3 DNA was digested with the same enzymes as above, and processed similarly except that 50 x 10^6 cpm of probe was used and exposure was for 18 hours.

unamplified clones in the debilitated host JC 7623 were screened resulting in a single positive clone (designated mgEgr-1.1) that spanned the 5' end of the cDNA and included about 1 kb of flanking sequence (see below under Restriction Mapping). Southern analysis of genomic liver DNA and phage DNA from mgEgr-1.1 using a full length 3.1 kb cDNA Egr-1 probe (OC 3.1) was used to verify the authenticity of the phage isolate. Figure 1 shows that identical fragments hybridize in both genomic and phage DNA when it is digested with Rsal. Note that in distinction to the Rsal results, the upper band in the BglII and PvuII digests differ from each other because mgEgr-1.1 contains only 1 kb of 5' flanking sequence which has no sites for either BglII or PvuII whereas it does for Rsal (see Figure 2).

Restriction Map of Mouse Egr-1 Gene

Figure 2 shows a partial restriction map of phage mgEgr-1.1 obtained by analysis of fragments by agarose gel electrophoresis and subsequent Southern blot analysis using Egr-1 cDNA fragments.
In particular, a 6.6 kb XbaI - XbaI fragment from mgEgr-1.1, which hybridized to both 5' and 3' fragments of the Egr-1 cDNA, was subcloned into the XbaI site of pUC18. Similarly, a 2.4 kb PvuII - PvuII fragment from mgEgr-1.1 which hybridized to the 5' end of the Egr-1 cDNA, was subcloned into the SmaI site of pUC13. Both plasmids were used for mapping and sequence analysis.

Gene Organization

To obtain genomic sequence, specific oligonucleotides (17-mers at positions 83, 122, 174, 200, 379, 543, 611, 659, 905, 920, 1000, 1200, 1400, 1600, 1800, 2100, 2353, 2650, 2825 of the cDNA sequence (1)) were used as primers for double stranded sequencing of plasmids p2.4 and x6.6. Comparison of the Egr-1 genomic sequence with that of the cDNA sequence showed the Egr-1 gene consists of 2 exons and a single 700 bp intron (between nucleotide position 556 and 557 as numbered in the cDNA sequence (1)) as shown in Figure 2. Both the 5' and 3' splice junction
### TABLE 1

**Mouse Egr-1 Exon-Intron Junction Sequences**

| Exon 1 - ACAG[GTAAGCGGTGG - - - TCCCTTTCCTGCCAG]AGTCC - Exon 2 |
| Exon-Intron Splice Junction Sequences in the Mouse Egr-1 Gene. The exon-intron boundaries were determined by comparison of the cDNA and genomic sequence. The intron length is about 700 bp. The numbers on top of the sequence refer to the nucleotide position in mouse Egr-1 cDNA (1). The consensus boundary sequences (12) are also shown for comparison.

sequences are in excellent agreement with the consensus boundary sequences (12) as shown in Table 1.

**5' Flanking Sequence of Mouse Egr-1 Gene**

From plasmid p2.4, we obtained approximately 1 kb of 5' flanking sequence as shown in Figure 3. In order to map the transcription initiation site, we used both primer extension and S1 nuclease analysis. Results of primer extension experiments (Figure 4, left panel), using a 29-mer whose sequence was derived from a region near the 5' end of the cDNA, indicate that the 5' end of our published cDNA sequence starts 24 nucleotides inside the mRNA. We have chosen as +1 (Figure 3) the position predicted by the most intense band from the oligomer extension data (arrow in Figure 4, left panel). In addition, the 5' end of the mRNA was also determined by S1 nuclease analysis (Figure 4, right panel). In this experiment, a 482 bp SmaI - SmaI fragment derived from plasmid p2.4 was end-labeled and hybridized to RNA. The size of the S1 nuclease resistant DNA indicated a location for transcription initiation site in agreement with data obtained from the primer extension studies. The two brightest bands seen (arrows in Figure 4, right panel) in the S1 nuclease analysis data correspond to two nucleotides 5' and 3' of the cap site as determined by oligomer extension.

A number of possible regulatory elements were identified in the 5' flanking sequence of mgEgr-1.1 (Table 2). A putative TATA motif (AAATA) is located 26 nucleotides upstream of the the transcription start site. A "CCAAT" type sequence starts at nucleotide -337. Five different regions, each 10 nucleotides in
length, located at -110, -342, -358, -374, and -412, are nearly identical to the inner core of the c-fos serum response element (13). Each has a 5-6 nucleotide AT rich stretch and is surrounded by the dinucleotide CC on 5' side and GG on the other. Two potential TPA responsive elements (14,15) are located at nucleotides -610 and -867. Four consensus Spl (16) binding
Figure 4. Primer Extension and S1 Nuclease Analyses. Primer extension analysis on mouse Egr-1 mRNA is shown in the left panel. Lane 1: extension products obtained using control (kidney) RNA; lane 2: extension products obtained using RNA
prepared from Balb/c 3T3 fibroblasts stimulated by 20% fetal calf serum in the presence of 10 μg/ml of cycloheximide for 3 hours. P marks the position of the unextended 29-mer. S1 nuclease analysis is shown in the right panel. The four lanes marked A, C, G, T are sequencing reaction products used as markers. Lanes 1 and 2: 350 and 482 bp markers respectively; lanes 3 and 5: blank; lane 4 is control (kidney) (see above); lane 6: arrows point to the protected fragment when RNA from serum and cycloheximide stimulated cells is used.

sequences are at position -285, -649, -700 and -719. In addition, we have identified two sequences that might serve as a cAMP response elements (17) (-138 and -631).

DISCUSSION
We have presented the genomic organization and about 1 kb of 5' flanking region sequence of murine Egr-1, a recently identified early growth factor responsive gene. The importance of this gene is that (a) its expression is modulated during diverse biological processes including cell growth, differentiation and following cellular depolarization and (b) it itself encodes a putative transcription factor by virtue of its zinc finger structure. These data suggest that Egr-1 may function in the broad role of a nuclear intermediary in signal transduction.

Egr-1 Flanking Sequence
In view of these results, it was of interest to begin to define the intracellular events that regulate Egr-1 expression and to understand the mechanisms by which Egr-1 and c-fos are often coregulated. We have identified several potential regulatory elements in the 5' upstream sequence of Egr-1 with similarities to those found in c-fos. In particular, deletion analysis and DNA footprinting have defined a serum response element (SRE) in c-fos consisting of 22 nucleotides which displays dyad symmetry (13). A similar sequence has also been found in the Xenopus laevis γ-actin gene promoter (18). Interestingly, the Egr-1 upstream region contains 5 domains of 9-10 nucleotides which are nearly identical to the inner core of these SREs. It is tempting to speculate that the presence of a large number of these elements might account for the 5 to 10 fold greater serum induction of Egr-1 compared to that noted for c-fos (1,2). Since the transcriptional response to TPA is mediated (at
### TABLE 2
Location and Identification of Potential Regulatory Elements

<table>
<thead>
<tr>
<th>Element</th>
<th>Sequence*</th>
<th>Location+</th>
</tr>
</thead>
<tbody>
<tr>
<td>TATA</td>
<td>AAATA</td>
<td>-26 to -22</td>
</tr>
<tr>
<td>CCAAT Box</td>
<td>CCAAT</td>
<td>-337 to -333</td>
</tr>
<tr>
<td>Serum Response Element</td>
<td>GATGTCCCATATTAGGACATC</td>
<td>(Ref. 13)</td>
</tr>
<tr>
<td>Consensus</td>
<td>CC TA AT GG</td>
<td></td>
</tr>
<tr>
<td></td>
<td>G</td>
<td></td>
</tr>
<tr>
<td></td>
<td>TCCTTCCATATTAGGCTTC</td>
<td>-110 to -91</td>
</tr>
<tr>
<td></td>
<td>GTGGCCC-AATATGGCCCTG</td>
<td>-342 to -324</td>
</tr>
<tr>
<td></td>
<td>CAGCGCTTTATATGGAGTGG</td>
<td>-358 to -339</td>
</tr>
<tr>
<td></td>
<td>ACAGACCTTTATTTGGGAGC</td>
<td>-374 to -355</td>
</tr>
<tr>
<td></td>
<td>AAACGCATATAAGGAGCAG</td>
<td>-412 to -393</td>
</tr>
<tr>
<td>TPA Responsive Element (API binding site)</td>
<td>C C</td>
<td>(Ref. 14,15)</td>
</tr>
<tr>
<td>Consensus</td>
<td>TGACTA</td>
<td></td>
</tr>
<tr>
<td></td>
<td>GA</td>
<td></td>
</tr>
<tr>
<td></td>
<td>CTGACTCG</td>
<td>-610 to -603</td>
</tr>
<tr>
<td></td>
<td>CTGACTTG</td>
<td>-867 to -860</td>
</tr>
<tr>
<td>Spl binding site</td>
<td>GGGCGG</td>
<td>-285 to -280</td>
</tr>
<tr>
<td></td>
<td>GGGCGG</td>
<td>-649 to -644</td>
</tr>
<tr>
<td></td>
<td>CCGCCC</td>
<td>-700 to -695</td>
</tr>
<tr>
<td></td>
<td>GGGCGG</td>
<td>-719 to -714</td>
</tr>
<tr>
<td>cAMP Response Element</td>
<td>TGACGTCA</td>
<td>(Ref. 17)</td>
</tr>
<tr>
<td>Consensus</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>TCACGTCA</td>
<td>-138 to -131</td>
</tr>
<tr>
<td></td>
<td>TGACCGTG</td>
<td>-631 to -624</td>
</tr>
</tbody>
</table>

**Location and Identification of Potential Regulatory Elements.** * The underlined bases in the mouse Egr-1 gene sequence are mismatched from the consensus sequence. + The numbers refer to the nucleotides of the mouse Egr-1 gene as indicated in Figure 3.

least partially) through binding of API (14,15), we searched for API binding sites in the 5' flanking region and two such potential elements were found. This finding may be of significance since TPA is a powerful inducer of Egr-1 expression in murine fibroblasts (2) and in embryonal carcinoma cells (data not published). Similarly, the putative cAMP response elements (-138 and -631) may mediate the Egr-1 induction seen in cardiac cells by β-adrenergic agonists (data not shown).
Genomic Organization

TFIIIA is the archetypical Cys$_2$-His$_2$ type of zinc finger motif (19). Mammalian structures with similar domains have been found recently in the transcription factor Spl (20), in GLI, a gene amplified in glioblastomas (21), in the testis-determining factor (TDF) (22), in a growth regulated gene (Krox-20) (23) as well as in Egr-1/NGFI-A (1,24). To the best of our knowledge, our study is the first to report on the genomic organization of a mammalian gene which encodes zinc fingers of the Cys$_2$-His$_2$ class. The Egr-1 gene has a surprisingly simple organization: it is composed of 2 exons and one intron. The first exon includes the first 99 amino acids of the deduced protein, while the second includes the rest of protein in which are located the three tandem zinc finger motifs. The TFIIIA gene structure (25) shows that the first six fingers are each separated by introns located at the base of each finger whereas the last three are not. The Egr-1 structure indicates that the fingers are all on one exon with no splice junction located just outside the finger region. Given that these fingers share 50-70% amino acid identity to each other, it is possible that they have evolved through gene duplication and that the introns joining them have subsequently been lost. We have recently identified two other genes which cross hybridize to the finger region of Egr-1 and which are also regulated by mitogenic stimulation (manuscript in preparation). Remarkably, the finger domain sequences are nearly identical to those of Egr-1 whereas the non-finger flanking regions diverge significantly. Thus, it will be of particular interest to define the genomic structure (and regulatory region sequences) of these genes in order to understand their evolutionary relationship.

ACKNOWLEDGEMENTS

We thank V. Rangnekar for help in the initial stages of this study. This work was supported by the Howard Hughes Medical Institute.

REFERENCES

2. Sukhatme, V.P., Kartha, S., Toback, G., Taub, R., Hoover,