Nucleic Acids Research, Vol. 20, No. 7 ***1477–1482***

Isolation of high affinity cellular targets of the embryonal LTR binding protein, an undifferentiated embryonal carcinoma cell-specific repressor of Moloney leukemia virus

Toshio Tsukiyama and Ohtsura Niwa
Department of Pathology, Research Institute for Nuclear Medicine and Biology, Hiroshima University, 1-2-3 Kasumi, Minami-ku, Hiroshima 734, Japan

Received February 17, 1992; Revised and Accepted March 16, 1992

ABSTRACT

ELP, the embryonal LTR binding protein, is a member of the nuclear receptor superfamily and a mouse homologue of *Drosophila* FTZ-F1. ELP is expressed specifically in undifferentiated mouse embryonal carcinoma cells and participates in suppression of the Moloney murine leukemia virus genome. The zinc finger domain of the protein was fused with glutathione S-transferase and was successfully used for isolating genomic targets. Sixteen genomic fragments were isolated and twelve of them strongly interacted with ELP. Six of the ELP binding fragments were analyzed further. All of these contained the multiple binding sites for ELP, which matched well with the consensus binding sequence for FTZ-F1, YCAAGGYCR. Among these, three fragments functioned as negative regulatory elements in response to ELP, when placed upstream to the promoter region of the Moloney leukemia virus. These results indicate that ELP may function as a negative transcription factor for a variety of cellular sequences, in addition to suppressing expression of Moloney leukemia virus in early embryonal cells. It was also shown that the procedure employed here works well for isolation of genomic targets of transcription factors.

INTRODUCTION

Transcription factors play a major role in determining the phenotypes of cells during development of higher animals. These factors bind to specific sequence elements located in the control region of genes. Because these elements consist of a short stretch of nucleotides, one particular DNA binding protein usually bind to regulatory elements of multiple genes and controls their activities (1). Therefore, in order to elucidate the complex network of gene regulation during development in which multiple genes are regulated in concert, it is important to identify a set of target genes for a particular transcription factors.

Several strategies have been employed for the isolation of cellular target sequences for various transcription factors. All the methods exploited the ability of transcription factors to bind particular sequence elements. DNA-protein complexes were recovered by either immunoprecipitation (2–4) or binding to nitrocellulose filters (5). Similarly, synthetic oligonucleotides bound to transcription factors were recovered by either nitrocellulose filter binding (6,7), Southwestern blotting (8) or by affinity column (9).

We have previously identified the embryonal long terminal repeat (LTR) binding protein, ELP, which negatively regulates expression of Moloney murine leukemia virus (Mo-MuLV) in undifferentiated mouse embryonal carcinoma cells (10,11). Analysis of the cDNA indicated that ELP belongs to the nuclear receptor superfamily (12). In addition, it was found that ELP is a mouse homologue of *Drosophila* FTZ-F1, which is also a member of the nuclear receptor superfamily and positively regulate transcription of the *fushi tarazu* gene (13,14). ELP and FTZ-F1 are quite unique in that they are highly conserved during evolution and that they are expressed in early embryonal cells. These data indicate the importance of the function of the two transcription factors in early embryogenesis. Therefore, identification of cellular targets of ELP is of prime importance. In this study, we describe an efficient method of cloning such sequences and demonstrate that some of these indeed function as negative elements when fused with the promoter regions of the LTR.

MATERIALS AND METHODS

Production of the DNA binding domain of ELP in bacteria

A DNA fragment of the ELP cDNA (12) from AvaII to AvaII (nucleotides 337 to 802), encoding amino acids 47 to 200, was cloned into bacterial expression vector pGEX2T (15). This DNA fragment contains the coding region for the zinc finger domain of ELP.

The resulting plasmid, pGST-EZ, was introduced into the HB101 strain of *Escherichia coli (E. coli)*, and production of protein was induced as described (15). The induced bacteria were collected by centrifugation, and washed once with phosphate buffered saline. The bacterial pellet was suspended in 1/20 volume of 10 mM Hepes (PH 7.9), 0.3 M NaCl, 0.5 M sucrose, 5 mM $MgCl_2$, 0.5% NP40, 1 mM DTT, and subjected to four

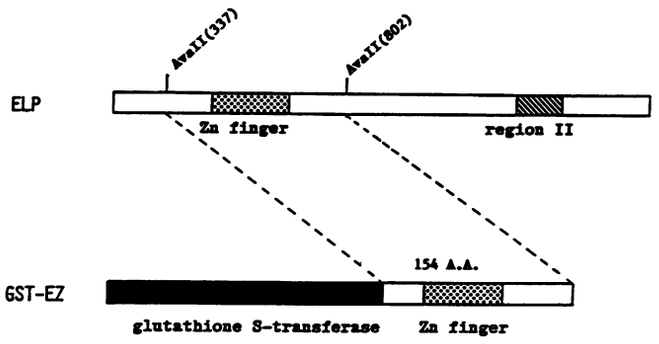

Fig. 1. Construction of the fusion gene. The zinc finger domain of ELP was fused with glutathione S-transferase. The dotted region indicates the zinc finger domain.

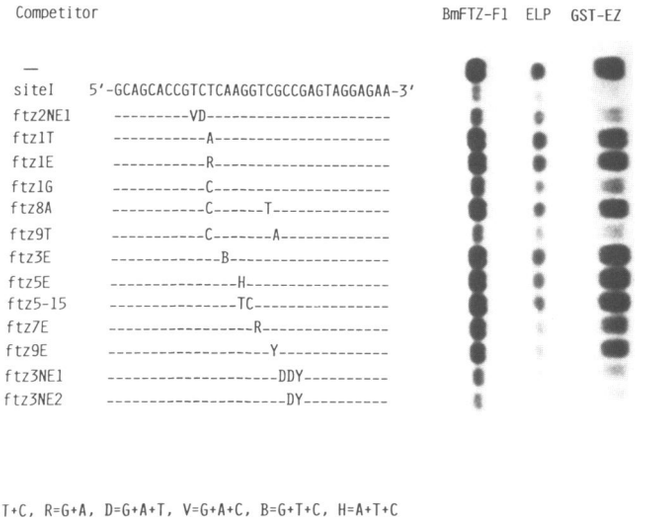

Fig. 2. Comparison of the specificity of sequence recognition by BmFTZ-F1, ELP, and GST-EZ. Only the bands corresponding to the complexes are shown. The site I of FTZ-F1 (13) was used as a probe. A 20-fold excess of mutated site I sequences were used as competitors.

rounds of freezing and thawing. The lysate was then sonicated five times for one minute each on ice, and centrifuged at 30,000 rpm for 30 minutes at 4°C in Beckmann SW 50 rotor. The supernatant was dialyzed against 10 mM Hepes (PH 7.9), 1 mM $MgCl_2$, 1 mM DTT, 50 mM NaCl, 0.5 mM EDTA, 50% glycerol for 5 hours at 4°C, and stored at −80°C.

GST-EZ protein was bound to glutathione agarose beads (Sigma) as described (15).

Gel retardation assay

Preparation of nuclear extracts (10) and the procedure for gel retardation assays (13) were as previously described. For the assays with the bacterial extract, 20 ng of the protein was used for each reaction.

DNaseI foot printing

The probes for DNaseI foot printing were terminally labeled to a specific activity of about 8×10^3 cpm/fmol as described (10). A 5 fmol of the probe was incubated at room temperature for 30 minutes with either 0, 100, 200, 500, 1000,or 2000 ng of the bacterial extract prepared as described above in 10 mM Tris HCl (PH 7.5), 30 mM NaCl, 1 mM DTT, 5% (vol/vol) glycerol, 1 μg of poly(dI-dC) poly(dI-dC) (Pharmacia), in a total volume of 40 μl. After the binding reaction, 4 μl of DNaseI (Sigma), diluted to 25 μg/ml with 10 mM $MgCl_2$, 5mM $CaCl_2$, was added to the mixtures and incubated for 1 minute at 20°C. The digestion was terminated by adding 60 μl of 20 mM Tris HCl (PH8.0), 1% SDS, 20mM EDTA, 0.02 μg/ul calf thymus DNA. The reaction mixtures were phenol extracted, and the probes were ethanol precipitated and then loaded onto sequence gels. The probe DNA was chemically cleaved as previously described (16) and run in parallel as markers.

Chloramphenicol acetyl transferase (CAT) assay

The reporter plasmid, pMolXKCAT, carries the promoter region [XbaI (−150) to KpnI (+31)] of Mo-MuLV upstream of the CAT gene. This region was shown to be sufficient for the basal transcriptional activity (11). In pSP8XKCAT, eight copies of ELP binding sites were inserted upstream of the promoter region of pMolXKCAT in the same orientation as in the original ELP site. The DNA fragments, isolated from mouse genomic DNA, were placed upstream of the promoter region of pMolXKCAT in both orientations.

The ELP expression vector, pRV-ELP, was described previously (12) and it contains the entire open reading frame for ELP under the control of the Rous sarcoma virus LTR. As for the negative control, a stop codon linker (New England Biolabs Inc.) was inserted between the PmaCI sites in the zinc finger region of ELP in pRV-ELP, and the resulting plasmid was designated pRV-ELP-ZFS. The effector plasmids, the reporter plasmids, and p-act-β-gal plasmid which served as an internal control, were transfected to NIH3T3 cells, and CAT assay was performed as previously described (12). The quantities of cell extracts used for CAT assays were normalized by the β-galactosidase activity. The assay was repeated three times, and radioactivities of two representative samples were quantitified by liquid scintillation counting. The average of the counts were presented.

RESULTS

The binding specificity of the ELP-GST fusion protein

A recombinant plasmid was constructed which expresses the zinc finger domain of ELP fused to the C-terminus of GST (Fig. 1). The fusion protein, designated as GST-EZ, was purified from the bacterial lysate and was tested for the specificity of the binding to the FTZ-F1 site I sequence by the gel retardation competition assay (12). The binding was competed to various degrees by mutant sequences and the patterns of the competition were identical among BmFTZ-F1, a *Bombix mori* homolog of FTZ-F1 (17), ELP and GST-EZ (Fig. 2). This verifies that the fusion protein can be used in place of ELP for the screening of genomic sequences with ELP binding sites.

Screening of mouse genomic library

Our strategy of cloning genomic sequence is essentially same as the one reported by others (18) and is shown in Fig. 3. A 20 μg of genomic DNA was digested with Alu I and the purified DNA was dissolved in in 200 μl of binding buffer [50 mM NaCl, 50 mM Tris HCl (PH8.0), 1 mM EDTA, 1 mM DTT, 20%

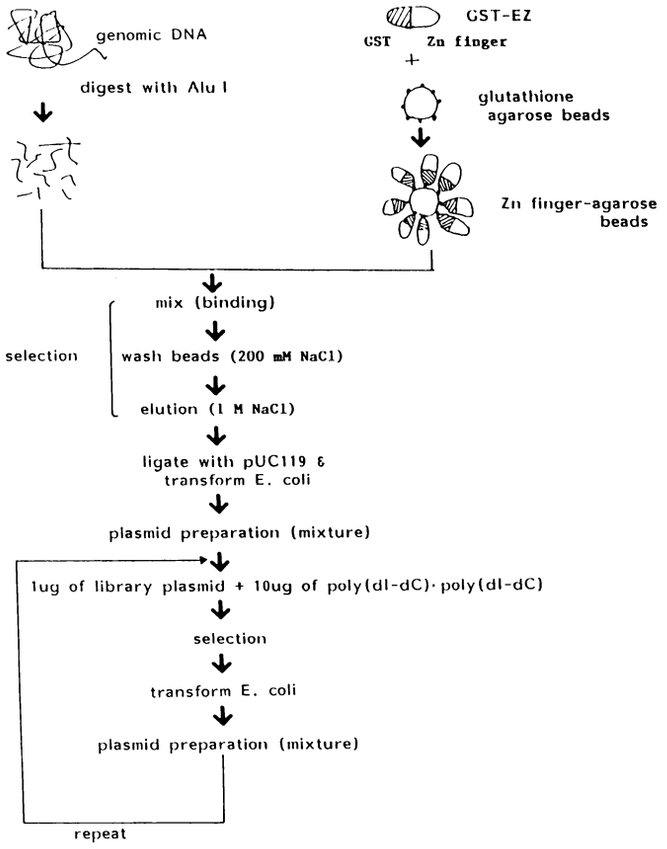

Fig. 3. The strategy employed to isolate the genomic DNA containing ELP binding sites. See text for detailed procedure.

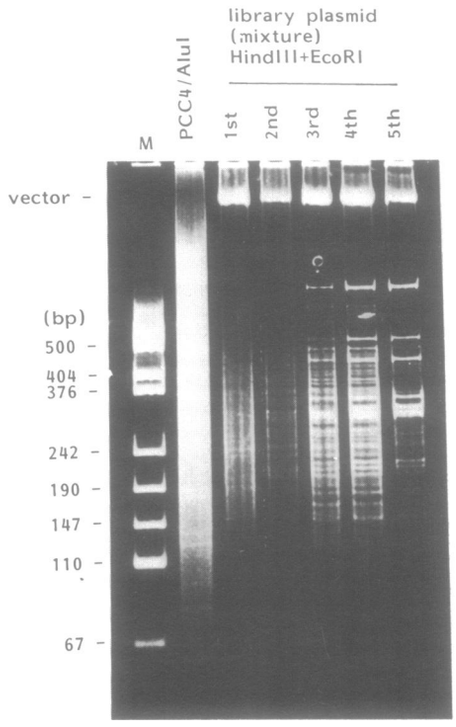

Fig. 4. Ethidium bromide staining pattern of the inserts of the selected ELP target libraries. PCC4/AluI indicates the starting material, DNA of mouse PCC4 EC cell line digested with AluI. Total library DNA was digested at HindIII and EcoRI sites in the polylinker region of pUC119.

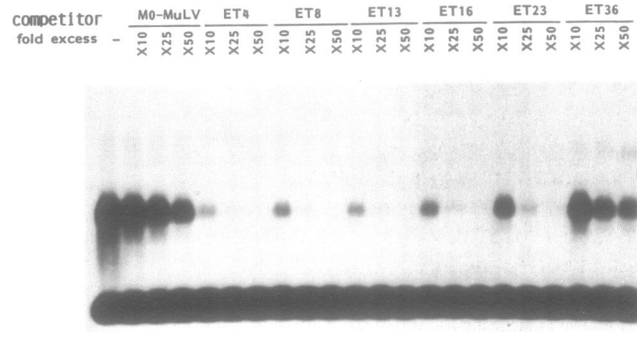

Fig. 5. Quantitative competition of ELP binding by the genomic ELP binding fragments. The site I of FTZ-F1 (13) was used as a probe. Mo-MuLV indicates the Sau3AI (−353) to XbaI (−150) fragment of the Mo-MuLV LTR, which contains single ELP site and was originally used for detection of ELP complex (10). The competitor fragments were prepared as described (10).

(vol/vol) glycerol, 0.1% (vol/vol) NP40, 2.5 μg/μl BSA]. Agarose beads of 100 ul bed volume bound with GST-EZ (about 0.5 μg of GST-EZ protein/μl of resin) was added to the mixture, and incubated on ice for 30 minutes. After washing of the resin with low salt buffer (binding buffer with 200 mM NaCl), bound DNA was recovered by elution with high salt buffer (binding buffer with 1 M NaCl) and cloned into the SmaI site of pUC119 vector.

After amplification of the plasmid library in *E. coli*, the selection was repeated several times. At each cycle of the selection, DNA was examined by polyacrylamide gel electrophoresis. The smear pattern of the ethidium bromide staining converged to that of less than 20 bands after 5 cycles of the selection (Fig. 4). This clearly demonstrates that particular clones were selected by this procedure. No change in the pattern was observed after 6th selection (data not shown), and the selection was terminated.

Binding of the cloned sequences to the FTZ-F1 site I

Sixteen independent clones were isolated after 6 cycles of the selection. These were first tested for the affinity to ELP, and twelve of them were shown to be positive (data not shown). Among them, six independent clones, ET4, ET8, ET13, ET16, ET23, ET36, were analyzed further. They were sequenced and the length of the inserts were shown to be 346 base pairs (bp), 247 bp, 273 bp, 246 bp, 244 bp, and 193 bp, respectively. Their affinity to ELP was assayed by quantitative competition in gel retardation assay (Fig. 5). The FTZ-F1 site I sequence was used as a probe and the excess amount of the unlabeled inserts were added to the reaction mixtures. All of these showed much stronger binding to the FTZ-F1 site I than the ELP site of Mo-MuLV, although affinity were different among the clones. Thus, the procedure used in this study indeed selected genomic sequences with high affinity to ELP. These inserts were analyzed further by DNase I foot printing.

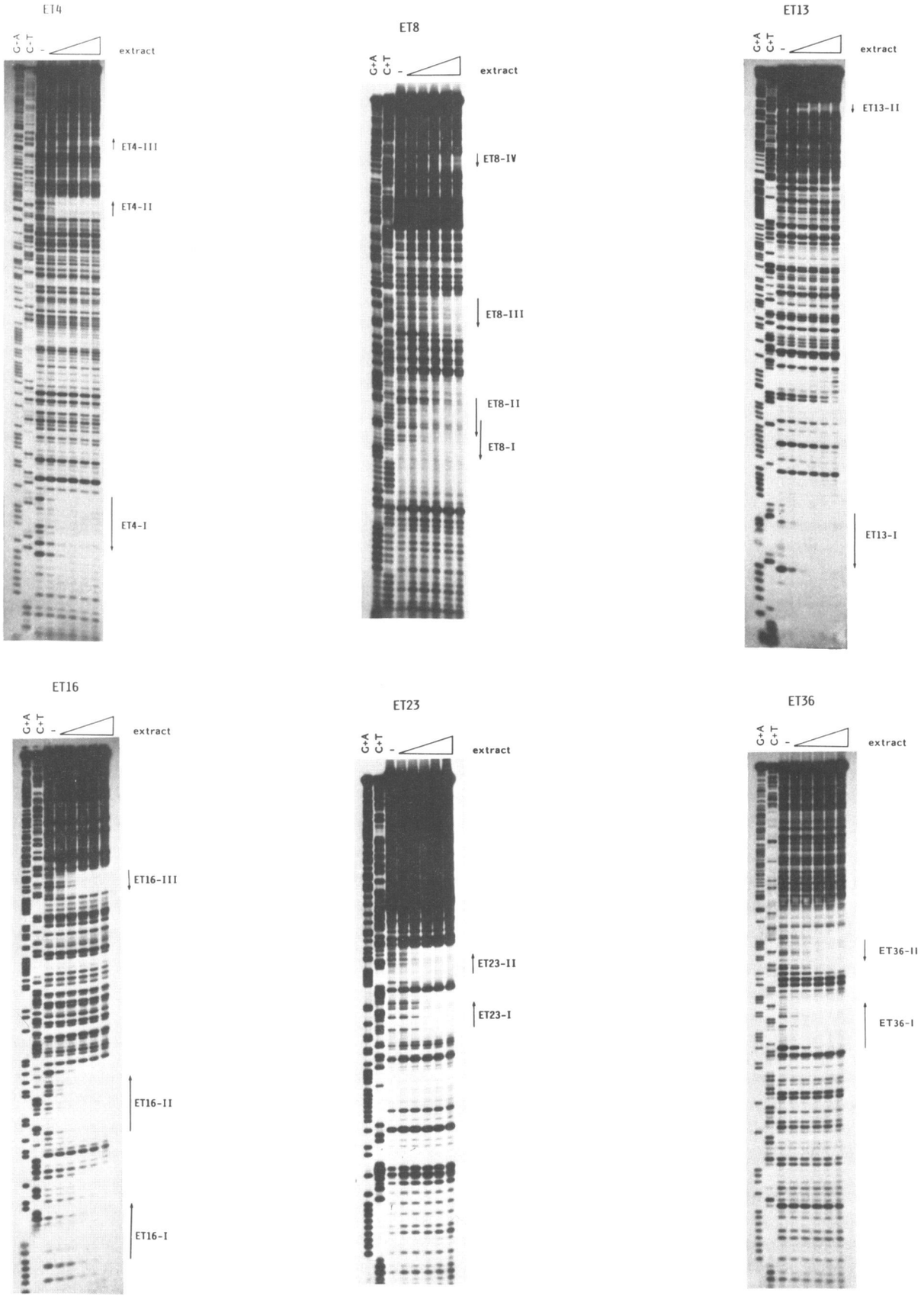

Fig. 6. DNaseI foot printing of genomic ELP targets by GST-EZ. The position and the orientation of FTZ-F1 consensus sequences in protected areas are indicated by arrows. They are temporary numbered for the convenience. The triangle above indicates the increase in amount of bacterial extract containing GST-EZ protein.

Fig. 7. Nucleotide sequences of genomic ELP binding sites. The sequences matching to the FTZ-F1 consensus are boxed. The open areas correspond to the polylinker region of the vector.

Foot printing analysis

Fig. 6 shows the foot print patterns of the clones. All the sequences matching to FTZ-F1 consensus were protected by GST-EZ. The number of the ELP binding sites located in the clones were, 4 in ET8, 3 in ET4 and ET16, and 2 in ET13, ET23 and ET36. It is interesting to note that the strength of the binding to ELP does not necessarily parallel to the number of the elements in the clones.

Sequences of the binding sites thus identified are shown in Fig. 7. Most of the sequences match with the consensus sequence motif for FTZ-F1 binding, YCAAGGYCR (19). Out of 16 binding sites, only 4 violated the general rule. Adenine was found in place of the pyrimidine at the first position of the motif in ET23-I site. Thymidine was found in place of cytosine at the 8th position of the motif in ET4-III, ET16-I and ET36-II sites. The sequence CAAGG was completely kept in all the sites, indicating the importance of the motif.

Functional analysis of the clones

Since all the clones possessed the binding sites for ELP, we further tested whether or not they function as regulatory elements responding to ELP. The reporter plasmids were co-transfected

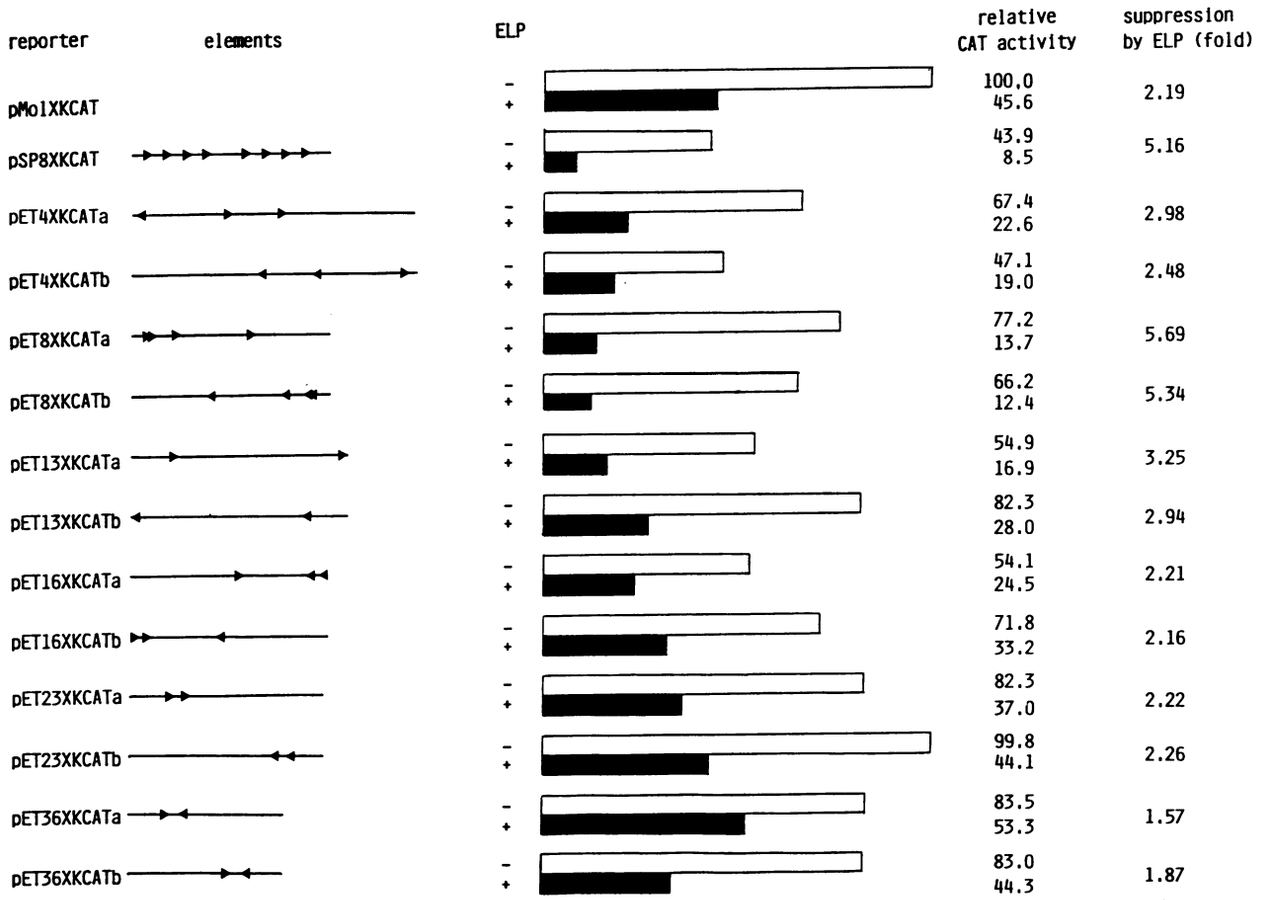

Fig. 8. Functional assay of genomic ELP binding fragments. ELP + and − denotes co-transfection with ELP expression vector pRV-ELP, and negative control vector, pRV-ELP-ZFS (12), respectively. The line on the column of elements represents the relative length of the ELP binding fragments, and the triangle on the line indicates position and orientation of ELP binding sites. Each CAT activity was normalized to that of pMolXKCAT without ELP, and shown as relative CAT activity. Fold suppression by ELP is the ratio of CAT activity of each reporter plasmids with and without ELP expression.

into NIH3T3 cells together with the effector plasmid expressing functional ELP driven by the RSV promoter and the plasmid carrying the β-galactosidase gene driven by the β-actin promoter. NIH3T3 cells lacked endogenous ELP activity (9) and suited for assessing the effect of ELP on the reporters. Experiments were repeated 3 times and the efficiency of transfection was always normalized by the level of β-galactosidase activity.

As shown in Fig. 8, the promoter region of the Mo-MuLV LTR (11) was used as a basal element for expression of the CAT gene. This element itself was negatively regulated by ELP (2.19 fold suppression), probably because of the presence of the cryptic ELP site in the region. Eight copies of the ELP site functioned as a negative element when placed upstream of the promoter region, and 5.16 fold reduction in the CAT activity was noted. ET8, which had 4 ELP binding sites, functioned in an orientation independent manner and suppressed the expression of the CAT gene as strongly as the 8 tandem copies of the ELP sites. The effect of ET13 which had 2 ELP sites was always stronger than that of ET4 which had 3 ELP sites. ET16, ET23 and ET36 had no effect on expression of the reporter gene in the presence of ELP. These data indicates that the degree of the response to ELP does not always parallel to the number of ELP sites nor the affinity to ELP of each clone in gel retardation assay.

DISCUSSION

ELP was initially identified as a repressor of transcription of the LTR of Mo-MuLV in EC cells (10). It was later found to be a mouse homologue of *Drosophila* FTZ-F1 (12). Strong conservation of the gene structure between the two during evolution of vertebrates and arthropods suggest functional importance of these genes. In order to elucidate the function of ELP in mouse embryogenesis, identification of cellular targets is necessary.

Nuclear receptors share sequence motifs which code for the zinc finger domain of the proteins. This domain was shown to be sufficient by itself for the sequence specific binding to target elements (20,21). Development of a new expression vector which directs synthesis of many polypeptides as soluble proteins fused with GST (15) enabled us to devise a new method of isolating genomic targets of ELP. The fusion protein, GST-EZ, bound on glutathione beads efficiently trapped genomic binding fragments. All the genomic fragments cloned after 6 cycles of selection and amplification of plasmids in *E. coli* had more than 2 copies of ELP binding motifs. They also bound to ELP much more strongly than did the authentic ELP site of Mo-MuLV. It is likely that 6 cycles of selection yielded only those with strong affinity to ELP. Numerous fragments were seen in the samples selected through 3 and 4 cycles and the final sample only consisted of a small number of bands. Therefore, we will be able to isolate genomic fragments with less number of the motif and less affinity to ELP from the pool of the plasmid selected with less number of the cycle.

Recently, Fainsod et al. (18) reported isolation of cellular targets for chicken homeobox gene using essentially the same method we employed here. Therefore, this method can be successfully applied for cloning targets of at least two different class of transcription factors.

The footprinting analysis identified the sequence of naturally occurring ELP binding sites in mouse genome. They matched well with the consensus sequence of FTZ-F1 (19), especially in the motif, CAAGG, which was conserved to 100%. The five nucleotide sequence was thus suggested to play a critical role in recognition of the element by ELP. Search of the EMBL nucleotide sequence data bank indicated that all the sequences isolated in our present work have not been registered in it (data not shown).

The CAT assay of the construct with the genomic fragments indicated that some of these, ET4, ET8 and ET13, did function as negative regulatory elements in the presence of ELP. This suggests that they may be part of regulatory regions of functional genes whose activity is regulated by ELP in EC cells. Analysis of flanking regions of these fragments is necessary to clarify this point.

FTZ-F1 activates expression of the *fushi tarazu* gene, while the ELP site of Mo-MuLV and the genomic sequences cloned in this communication indicate that ELP functions as a repressor of transcription. The cause of this discrepancy remains to be clarified. Identification of as many cellular target genes will reveal the mode of action of ELP and its role in the mammalian embryogenesis.

ACKNOWLEDGMENTS

We thank Drs. H. Ueda, S. Hirose and K. Umesono for valuable discussions. We also thank Mr. T. Nishioka for photographic works and Ms. Matsuura for typing the manuscript. This work is supported by a Grant-in-Aid from the Ministry of Education, Science and Culture, Japan. T.T. is a recipient of a post-doctoral fellowship from Japan Society for the Promotion of Science in Cancer Research.

REFERENCES

1. Wingender, E. (1988) Nucleic Acids Res., 16, 1879–1902.
2. Kinzler, K. and Vogelstein, B. (1989) Nucleic Acids Res., 17, 3645–3653.
3. Sompayrac, L. and Danna, K. J. (1990) Proc. Natl. Acad. Sci. USA, 87, 3274–3278.
4. Gould, A. P., Brookman, J. J., Strutt, D. I. and White, R. A. H. (1990) Nature, 348,308–312.
5. Inoue, S., Kondo, S., Hashimoto, M., Kondo, T. and Muramatu, M. (1991) Nucleic Acids Res., 19, 4091–4096.
6. Biedenkapp, H., Borgmeyer, U., Sippel, A. E. and Klempnauer, K-H. (1988) Nature, 335, 835–837.
7. Thiesen, H-J. and Bach, C. (1990) Nucleic Acids Res., 18, 3203–3209.
8. Keller, A. D. and Maniatis, T. (1991) Nucleic Acids Res., 19, 4675–4680.
9. Oliphant, A. R., Brandl, C. J. and Struhl, K. (1989) Mol. Cell. Biol., 9, 2944–2949.
10. Tsukiyama, T., Niwa, O., Yokoro, K. (1989) Mol. Cell. Biol., 9, 4670–4676.
11. Tsukiyama, T., Niwa, O. and Yokoro, K. (1991) J. Virol., 65, 2979–2986.
12. Tsukiyama, T., Ueda, H., Hirose, S. and Niwa, O. (1992) Mol. Cell. Biol., 12, 1286–1291.
13. Ueda, H., Sonoda, S., Brown, J. L., Scott, M. P. and Wu, C. (1990) Genes Develop., 4, 624–635.
14. Lavorgna, G., Ueda, H., Clos, J. and Wu, C. (1991) Science, 252, 848–851.
15. Smith, D. B. and Johnson, K. S. (1988) Gene, 67, 31–40.
16. Maxam, A. and Gilbert, W. (1980) Methods Enzymol. 65, 499–560.
17. Ueda, H. and Hirose, S. (1990) Nucleic Acids Res., 18, 7229–7234.
18. Fainsod, A., Margalit, Y., Haffner, R. and Gruenbaum, Y. (1991) Nuc leic Acids Res., 19, 4005.
19. Ueda, H. and Hirose, S. (1991) Nucleic Acids Res.., 19, 3689–3693.
20. Freeman, L. P., Luisi, B. F., Korszun, Z. R., Basavappa, R., Sigler, P. B. and Yamamoto, K. (1988) Nature 334, 543–546.
21. Eul, J., Meyer, M. E., Tora, L., Bocquel, M. T., Quirin-Stricker, C., Chambon, P. and Gronemeyer, H. (1989) EMBO J., 8, 83–90.