Structure and Expression of Mouse Cytokeratin Endo B Gene

A Doctoral thesis

submitted

by

Yuki Ichinose

to

Faculty of Science

Osaka University

1988

CONTENTS

1.ABSTRACT

In mouse early embryogenesis, the first differentiation was characterized as the appearance of two different types of cells, the inner cell mass and the trophectoderm during blastocyst formation. Cytokeratin endo B, the intermediate filament protein is expressed coordinately with the other cytokeratin (endo A) in trophectoderm cells specifically.

To study the molecular mechanisms of such differentiation, endo B gene which expression is characterized in stage specific and tissue specific manner was analyzed.

Endo B genes form a small gene family. By cloning of endo B genes, their structures were characterized and classified. From this result, one structural gene and one pseudogene were revealed. This structural gene consists of seven exons and six introns. The positions of six introns exactly coincided with those of introns previously reported in the other type I keratin genes, but the endo B gene lost the seventh intron which resided in the tail domains of other type I kertatins.

The comparison of the 5' flanking regions including the promoter of endo A and endo B genes revealed local homology around the respective TATA boxes. It's possible that these homologouse sequences might be involved in transcriptional and/or post transcriptional regulation. Endo B promoter region showed high homology with the bovine epithelial type I cytokeratin No.19, but doesn't have the consensus sequence, which observed in the 5' upstream region(-65 to -40) of epidermal cytokeratin

genes.

The regulation of endo B gene expression was investigated by the transfection of endo B/CAT fusion genes, and the assay of CAT activity. From such expriments it was found that the upstream sequence(s) had an enhancer-like activity to the endo B gene transcription.

2. INTRODUCTION

Intermediate filaments are major cytoskeletal proteins which constitute 8-10nm fibers in higher eukaryotic cells(see review Lazarides et al.,1980; Steinert et al.,1984; Fuchs et al.,1983). They are encoded by a large multigene family (Lazarides.1982; Steinert et al.,1985), and expressed in a tissue specific manner. Intermediate filaments are divided into five major groups: neurofilaments of neurons; desmin of myogenic cells; vimentin of mesenchymally derived cells; glial filament protein of astrocytes; and keratin of epithelial cells. Keratin proteins are subdivided into two groups, the acidic (type I) and basic(type II) keratins, A minimum of one acidic and one basic keratin is expressed in all epithelia. For assembly in vitro, keratin intermediate filaments are obligate hetropolymers (Steinert et al.,1982). Assembly both in vitro and in vivo requires one type I and one type II subunit, especially for the two-chain coiled-coil molecule(Woods and Inglis,1984; Parry et al.,1985; Margin et al.,1987). These two types of keratins share only limited sequence homology(<30%) which is restricted to the coiled-coil central domain(Fuchs et al.,1981,1984; Crewther et al.,1983; Hanukoglu and Fuchs,1983; Fuchs et al.,1984; Jarcano et al.,1984; Quinlan et al.,1984; Steinert et al.,1984; Sun et al.,1985)

The first differentiation step in mouse embryogenesis is characterized by the appearance of two different cell types, the inner cell mass and the trophectoderm. Both murine endo B and

endo A cytokeratin are coordinately expressed in trophectoderm cells specifically. These two cytokeratins form the intermediate filament(Duprey et al.,1985). A complete cDNA sequence of the endo B expressed in endodermal teratocarcinoma cells, has been described which classified endo B as a type I keratin-like protein. Recently Morita et al. (1988) classified the endo A protein as a type II keratin-like protein from its cDNA deduced amino acid sequence. The type I keratin-like protein endo B and the type II keratin-like protein endo A form a intermediate filament complex in the cytoplasm of epithelial cells (Oshima et al.,1981). The endo A genes were composed of one structural gene with one pseudogene (Vasseur et al.,1985; Brulet et al.,1986). Isolation of an endo B cDNA has enabled Singer to suggest that the mouse genomic contains about five copies of sequences homologous to endo B cDNA(Singer et al.,1986). Prior to this paper, however, the endo B structural gene was neither identified or characterized.

The present paper describes the structure of the endo B gene subfamily and the regulation of this gene expression, i.e., the endo B gene subfamily is characterized and classified, and the sequences of one structural gene and one pseudogene are analyzed and revealed the figure of the endo B gene. Furthermore, I investigate the regulation of the endo B gene expression, and discuss the mechanism of this gene expression.

3. MATERIALS AND METHODS

(a) Construction of the cDNA library prepared from PYS-2 cells and and isolation of endo B cDNA

Total RNA was prepared from mouse parietal yolk sak like teratocarcinoma cells(PYS-2), and selection of poly(A^+) RNA on oligo(dT) cellulose column was performed. 2 μg of poly(A^+) RNA was used for cDNA synthesis. Double stranded cDNA was made by the method of Gubler and Hoffman(1983). Internal EcoRI sites were methylated and attached with EcoRI linkers. After digestion with EcoRI, resulted DNA were passed through a Sepharose CL-4B column and ligated to λgt10 phage vector. DNA was packaged in vitro and transfected to E.coli C600 hfl cells(Huynh et al.,1985).

Phage DNAs of independent plaques($6X10^5$) were transfered to nitrocellulose filters, and hybridization were performed with 5' end-labeled synthetic oligonucleotides(63 mer) that was complementary to 3' untranslated sequence of endo B gene(Singer et al.,1986). After hybrididdation at 60°C for overnight nitrocellulose filters were washed by 6XSSC at 60°C for 15 min twice. After screening, fifteen positive plaques were obtained. Futhermore, isolated these positive plaques were hybridized with 5' end-labeled synthetic oligonucleotides(60 mer) which was complementary to 5' leader sequence of endo B gene, and three positive plaques were obtained. Inserted DNAs of these clones were excised with EcoRI, and subcloned into plasmid vector pUC19.

(b) Construction of the mouse genomic DNA library and isolation of endo B gene

Total DNA was isolated from the liver of a 129/SvJ mouse. The mouse DNA was digested partially with EcoRI, and the resulting fragments were fractionated by equilibrium centrifugation in sucrose gradients as described (Maniatis et al. 1986). DNA fragments of 15-20 kb size were selected, and the fragments were ligated to λ EMBL4 phage vector which was digested with EcoRI and SalI in order to remove polylinkers. The ligated DNA was packaged in vitro and transfected to E.coli P2392 cells. $6X10^5$ plaques were transfered to nitrocellulose filters and they were hybridized with a 5' end-labeled synthetic oligonucleotide (60 mer) that was complementary to the 5' leader sequence of the endo B gene(Singer et al.,1986). Hybridization was carried out at 60°C for 16 hr, and nitrocellulose filters were washed twice with 6XSSC for 15 min at 60 °C. After a second screening seventeen positive clones were obtained. The phage DNAs were extracted and their insert DNAs were analyzed by electrophoresis and southern hybridization.

Synthetic oligonucleotides of mouse B1(5'-GGCCTCGAACTCAGAAAT-3') and B2 (5'-TGCTCTTCCGAAGGTCCA-3') were also used as probes for mapping these repetitive sequences in the clones(Krayev et al.,1980, 1982).

(c) DNA sequencing

Restriction fragments of clone 23 were subcloned into M13 phage and sequenced by the dideoxynucleotide sequencing method (Sanger et al., 1977). The DNA sequence was analyzed and compared with published sequences using the SEQF, SEQH and SEQA programs of the NINCDS VAX(Digital Equipment Corporation) system.

(d) S1 Nuclease Analysis

S1 nuclease mapping was performed according to standard procedures(Berk and Sharp.,1977). Using the M13 putative promoter region(762 bp from position 1723 to 2485 in Fig.5) of the endo B gene as template a single stranded ^{32}P-labeled DNA complementary to the endo B RNA was synthesized. This cDNA was prepared using the method of Sanger and Coulson (1978), without the dideoxy nucleotides. After the resulting double stranded DNA was digested with EcoRI and HindIII, the single stranded labeled DNA was purified from a 7 M urea-4% polyacrylamide gel. 20 µg of RNAs were mixed with 100,000 cpm of ^{32}P-labeled probe. Hybridization was performed in 40 µl of hybidization buffer composed of 75% formamide, 0.4 M NaCl, 20 mM Tris-HCl(pH 7.4), 1 mM EDTA, and 0.1% SDS(Davis et al.,1986). Hybridization mixtures were first heated for 20 min at 75°C, prior to overnight incubation at 52°C. The hybridized DNAs were digested with S1

nuclease (50 U of S1 nuclease, 2 µg denatured salmon sperm DNA, 1 M NaCl, 0.6 M Sodium Acetate(pH 4.5), 0.04 M Zinc Sulfate) at 30°C for 1 hr. The S1 nuclease digestion was terminated and extracted twice using phenol:choloroform:isoamylalchol(25:24:1), The ethanol prepicitated samples were resuspended in 80% formamide, boiled for 3 min, and electrophoretically separated on a 7 M urea-6% polyacrylamide gel along with ^{32}P-end labeled pBR322 HpaII fragments as size markers. S1 digested fragments were visualized by autoradiography.

(e) Cell lines

The following cells were used: embrional carcinoma(EC) cells F9 (Bernstine et al.,1973), parietal yolk sac like cells PYS-2 (Lehman et al.,1974), placenta cell line(Log et al.,1981) and BALB/c 3T3 fibroblast cells.

Cells were maintained in Dulbecco's modified Eagles medium supplemented with 0.2% $NaHCO_3$, 0.5 mM glutamine, 10% (vol/vol) fetal carf serum, and 100 units/ml of penicilin, 0.05 mg/ml of streptmycin. F9 cells were maintained on gelatin-coated dishes.

To induce differentiation in retinoic acid(RA), F9 cells were plated at a density of 2 X 10^5 cells per 100 mm dishes, and 12-18 hr later RA and dibutyl-cAMP were added to final concentration 1 X 10^{-7} M and 1 X 10^{-3} M, respectively.

(f) Northern hybridization

Total RNAs were prepared from cultured cell lines by the guanidinium isothiocyanate procedure. Ten microgram of total RNAs were run on a 1% agarose, 1 X MOPS and 2.2 M formaldehyde gel, and transfer to the nitrocellulose filter. RNA hybridization to ^{32}P-labeled DNA was performed in 10 ml of hybridization buffer: 50% formamide, 4 X SSC, 5 X Denhardt's, 0.12 mg/ml of salmon sperm DNA and 0.2% SDS at 42°C for overnight. Hybridization washes were in 2 X SSC, 0.5% SDS at room temperature for 15 min twice, then in 0.1 X SSC, 0.1% SDS at 55°C for 15 min twice. Hybridized RNA fragments were visualized by autoradiography.

(g) Plasmid construction

The endo B promoter was isolated as AluI fragment. This 569 bp fragment was cloned into the HincII site of pUC18, and orientation at the cloning site was checked by digestion with appropriate restriction enzymes and by sequencing analysis. The endo B gene promoter was digested with BamHI and HindIII and the resulted 605 bp was ligated into plasmid pA10cat$_2$ digested with BglII and HindIII to eliminate the SV40 early promoter. Recombinant clone which digested with BamHI was filled by T4 DNA polymelase, then treated with BAP(bacterial alkaline phosphatase) at 37°C for 30 min and at 55°C for 30 min. Futhermore the endo B

genomic DNA(9.7 kb) containing the 5' and 3' flanking regions was digested with appropriate restriction enzymes, and separated to the six DNA fragments. No.1, No.2 DNA fragments covered 2.3 kb of the 5' upstream region, No.3, No.4 DNA fragments covered all of the introns and No.5, No.6 covered 3.6 kb of the 3' downstream region. Then these genomic DNA fragments of endo B gene were inserted into T4-polymerized BamHI site of recombinant vector (see Fig.11).

On the contraly, pA10cat$_2$ plasmid was digested with BglII and HindIII. After elimination of a short DNA fragment which is SV40 early promoter region, the produced cohesive ends of BglII and HindIII were polymerized and self-ligated. The obtained plasmid was designated as a pA10CATΔPRO, and used as a promoter-less CAT plasmid.

(h) Transfection and analysis of CAT activity

Each type of cells were grown to mid-phase and plated in 10 ml on 100 cm^2 dishes in 4-5 x 10^5 cells. Six from twelve hours after plating, the cells were transfected with 10 µg of virious types of CAT plasmids by DNA-$CaPO_4$ coprecipitation method(Graham and Van der EB, 1973; Wigler et al., 1978). The cells were incubated 4 hours and added Sodium Butylate to 5 mM, then incubated untill replacemant of the DNA-$CaPO_4$ containing medium with fresh medium. Forty eight hours after exchange to the fresh medium,

transfected cells were harvested, lysed by five cycles of freeze-thaw. And the enzyme solutions were prepared from the supernatant after centrifugation. All enzyme solutions preparated as cellular extract were analyzed for CAT activity by the method of Gorman et al.(1982).

4. RESULTS

(a) Isolation of complete cDNA clone for endo B

Complete cDNA sequences of subcloned endo B gene was shown in Fig.1. The cDNA sequence linked with EcoRI sites and the sequence is 1,374 nucleotides long, which is the same sequence as one reported by Singer et al.(1986) except two nucleotides substitution. One diffrence is AAG for AAA at position 261, this substitution resulted in no change of amino acid sequence (Lysine). But another substitution TTT for CTT at position 442 resulted in the change of amino acid Leucine for Phenylalanine. These two substitutions of nucleic acid may depend on the difference of the cultured cell line. The deduced amino acid sequence also supportes that the structure of endo B protein is composed of a central, α-helical rod domain flanked by non α-helical head and tail domains which vary between the particular proteins.

(b) Isolation of the mouse endo B cytokeratin gene

As described by Singer et al. (1986), Southern blot analyses of EcoRI or BamHI digested 129/SvJ mouse DNA, revealed that 4-5 genes were strongly homologous with the endo B cDNA probe(Singer et al. 1986). My genomic DNA library from a 129/SvJ mouse was

screened with endo B cDNA. From 17 positive plaques, we obtained 4 independent genes which were homologous to endo B. The maps of clone 23, 4, 10, and 14, which represent each of the four genes, are shown in Fig. 3. The southern blot suggests that one more gene homologous to endo B may also exist, with a 6.8 kb size estimated from the EcoRI fragment(Singer et al. 1986). The fifth gene was not isolated, but four of the five genes homologous to endo B were isolated and characterized. Using the 5' probe (synthetic 60 mer complementary to the 5' leader and N-terminal sequence of endo B gene) and the 3' probe (synthetic oligonucleotide corresponding to the 3' noncoding region), the DNAs of four clones (23, 4, 10, 14) were hybridized. The results showed that all of the four clones hybridized with the 5' probe, clone 23,10 14 hybridized with the 3' probe, and only two clones (23 and 4) had internal EcoRI sites between the 5' and 3' hybridizing portions. Because the endo B cDNA from parietal endoderm cells has an internal EcoRI site, clones 23 and 4 were the only possible candidates for the endo B gene expressed in parietal endoderm. The fifth clone was predicted not to have such internal EcoRI sites by Southern blot analysis. As shown in Fig.3 the gene of clone 23 extends 9.2 kb, while the clone 4 gene is extremely compact(2.1kb), including the 5' and 3' hybridizing portion. To confirm which of these clones is a endo B structural gene, sequencing analyses were performed.

(c) Nucleotide Sequence of the endo B cytokeratin gene

By the sequencing analyses, it was revealed that clone 4 was a processed pseudogene corresponding to the endo B cDNA without intron, and clone 23 corresponds to the endo B gene which is expressed in parietal endoderm.

As shown in Fig.4, homologous region between the sequences of clone 4 and clone 23 corresponding to the endo B cDNA extends for 460 nucleotides, including 430 matches, 29 replacements, 2 deletions and 1 insertion in clone 4. There is 93% overall homology for the 462 bp compared between clone 4 and clone 23, while the sequence of clone 23 is exactly the same as endo B cDNA. Since evidence has been accumulated which shows a rapid and constant rate of divergence of psuedgene sequences of 400 PAMs (1 PAM = 1 accepted point mutation per 100 residues per 100 million years) calculated amino acid data. I could culculate the approximate time of divergence of this pseudogene(clone 4) from its parent endo B gene(clone 23). Translating the coding sequence into amino acid, and comparing this sequence to the murine endo B amino acid sequence deduced from cDNA sequence of PYS-2 cells, I found 15 substitution and 1 nonsense codon , for a total of 16 mutational events. For a protein which originally had 137 amino acids, I estimated the time of divergence as 2.9 million years ago.

The nucleotide sequence of the genomic clone 23(λEB23) which contains the complete mouse endo B gene is shown in Fig.5. In order to define the exon-intron organization of the endo B gene

and clarify the structural basis of endo B gene expression, I determined the nucleotide sequence of λEB23, including both the 3'(2kb) and 5'(1.7kb) flanking regions. From the comparison of the sequence with that of endo B cDNA, we identified the 7 exons and 6 introns of the endo B gene. As shown in Fig.6, the first, second and third exons encode the head and "1A" and "1B" domain of the α-helical coiled coil, while the forth, fifth and sixth exons encode the linker and the latter part of α-helical domain "2"(Singer et al.,1986). The seventh exon encodes for the non α-helical tail region and non-translated 3' sequence. The consensus sequences for splicing junctions, GT and AG (Breathnbach et al., 1978), are conserved at each site.

(d) Transcription initiation sites of the mouse endo B cytokeratin gene

S1 nuclease mapping was carried out to determine the transcriptional start point in PYS-2 cells. Mouse differentiated teratocarcinoma PYS-2 mRNA was hybridized with the M13 single stranded probe, treated with S1 nuclease, and subjected to gel electrophoresis as described. The major protected bands appeared, and it shows the cap site of the endo B mRNA (Fig.7, lane 2). Thus, the transcription of this gene was expected to start at 58 bp upstream from the initiation codon ATG. About 26 bp upstream from the putative cap site locates the Goldberg-Hogness

box-like sequence GATATAA which is commonly found in eucaryotic genes (Sadler et al.,1983; Efstratiadis et al.,1980), but we did not find a CCAAT box in the 5' flanking region. Sixteen and twenty-one nucleotides downstream from the cap site are found the sequences CTCCTG, GTTCTG, respectively. These sequences are homologouse to the sequence CTTCTG, a consensus sequence that frequently appears in the 5' noncoding region of eucaryotic mRNA (Baralle and Brownlee, 1978).

(e) B1 repeat sequences in the 5' and 3' flanking regions of the endo B gene

In the 5' and 3' flanking regions of the endo B gene, we found four B1 repeat sequences which were first described by Krayev et al. (1980) without an ascribed function.

The first B1 repeat was found about 0.3 kb upstream of the initiation codon. The second and third B1 repeats were clustered in the 5' flanking region 1.2 kb downstream from the polyadenylation site. The forth B1 repeat was located more than 2 kb downstream of the other B1 clusters. Three of the B1 repeats were sequenced, revealing that the two had the same orientation in the endo B gene, while the third had the reverse orientation. The homology of these B1 repeats to the consensus sequence is shown in Fig.8.

(f) Expression of the endo B gene and B1 repetitive elements in mouse cultured cell lines

To estimate the level of expression of the endo B mRNA in various cell lines by northern blot hybridization(Fig.9) the 3' region (ca.1000 bp) of the endo B cDNA was used as a probe. Strong hybridization signals were observed in PYS-2 and placenta cells, but in F9 cells only weak signal was observed and in 3T3 cells such a signsl didn't appear at all. These sizes of hybridization signals were estimated by the mobility and RNA staining, and showed 1.5 kb size, which consisted with the described data(Singer et al.,1986). It was known that retinoic acid treatment of F9 cells induced its differentiation in vitro. The longer F9 cells exposed to retinoic acid, the more F9 cells expressed the endo B mRNA. In fact, the endo B transcripts seems to be derived from the differentiated cells after retinoic acid treatment. The appearance of differentiated cells were tested by indirect immunofluorescence staining using an antibody to keratin filament(TROMA-1).

After the hybridization with the endo B cDNA, the same nitrocellulose filter was hybridized with plasmid pUBPRO containig B1 repetitive sequence derived from 5' upstream region of the endo B gene as AluI DNA fragment(see Fig.11). In all of cell lines examined, many sizes of B1 repeat transcripts were observed, and in F9 cells B1 repeats were transcribed much actively as retinoic acid treatment was continued. However, there is no significanct difference between the cells to express the

endo B and the ones not to express.

(g) Regulatory regions for the transcription of mouse endo B gene

To test if any cell type specific enhancer element, acting in cis, was located in the 5' and 3' flanking regions and introns of the endo B gene, six kinds of endo B DNA fragments were inserted into BamHI site downstream of CAT gene, and pABCAT plasmid series were constructed. pSV_2CAT plasmid, which had a SV40 enhancer element, early promoter and CAT gene, was used as apositive control. $pA10CAT_2$ plasmid, which had a only SV40 eaely promoter and CAT gene, was used as a enhancer-less plasmid. The CAT activity of the plasmid pABCAT, which contained the 5' upstream region of the endo B gene(-506 to +63) as an endo B promoter, was compared with those of the plasmid pAB1CAT to pAB6CAT, which were constructed by the ligations of the pABCAT and endo B DNA fragments. The plasmid pA10CAT PRO was constructed from the $pA10CAT_2$ plasmid by the elimination of the SV40 early promoter, and used as a promoter-less negative control. These ten kinds of CAT plasmid were transfected to the various types of cell lines, PYS-2, placenta cells, F9 cells and 3T3 fibroblast. As a preparatory expriments, I transfected various amount of CAT plasmid, pSV_2CAT, $pA10CAT_2$, pABCAT and pA10CAT PRO, to each kind of cell lines. In all of cell lines, 3 μg, 5 μg, 10 μg of plasmid

DNA was transfected. From these results, I selected the condition that 10 ug of plasmid DNA was used for transfestion, because in this condition a weak CAT activity was observed in the plasmid pABCAT. In 3 μg and 5 μg of transfection conditions, such an activity could not be observed. When I used 20 μg and 30 μg of plasmid DNA for transfection, all of test plasmids represented high level of CAT activity, and I couldn't obtain the significant difference between the test plasmids. Then 10 μg of CAT plasmid was used for transfection.

In PYS-2 cells, from the comparison of the CAT activity between pABCAT and other test plasmids, pAB1CAT to pAB6CAT, it was clear that one of those plasmid, pAB1CAT had an activity to enhance the endo B transcription(Fig.12). In placenta cells and F9 cells, similar CAT activities were also observed, i.e., the CAT activity of the plasmids, pAB1CAT, pAB5CAT and pAB6CAT are stronger than control plasmid, pABCAT(Fig.13 and Fig.14). On the contraly, in 3T3 cells the CAT activity of pABCAT1 is a little stronger than pABCAT, but the CAT activity of pABCAT5 and pABCAT6 showed only the same degree with pABCAT(Fig.15).

In all types of cell lines, the 5' upstream region of pAB1CAT plasmid showed enhancer-like activity. This activity didn't show the cell type specificity of the endo B gene expression. pAB5CAT and pAB6CAT plasmid also showed a weak level of enhancer-like activities in placenta cells, which express the endo B gene, and in F9 cells, which don't express endo B gene. Through these CAT assay experiments, pAB3CAT and pAB4CAT, which contain the introns region, showed the decreased CAT activity by the

comparison with pABCAT. This fact may suggest the presence of some negative regulatory element in these regions. From the survey of the endo B gene for enhancer elements, I identified two types of regulatory elements, i.e., enhancer-like element and silencer-like element.

5.DISCUSSIONS

(a) Conservation of exon and intron patterns

In the multigene family of the intermediate filaments, many researcheres reported that the genes for type I,II,III intermediates filament arose from a common ancestor (Krieg et al., 1985). As shown in Fig.16 a comparison of the nucleotide sequences of the murine cytokeratin endo B gene with the human, murine and bovine type I keratin genes revealed that the locations of the first six introns were conserved. Present data of endo B gene also supports this evolutional model. Other type I keratin genes (human 50 kD keratin Marchuk, 1984, 1985; mouse 59kD keratin Krieg, 1985; and bovine cytokeratin BVI Rieger, 1985) contain the additional seventh intron in their tail regions. But the regions were less conserved and more characteristic to their respective genes. However as a peculiarity of endo B gene, this gene lost a seventh intron.

(b) Significance of B1 repetitive sequence

The murine B1 repeat presents 10^5 copies per haploid mouse genome, so that on average, every 10-30 kb of mouse genomic DNA should contain one B1 repeat. Thus, the four B1 repeats in the 16 kb length of clone 23 showed an unusal abundance in endo B gene.

The other clones representing the endo B subgene family were also hybridized with a synthetic oligonucleotide corresponding to the B1 repeat(Fig.3). Three out of the four clones contained the B1 repeat. On the contrary, the endo A gene contains the B2 repeat in the third intron, and its involvement in endo A gene expression has been discussed(Vasseur et al., 1985).

The mouse endo B gene and its subfamily is accompanied by a B1 repeat but not by a B2 repeat(Fig.3). Though there is no significant difference between undifferentiated cells and differentiated cells in the B1 expression, and the function of the B1 repeat has not been clarified yet, the B1-1 repeat located immediately 0.3 kb upstream from the first exon might influence the regulation of the expression of endo B gene.

(C) Homology of the 5' upstream regions between endo B gene and the other intermediate filaments

I compared the sequence of the upstream region of the endo B gene with those of other intermediate filaments. Blessing et al. (1987) described that the 5' region of the epidermal keratin genes are remarkably conserved between different species (bovine, murine and human). They also found a consensus sequence AAPuCCAAA located upstream(-65 to -40) of the TATA box, specifically in epidermal cytokeratin genes.

The endo B and endo A are known to be cytokeratins only in simple epithelia and not that of the epidermal type. Though we identified similar sequences (AAACAAAA,-2064 to -2057), AAGCCAGA (-1320 to -1313), AAAACCAAG(-609 to -602) in the upstream region of the endo B gene, their locations were rather far from the TATA box.

I found local homologies around the promoter region between the endo B gene and other type I keratin genes. In particular, bovine type I cytokeratin No.19, which is expressed in the simple epithelial cells(Bader et al., 1986) contains sequences, which are not only highly homologies to the endo B gene, but also these homologies are located in the same way for both genes. As shown in Fig.17, the 5' upstream regions of the two type I cytokeratin genes were aligned to get the optimal homology. The sequences around promoters in comparable positions were found to share about 67% homology. The sequences near the cap sites and the transcribed 5' noncoding regions also shared substantial

homology. These two type I cytokeratins, endo B and bovine No.19, are expressed in epithelial cells. It is possible that the expression of these genes are regulated under the similar mechanisms.

On the contrary, the mouse endo A gene is expressed with the endo B gene in the same cells and in the same stages of the development. The homologous upstream sequences of these genes may explain their coordinate expression.

As shown in Fig.18, the upstream regions of the endo A and endo B genes were aligned. The endo A and endo B genes share two concentrated homologous sequences around the TATA boxes. One sequence is CTGGGGCGTGGCCT, which is located in the upstream regions of the TATA box. Another homologous sequence is GCTTCGCTCTCCTCTC, which is located betwen the TATA box and initiation site, ATG. Since this latter sequence is transcribed, it might be involved in post-transcriptional regulation. Whether any of these similarities are significant in the coordinate regulation of these genes remains to be investigated.

(d) Regulatory elements for the transcription of endo B gene

Regulation of gene expression in the mouse early embryogenesis is not understood well. In particular, the cytokeratin endo B and endo A seem to be expressed in coordinate manner, because these types of cytokeratins make coiled coil heterotetramer to form intermediate filament. However, the investigations of the regulatory mechanisms of cytokeratin gene expression, have not been successful including the promoter assay using CAT reporter gene so far. The endo A and endo B are the major proteins of early mouse embryos by comassie-blue staining of two-dimential gels, so that It is possible that there may be some enhancer element for the strong transcription of these genes during the embryonic development.

From the survey of the enhancer element, it was observed that pAB1CAT had an activity to enhance the endo B transcription. Another two plasmid, pAB5CAT and pAB6CAT represent a weak enhancer-like activity, while pAB3CAT and pAB4CAT represent a silencer-like activity. It is possible that regulatory element(s) consist in the plural regions.

The pAB1CAT which contains the 5' upstream region of the endo B promoter gave the enhancer-like activity also in all of types cells which were tested so far. This activity didn't show the cell type specificity of endo B gene expression.

From these results, it is supposed that there may be two regulatory factors at least. One is a positive factor of the endo B gene expression, which enhances the endo B gene transcription,

and another one is a negative regulatory factor.

Cell type specificity of the endo B gene expression depends on the inner-nuclear circumastance, i.e., the presence of some positive regulatory factor(s) and negative regulatory factor(s). In murine undifferentiated embryonal carcinoma cells, the presence of such negative regulatory factor(s) have been suggested. For examle, Crèmisi and Duprey reported that the transcription of endo A gene which expressed coordinately with endo B, is blocked by a labile inhibitor in murine undifferentiated embryonal carcinoma cells, PCC4(1987). It is possible that endo B gene expression is also blocked by a similar or the identical labile inhibitor(negative factor) in the undifferentiated F9 cells.

On the other hand, it is known that embryonal carcinoma cells are resistant to infection by polyoma virus and adenovirus (Herbomel et al.,1984; Cremisi and Babinet,1986). In the case of adenovirus infection, E1a which is a virul product from adenovirus early promoter can activate E2 late promoter and repress the E1a enhancer itself(Imperiale et al., 1984; Borrelli et al., 1984). Moreover, such an adenovirus E1a-like activity was suggested to exist as a cellular transcriptional factor in normal cells(Imperial et al.,1984). Just as the adenovirus E1a transcriptional factor can repress enhancer-dependent transcription in differentiated cells, this E1a-like transcriptional factor might repress some genes during the differentiation of murine embyonal carcinoma stem cells.

In the regulation of the endo B gene expression, the negative

regulatory factor(s) are supposed to be present only in an undifferentiated cells(F9), and repress the gene expression of the endo B. While in a differentiated cells, PYS-2 and placenta cells, a positive regulatory factor(s) are present instead of the negative regulatory factor(s). Then, endo B gene can be transcribed in a differentiated cells. It is possible that the expression of the endo A gene is also regulated by the same mechanisms.

I surveyed the homologous sequences with the published core sequences of enhancer elements, and found the several similar sequences in the 5' upstream regions, which had an enhancer-like activity to the endo B transcription in the CAT assay experiments. As shown in Table 1, the homologouse sequences with adenovirus E1a (Hearing and Shenk, 1983), polyoma virus enhancer (Melin et al., 1985; Veldman et al., 1985) and Human HSP70 enhancer core sequences (Wu et al., 1985) were found in the 5' upstream regions, AGAACTGAAA(31 to 40), AGAAGTGAAA(215 to 214), and in the 3' downstream regions, GGAGGTGAAA(6253 to 6262) and AAAAGGAAA(7336 to 7345). It is posible that these sequences may act as an acceptor elements for the cellular E1a-like regulatory factor(s).

As the enhancer-like activity in PYS-2 cells is much stronger than in F9 cells, it was suggested that the positive regulatory factor(s) also present in F9 cells, but these amounts were very different.

Furtheremore, in 3T3 fibroblast cells, low level of enhancer-like activity was observed. It suggests that the 5' upstream

region which is inserted as a endo B promoter regions, also had a non-specific positive regulatory elements. In this region, I found a little homologous sequences with SV40 enhancer core sequence (Weiher et al., 1983) at just upstream of TATA box, GTGGCTGG(-75 to -68) and GTGGCCCCTG(-64 to -57). However these levels of homologies were able to find in overall sequences of the endo B gene, even in the coding regions.

At the moment, this type of research, the molecular mechanisms of the endo B and endo A gene expresion in the mouse esrly embryogenesis, has just been started. The isolation of the concrete molecules of the positive and negative regulatory factors, and the detailed identification of the enhancer and silencer elements, and their functions must await further studies.

6.ACKOWLEDGEMENTS

I would like to thank Dr.A.Mastushiro, Dr.T.Morita and Dr.M.Nozaki for thier valuable discussions advices and technical assistance. I am grateful to the other members of my laboratories for useful suggestions, in particular Mrs. Zhang Fengyn for the continuing help . I thank Dr.T.Log for the gift of mouse placenta cell line.

7.REFERENCES

Bader,B.L.,Margin,T.M.,Hatzfeld,M. and Franke,W.W.(1986) EMBO J, 5, 1865-1875.

Baralle,F.E., and Brownlee,G.G. (1978) Nature,274,84-87.

Berk,A.J. and Sharp,P.A. (1977) Cell,12,721-732.

Bernstine,E.G., Hooper,M.L., Grandchamp,S. and Ephrussi,B. (1973) Proc.Natl.Acad.Sci. USA 70, 3899.

Blessing,M., Zentgraf,H. and Jarcano,J.L. (1987) EMBO J.,6 567-575.

Borrelli,E., Hen,R. and Chambon,P. (1984) Nature, 312, 608-612.

Breathnach,R. and Chambon,P. (1981) Ann.Rev.Biochem.,50,349-383.

Brûlet,P., Duprey,P., Vasseur,M., Kaghad,M., Morello,D., Blanchet,P., Babinet,C., Condamine,H. and Jacob,F. (1985). In Cold Spring Harbor Symposia on Quantitative Biology,vol.50: Molecular Biology of Development.pp51-57. Cold Spring Harbor Laboratory, New York.

Crémisi,C. and Babinet,C. (1986) J.Virol. 59, 761-763.

Crémisi,C. and Duprey,P. (1987) Nucle.Acids.Res. 15, 6105-6116.

Crewther,W.G., Dowling,L.M., Steinert.P.M. and Parry,D.A.D., (1983) Int. J. Biol. Macromol. 5, 267-274.

Davis,L.G., Dibner,M.D., Battey,J.F. (1986) Elsevier Science Publishing Co., Inc. 276-284.

Duprey,P., Morello,D., Vasseur,M., Babinet,C., Condamine,H., Brulet,P. and Jacob,F. (1985) Proc. Natl. Acad. Sci. USA. 82, 8535-8539.

Efstratiadis,A.,Posakony,J.W.,Maniatis,T.,Lawn,R.M.,O'Connell,C., Spritz,R.A.,DeReil,J.K.,Forget,B.G.,Weissman,S.M.,Slightom,J.L., Blechl,A.E.,Smithies,O.,Baralle,F.E.,Shoulders,C.C.& Proudfoot, N.J. (1980) Cell 21, 653-668.

Fuchs,E. and Green,H. (1981) Cell, 25, 617-625.

Fuchs,E. and Hanukoglu,I. (1983) Cell, 34, 332-334.

Gorman,C.M., Moffat,L.F. and Howard,B.H. (1982) Mol. Cell Biol., 2, 1044-1051.

Graham,F. and A.van der Eb. (1973) Virology, 52, 456-457.

Gubler,U. and Hoffman,B.J. (1983) Gene, 25, 263-269.

Hearing,P. and Shenk,T. (1983) Cell, 33, 695-703.

Herbomel,P.,Bourachot,B., and Yaniv,M. (1984) Cell,39,653-662.

Huynh,T., Young,R.A. and Davis,R,W. (1985) In DNA cloning, vol.I: a practical approach. (ed.D.M.Glover), pp49-78. IRL Press, Oxford.

Imperiale,M.J. and Nevins,J.R. (1984) Mol.Cell.Biol. 4, 875-882.

Jorcano,J.L., Rieger,M., Franz,J.K., Schiller,D.L., Moll,R., and Franke,W.W. (1984) J.mol.Biol. 179, 257-281.

Krayev,A.S.,Kramerov,D.A.,Skryabin,K.G.,Ryskov,A.P.,Bayev,A.A. and Georgiev,G.P. (1980) Nucle. Acids. Res. 8,1201-1215.

Krayev,A.S., Markusheva,T.V., Kramerov,D.A., Ryskov,A.P., Skryabin,K.G., Bayev,A.A. and Georgiev,G.P. (1982) Nucle. Acids.Res. 10, 7461-7475.

Krieg,T.M., Schafer,M.P., Cheng,C.K., Filpula,D., Flaherty,P., Steinert,P.M. and Roop,D.R. (1985) J.Biol.Chem.,260,5856-5870.

Lazarides,E. (1980) Nature, 283, 249-256.

Lazarides,E. (1982) Ann. Rev. Biochem., 51,219-250.

Lerman,J.M., Speers,W.C., Swartzendruber,D.E., and Pierce,G.B. (1974) J.Cell.Physiol. 84,13-28.

Log,T.,Chang,K.S.S. and Hsu,Y.C. (1981) Int.J.Cancer.,27,365-372.

Maniatis,T., Frisch,E.F. and Sambrook,J. (1982) Molecular Cloning. A Laboratory Manual. Cold Spring Harbor Laboratory Press, NY.

Margin,T.M., Hatzfeld,M. and Franke,W.W. (1987) EMBO J, 6, 2607-2615.

Marchuk,D.,McCrohon,S. and Fuchs,E.(1984) Cell. 39, 491-498.

Marchuk,D.,McCrohon,S. and Fuchs,E.(1985) Proc.Natl.Acad.Sci.USA, 82,1609-1613.

Melin,F.,Pinon,H.,Reiss,C.,Kress,C.,Montreau,N. and Blandy,D. (1985) EMBO J, 4, 1799-1803.

Morita,T., Tondella,M.L.C., Takemoto,Y., Hashido,K., Nozaki,M., Matsushiro,A. (1988) Gene, in press.

Oshima,R.G. (1981) J.Biol.Chem.,256,8124-8133.

Parry,D.A.D., Steven,A.C. and Steinert P.M. (1985) Biochem. Biophys. Res. Commun. 127, 1012-1018.

Quinlan,R.A., Schiller,D.L., Hatzfeld,M., Achtstatter,T.,Moll,R., Jarcano,J.L.,Magin,T.M. and Franke,W.W. (1985) In Wang,E., Fischman,D.,Liem,R.K.H. and Sun,T.-T. (eds) Intermediate Filaments. Ann. New. York Acad. Sci., 455,282-306.

Rieger,M.,Jorcano,J.L. and Franke,W.W.(1985) EMBO J.4,2261-2267.

Sadler,J.R.,Waterman,M.S. & Smith,T.S. (1983) Nucleic Acids Res. 11,2221-2231.

Sanger,F., Nicklen,S. and Coulson,A.R.(1978) Proc.Natl.Acad.Sci. USA.74,5463-5467.

Singer,P.A.,Trevor,K. and Oshima,R.G. (1986) J. Biol. Chem,261, 538-547.

Steinert,P.M., Idler,W.W., Aynardi-Whitman,M., Zackroff,R V. and Goldman,R.D. (1982) Cold Spring Harbor Symp. Quant. Biol. 46, 465-474.

Steinert,P.M., Parry,D.A.D., Racoosin,E.L., Idler,W.W., Steven,A.C., Trus,B.L. and Roop,D.R. (1984) Proc. Natl. Acad. Sci. USA, 81, 5709-5713.

Steinert,P.M.,Steven,A.C. and Roop,D.R. (1985) Cell, 42, 411-419.

Sun,T.-T., Tseng,S.C.G., Huang,A.J.-W., Cooper,D., Schermer,A., Lynch,M.K., Weiss,R. and Eichner,R. (1985) In Wang,E., Fischman,D., Liem,R.K.H. and Sun,T.-T. (eds), Intermediate Filaments. Ann. New York Acad. Sci., 455, 307-329.

Vasseur,M., Duprey,P., Brulet,P. and Jacob,F. (1985). Proc. Natl. Acad. Sci.USA.82,1155-1159.

Vasseur,M., Condamine,H. and Duprey,P. (1985) EMBO J.4,1749-1753.

Veldman,G.M.,Lupton,S. and Kamen,R. (1985) Mol.Cell.Biol. 5, 649-658.

Weiher,H., Konig,M. and Gruss,P. (1983) Science, 219, 626-631.

Wigler,M., Pellicer,A., Silverstein,A. and Axel,R. (1978) Cell, 14, 725-731.

Woods,E.F. and Inglis,A.S. (1984) Int. J. Biol. Macromol. 6, 277-283.

Wu,B., Kingston,R. and Morimoto,R.I. (1986) Proc. Natl. Acad. Sci. USA, 83, 629-633.

8.REGENDS TO FIGURES

Fig.1. Nucleotide sequence and deduced amino acid sequence of the endo B cDNA. Central α-helical coiled coil domains were boxed.

Fig.2. Southern blot analysis of mouse genomic and cloned phage DNAs (clone number 23,4,10,14). Mouse genomic DNA and cloned phage DNAs were digested with EcoRI, hybridized with 5' region of endo B cDNA about 500 bp(A) and 3' region about 1000 bp(B).

Fig.3 Genes homologous to endo B. Positions where the 5' and 3' specific probes derived from cDNA hybridized are marked 5' and 3', respectively. Hybridized DNA fragment sizes are represented under the white boxes(cloned genes) and the dashed box(not isolated). Internal EcoRI sites corresponding to the endo B cDNA sequences are indicated by arrows. The ability of each of the clones to hybridize to synthesized B1 and B2 repetitive sequences is shown as positive(+) or negative(-).

Fig.4. Nucleotide sequence of the mouse cytokeratin endo B and the pseudogene(clone 4). Sequence of the structural gene(clone 23) from transcription initiation site is shown under the pseudogene sequence. The deletions of the nucleotides are shown as dashes in the sequence.

Fig.5. Nucleotide sequence of the gene coding for mouse cytokeratin endo B. The 5' and 3' flanking regions of endo B gene are also shown. The amino acid sequence encoded by the exons are indicated in single letter codes. Nucleotides are numbered from 1 to 7878. The sequence of synthetic oligonucleotide (60mer) complementary to the 5' region of endo B gene is shown by a dashed underline. The introns are marked by arrows and numbered. Putative promoter sequence, GATATAA and the polyadenylation signal are boxed. The transcription initiation site(cap site) is shown by an arrow head. The mouse B1 repetitive elements are boldly underlined.

Fig.6 Structure of the mouse endo B gene (λEB23). (a) Structure of the genomic DNA. The sites for EcoRI(E) and HindIII(H) are shown. Seven exons are boxed, The 5' and 3' flanking regions and introns are shown by thin lines. Black arrows represent B1 repetitive elements. (b) The endo B mRNA. (c) Positions of intron-exon junctions in the endo B protein which consists of α-helical central domains(shown by white boxes) and head, tail and linker regions(shown by lines).

Fig.7 S1 nuclease mapping of the endo B transcription initiation site. Experimental details are in the MATERIAL AND METHOD. Lane 1:molecular weight maker, pBR322 digested with HpaII; Lane 2:^{32}P labeled DNA probe + 20μg of total RNA from PYS-2 cells; Lane 3:^{32}P labeled DNA probe + 20μg of yeast tRNA. Transcription initiation site(cap site) is shown by arrow head.

Fig.8. Mouse B1(Alu-like) repetitive sequences which reside in the flanking region of the endo B cytokeratin gene. Only different nucleotides from consensus sequence are indicated. The synthetic oligonucleotide(18 mer) sequence which was used as a DNA probe is boxed.

Fig.9. Northern blot hybridization of endo B mRNA. 10 μg of each RNA were run on a 1% agarose gel. The probe used was the 3' region of endo B cDNA(about 1 kb). Nick-translated DNA prove has 1×10^8 cpm/μg of specific activity. F9, embryonal carcinoma cells; F9 cells treated with 1×10^{-7} M retinoic acid for 1,2 or 3 days; PYS-2, parietal yolk sac like cells; Placenta, placenta cells; 3T3, fibroblastic cells.

Fig.10. Northern blot hybridization of B1 repeat. After hybridization with endo B cDNA(Fig.9), the same nitrocellulose filter was hybridized with B1 repeat containing plasmid, pUBPRO without washing. Nick-translated DNA probe has 3×10^8 cpm/μg of specific activity.

Fig.11. Construction of the plasmids containing endo B gene promoter to test enhancer activity by CAT.

Fig.12. Survey of the endo B gene for enhancer elements. A representative CAT assay of the endo B gene fragment is shown. Each recombinant plasmid was transfected to PYS-2 cells.

Fig.13. Survey of the endo B gene for enhancer elements. A representative CAT assay of the endo B gene fragment is shown. Each recombinant plasmid was transfected to placenta cells.

Fig.14. Survey of the endo B gene for enhancer elements. A representative CAT assay of the endo B gene fragment is shown. Each recombinant plasmid was transfected to F9 cells.

Fig.15. Survey of the endo B gene for enhancer elements. A representative CAT assay of the endo B gene fragment is shown. Each recombinant plasmid was transfected to 3T3 cells.

Fig.16. Schematic diagram showing the location of introns in type I cytokeratin genes of known structure. The α-helical domains(coil 1a, 1b and 2) are shown by shaded boxes, and head, tail and linker regions are shown by lines. The locations of introns are shown by arrow heads.

Fig.17. Alignment of the 5' upstream regions of the two type I cytokeratin genes, the mouse endo B and the bovine No.19(Bader et al.,1986). The 5' upstream regions of the endo B and the bovine No.19 cytokeratin genes were aligned for optimal homology. Putative promoter sequences and the translation initiation sites ATG are boxed. The cap sites are underlined.

Fig.18. The similarities in the 5' upstream regions of the endo A and endo B genes. For both endo A and endo B genes the sequences of 100 nucleotides preceding the cap sites were aligned and numbered relative to the capped nucleotide(position +1), and the putative promoter sequences are placed between brackets. The translation intiation sites ATG stressed by double lines. The substantial degree of local homology are shown by the boxed sequences in the 5' flanking regions, and by the boldly underlined sequences in the 5' noncoding regions.

Fig.1


1
CTCCTGTTCTGGTCTCTCGCTTCGCTCTCCTCTCCAGACAAG
43
ATG AGC TTC ACA ACT CGC TCC ACC ACC TTC TCC ACC AAC TAC CGG TCC CTG GGC TCT GTG CGA ACT CCC AGC CAG CGG GTC CGG CCT GCC
1
M S F T T R S T T F S T N Y R S L G S V R T P S Q R V R P A
133
AGC AGC GCA GCC AGC GTC TAT GCA GGT GCT GGG GGC TCC GGG TCC CGG ATA TCC GTG TCC CGC TCT GTC TGG GGT GGC TCT GTG GGG TCC
31
S S A A S V Y A G A G G S G S R I S V S R S V W G G S V G S
223
GCA GGC CTG GCG GGA ATG GGT GGA ATC CAG ACC GAG AAG GAG ACC ATG CAA GAC CTG AAC GAT CGC CTG GCC AGC TAC CTA GAC AAG GTG
61
A G L A G M G G I Q T E K E T M Q D L N D R L A S Y L D K V
1A
313
AAG AGC CTG GAA ACT GAG AAC AGG AGA CTG GAG AGC AAA ATC CGG GAA CAT CTG GAG AAG AAG GGG CCC CAG GGC GTC AGA GAC TGG GGC
91
K S L E T E N R R L E S K I R E H L E K K G P Q G V R D W G
403
CAC TAC TTC AAG ATC ATC GAA GAC CTG AGG GCT CAG ATC TTT GCG AAT TCT GTG GAC AAT GCC CGC ATC GTC TTG CAG ATC GAC AAT GCC
121
H Y F K I I E D L R A Q I F A N S V D N A R I V L Q I D N A
493
CGC CTT GCC GCC GAT GAC TTT AGA GTC AAG TAT GAG ACA GAA CTA GCC ATG CGC CAG TCT GTG GAG AGC GAC ATC CAT GGA CTC CGC AAG
151
R L A A D D F R V K Y E T E L A M R Q S V E S D I H G L R K
1B
583
GTG GTA GAT GAC ACC AAC ATC ACA AGG CTG CAG CTG GAG ACA GAA ATC GAG GCA CTC AAG GAA GAA CTT CTG TTC ATG AAG AAG AAT CAT
181
V V D D T N I T R L Q L E T E I E A L K E E L L F M K K N H
673
GAA GAG GAA GTC CAA GGT CTG GAA GCC CAG ATT GCC AGC TCT GGA TTG ACT GTG GAA GTG GAT GCC CCC AAA TCT CAG GAC CTC AGC AAG
211
E E E V Q G L E A Q I A S S G L T V E V D A P K S Q D L S K
763
ATC ATG GCG GAC ATC CGC GCC CAG TAT GAA GCG CTG GCT CAG AAG AAC CGC GAG GAA CTG GAC AAG TAC TGG TCT CAG CAG ATT GAG GAG
241
I M A D I R A Q Y E A L A Q K N R E E L D K Y W S Q Q I E E
853
AGT ACC ACA GTT GTC ACC ACC AAG TCT GCC GAA ATC AGG GAC GCT GAG ACC ACA CTC ACG GAG CTG AGA CGC ACC CTC CAG ACC TTG GAG
271
S T T V V T T K S A E I R D A E T T L T E L R R T L Q T L E
2
943
ATT GAC TTG GAC TCC ATG AAA AAC CAG AAC ATC AAC TTG GAG AAC AGC CTC GGG GAT GTG GAG GCC CGA TAC AAG GCA CAG ATG GAG CAG
301
I D L D S M K N Q N I N L E N S L G D V E A R Y K A Q M E Q
1033
CTC AAT GGG GTC CTT CTG CAT CTG GAG TCA GAG CTG GCA CAA ACT CGG GCA GAG GGC CAG CGC CAG GCC CAG GAA TAT GAA GCC CTC TTG
331
L N G V L L H L E S E L A Q T R A E G Q R Q A Q E Y E A L L
1123
AAC ATC AAG GTG AAG CTT GAG GCA GAG ATT GCC ACC TAC CGC CGC TTG CTG GAG GAT GGA GAA GAT TTC AGT CTC AAC GAT GCC CTG GAC
361
N I K V K L E A E I A T Y R R L L E D G E D F S L N D A L D
1213
TCC AGC AAC TCC ATG CAA ACT GTG CAG AAG ACA ACT ACC CGT AAG ATC GTG GAT GGC AGA GTG GTG TCC GAG ACT AAT GAC ACC AGA GTT
391
S S N S M Q T V Q K T T T R K I V D G R V V S E T N D T R V
1303
CTG AGG CAC TGA GGCAGAGAAGGAGGGAACCCCTGGGAACTGAGGGACCAATAAAAGTTGAGAGCTCACTGG
421
L R H

Fig.2

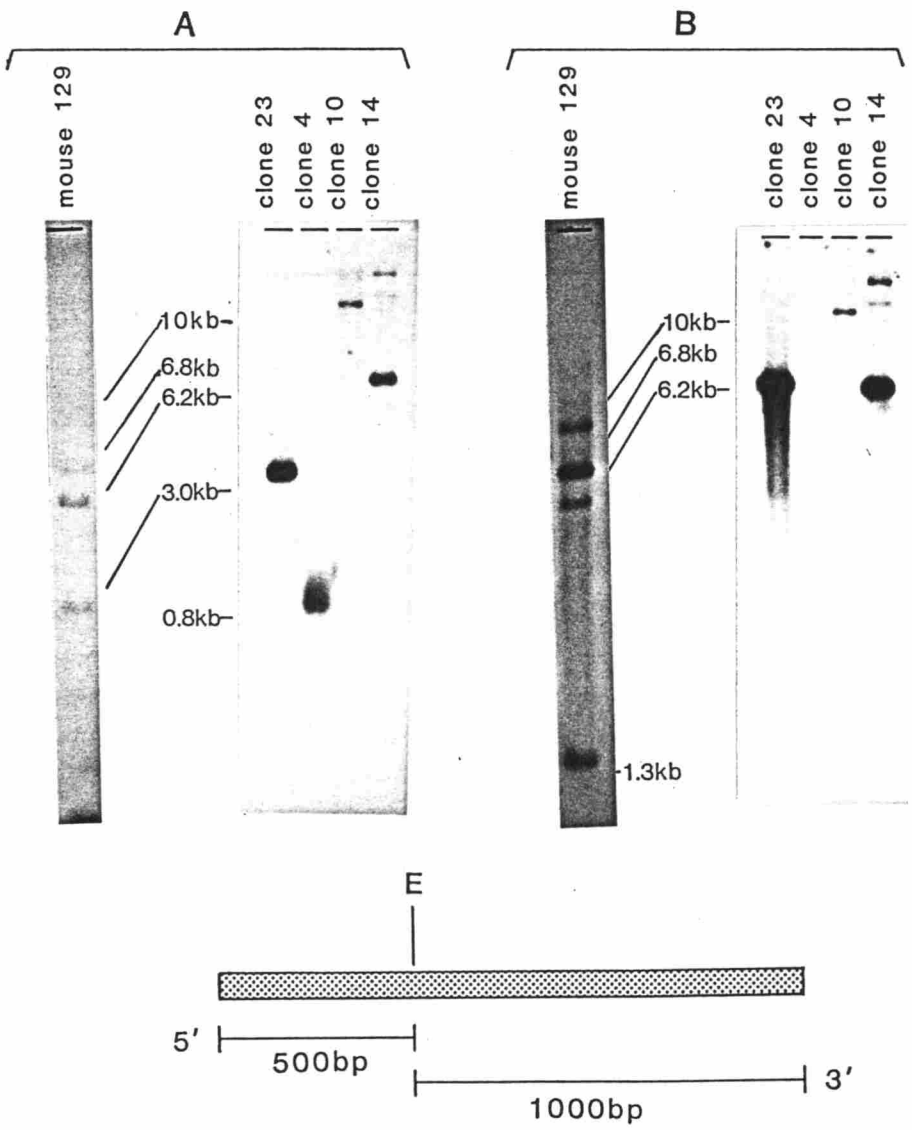

A
B
mouse 129
clone 23
clone 4
clone 10
clone 14
10kb
6.8kb
6.2kb
3.0kb
0.8kb
1.3kb
E
5'
500bp
1000bp
3'
Probe
A
B

Fig.3

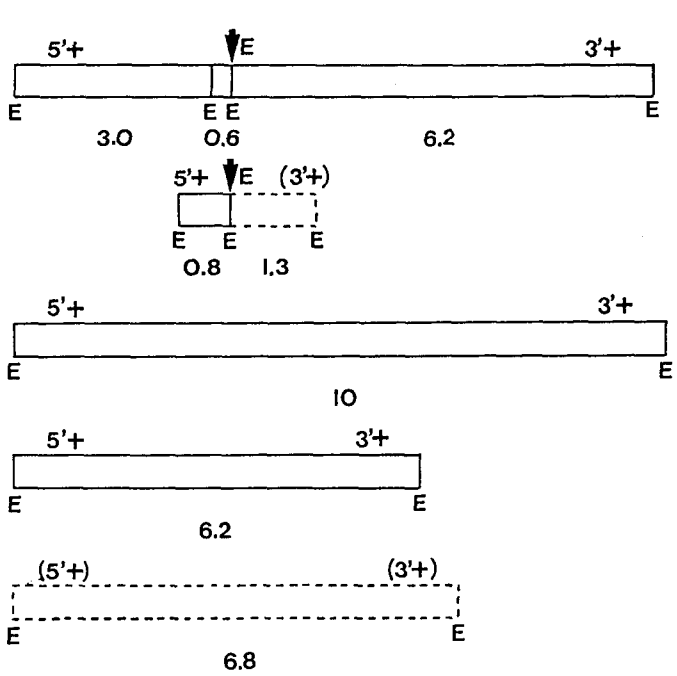

clone No.
B1
B2
23
5'+
E
3'+
E
E E
E
3.0
0.6
6.2
+
−
4
5'+
E
(3'+)
E
E
E
0.8
1.3
+
−
10
5'+
3'+
E
E
10
+
−
14
5'+
3'+
E
E
6.2
−
−
−
(5'+)
(3'+)
E
E
6.8

Fig.4

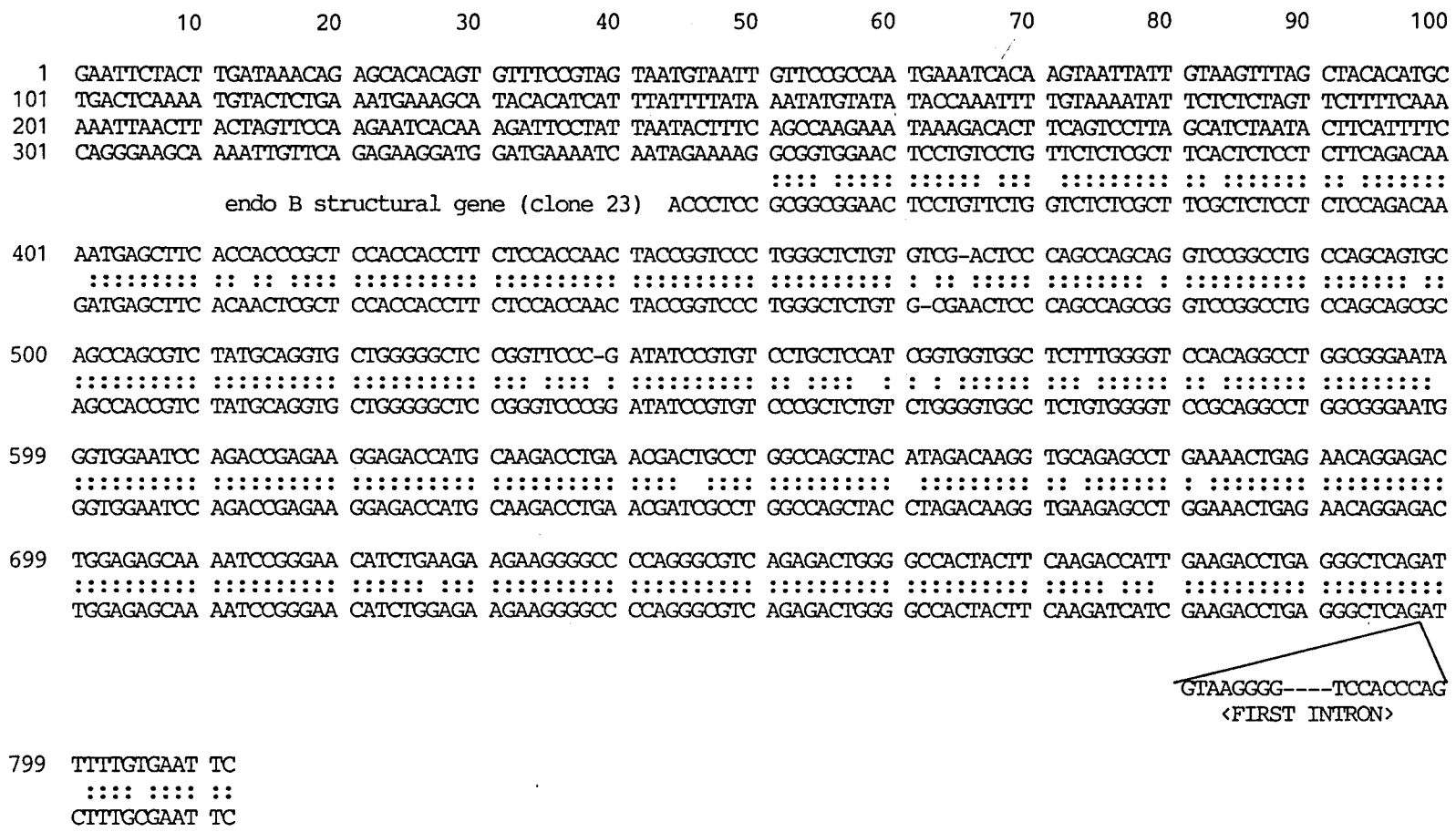

10 20 30 40 50 60 70 80 90 100
1 GAATTCTACT TGATAAACAG AGCACACAGT GTTTCCGTAG TAATGTAATT GTTCCGCCAA TGAAATCACA AGTAATTATT GTAAGTTTAG CTACACATGC
101 TGACTCAAAA TGTACTCTGA AATGAAAGCA TACACATCAT TTATTTTATA AATATGTATA TACCAAATTT TGTAAAATAT TCTCTCTAGT TCTTTTCAAA
201 AAATTAACTT ACTAGTTCCA AGAATCACAA AGATTCCTAT TAATACTTTC AGCCAAGAAA TAAAGACACT TCAGTCCTTA GCATCTAATA CTTCATTTTC
301 CAGGGAAGCA AAATTGTTCA GAGAAGGATG GATGAAAATC AATAGAAAAG GCGGTGGAAC TCCTGTCCTG TTCTCTCGCT TCACTCTCCT CTTCAGACAA
endo B structural gene (clone 23) ACCCTCC GCGGCGGAAC TCCTGTTCTG GTCTCTCGCT TCGCTCTCCT CTCCAGACAA
401 AATGAGCTTC ACCACCCGCT CCACCACCTT CTCCACCAAC TACCGGTCCC TGGGCTCTGT GTCG-ACTCC CAGCCAGCAG GTCCGGCCTG CCAGCAGTGC
GATGAGCTTC ACAACTCGCT CCACCACCTT CTCCACCAAC TACCGGTCCC TGGGCTCTGT G-CGAACTCC CAGCCAGCGG GTCCGGCCTG CCAGCAGCGC
500 AGCCAGCGTC TATGCAGGTG CTGGGGGCTC CGGTTCCC-G ATATCCGTGT CCTGCTCCAT CGGTGGTGGC TCTTTGGGGT CCACAGGCCT GGCGGGAATA
AGCCACCGTC TATGCAGGTG CTGGGGGCTC CGGGTCCCGG ATATCCGTGT CCCGCTCTGT CTGGGGTGGC TCTGTGGGGT CCGCAGGCCT GGCGGGAATG
599 GGTGGAATCC AGACCGAGAA GGAGACCATG CAAGACCTGA ACGACTGCCT GGCCAGCTAC ATAGACAAGG TGCAGAGCCT GAAAACTGAG AACAGGAGAC
GGTGGAATCC AGACCGAGAA GGAGACCATG CAAGACCTGA ACGATCGCCT GGCCAGCTAC CTAGACAAGG TGAAGAGCCT GGAAACTGAG AACAGGAGAC
699 TGGAGAGCAA AATCCGGGAA CATCTGAAGA AGAAGGGGCC CCAGGGCGTC AGAGACTGGG GCCACTACTT CAAGACCATT GAAGACCTGA GGGCTCAGAT
TGGAGAGCAA AATCCGGGAA CATCTGGAGA AGAAGGGGCC CCAGGGCGTC AGAGACTGGG GCCACTACTT CAAGATCATC GAAGACCTGA GGGCTCAGAT
GTAAGGGG----TCCACCCAG
<FIRST INTRON>
799 TTTTGTGAAT TC
CTTTGCGAAT TC

Fig.5

```
                    10        20        30        40        50        60        70        80        90       100       110       120

   1 GAATTCTAGGGTCTTCTAGTCATTACTGAAAGAACTGAAAGTCCACCAGTGGTTGGGGGTGACTTGAAGCTTCCTTCCCGTCCCCTCCTGACGGAGGCCTCACTCTGCAATAAGTAACTC
 121 CTATTTCTTCAGCAGTGAGTGCTGTATTAACATTCCCGTGCACTGCCTCCTAGAAGATAGACAAACGATTCTTTTTTTTTTTTCTTCTACCCATAGAAGTGAAAAACAAAACGGAAGTTC
 241 AGTTTAGTGACTTTCTGAAGTCACCCAACCAGCAAAGTATCCAAGTTGGATAACAAACGTAAAGCTTAAAAACAGCGAAAAGTTGCCCATCGGTACACTTCATTGTATAGTTCTATAAAA
 361 TAGAGCAATTGTGGAACTCGGGTGGGGGGGGTGGGGGGGGGAGAGTTTGTAAAATTGAAAGCCTGGGATCTCCAAACACACCCCCTCCCCAGGAGCTTTTAGGCAAAGCCAGTGGGCATT
 481 TGTTTGTATGTTATTTTGCATTCCATTTGCCTGGCTGTCTCTGGCAACACCCAGGCTCCCCAGCCAAGTCCCACGGAGTCGCTGAGGTTTTCCGGTCCAGAAGGAGGCAGGCTTCGAACA
 601 ATCCATCACACCCAGCATGCAGTCAGCCCTGTGTTTGCTTCCCTCTAGTCTGATCTGCTGACCTAGGTTCTTCGTAGCCCCTCCCCCTCCCCCTTGCTCTTCTGATTTTACTTTTCTGTG
 721 TGCTCGCACTCACACAGTCAGTTCAGCTGGCCGCTCGCTCGATAAGGATCAATAGGTCCACTCCAGAGCAAAGAGCATGGCCACAGGTATAGCAAACACCAAGTCCTGCCGGAAAAACGG
 841 GCTCCTGGTGCTGACCGATTCCAACCTCAGGGTTTCTCCCTCCTCATTCAGCAGGAGTGGTCATCCAGCTGGCGAGGAGGGGGTGCCGGGGGTGCAGAGGGGCTCTTTCCCCCAGAAAACA
 961 AGCAGGAAAGCCAGAAAGCATTATAAAAATTGACCAGGAAGGACCGAAGAAAAGGCCCGGGGAGCAGGAGAAGGGGACAGATGCAAGGGGGTGGCAGAGGTTGAGGTGGGGGATGGAGAG
1081 GCAGGCAATCTGGTTCAAACTTAGCTGTTTCTACATATCTGGTCTATGGACTCAGACCCCAACCTGTCCAGACACAGTATAATCTACAGGAAGGACAGTGGTTCTCTCTTCCTTGCTGTT
1201 CCCATAGCTACTAGAAGGTGTCTAGCAAATACTGGTTAAATCTATGTGCTCTTGGGTTGGCTGTGTTCTCAGCTGGTAAAGTGGTCTGCCTGAGGAGTATGAGGCCTTGGGTTTGGTCCT
1321 CAGCACCCTGTAACCTGAAAGTGATGGAGTCTCATGCCTGGGATCCCAGCACTTCAAAAGTGGGACGGGAAGATCAGGAGTTGAAGGCCATCCTCAGCTACATGGAAAGTTCAAGACCAG
1441 CCCGAGCTACATGAGACCCTGTCTCAAAAAAGAAGGAGGAGGAGGAGGAGGAGGAGGAGGTAAATGGTAGGTGTCCAGACATAGGGATACCTCTTCTACATAATCACAAAGGCCCCGATA
1561 AAGCCTCCAGGGAATGAGTTGAATGAGGACATGTGTACCCACGCGTTTGAGGATATGTATTTCCACGGTTCCCTATGGAGCCACAGCAGTGTCTTGGAACCATGGGGCCGGAGCAGAAAA
1681 ACCAAGAGACCTGCCCTCAGGTTCGTGTTTCTCTTCTTTTTGAGCTCCCAATTTTCTTTCCCCAGGATTCCTGGGGTTCTAGGACTGCCCCCAGACTTAAGCTGGCGACATGCCACAAAA
1801 TCTTCACCCTAGTGAACCCCTCTTCTGTCTGTGTAAATGGAGTTTGTTTGTTTGTTTGTTTGTTTGTTTCTCGAGATAAGGTTTCTCTGTGTAGCCCTGGCTGTCCTGGAACTCACT
1921 CTGTAGCCCAGGCTGGCCTCGAACTCAGAAATTCGCCTGCCTCTGCCTCCCAAGTGCTGGGATAAAGGCGTGCGCCACCACGCCCGGAATAAACGGATATTTAAAACGCCATTTTTTGCC
2041 CTGAAAACAGACACTCAGGGGCTAACACCACCTGCTATCCTTTGTGCACCCCTTAGTTATCCACTCCCTGTGTCGGTGAGGATGGCAGGTATGCTGAGGGCGTGTGCAGAAGTGGAGAAG
2161 GGCTGCGACCTGTTAGGTCACAGGAATCCCGATCCTGGCGCCCCCGGCTCCGGTGGCTGGGGCGTGGCCTGCTGGGGAGGTCCCTACCTCCTCCCGCGGCGATATAACAACAGGTCGAG
2281 GACTGCCACCCTCCGCGGCGGAACTCCTGTTCTGGTCTCTCGCTTGGCTCTCCTCTCCAGACAAGATGAGCTTCACAACTCGCTCCACCACCTTCTCCACCAACTACCGGTCCCTGGGCT
   1                                                                                                  M  S  F  T  T  R  S  T  T  F  S  T  N  Y  R  S  L  G  S
2401 CTGTGCGAACTCCCAGCCAGCGGGTCCGGCCTGCCAGCAGCGCAGCCAGCGTCTATGCAGGTGCTGGGGGCTCCGGGTCCCGGATATCCGTGTCCCGCTCTGTCTGGGGTGGCTCTGTGG
  20  V  R  T  P  S  Q  R  V  R  P  A  S  S  A  A  S  V  Y  A  G  A  G  G  S  G  S  R  I  S  V  S  R  S  V  W  G  G  S  V  G
2521 GGTCCGCAGGCCTGGCGGGAATGGGTGGAATCCAGACCGAGAAGGAGACCATGCAAGACCTGAACGATCGCCTGGCCAGCTACCTAGACAAGGTGAAGAGCCTGGAAACTGAGAACAGGA
  60  S  A  G  L  A  G  M  G  G  I  Q  T  E  K  E  T  M  Q  D  L  N  D  R  L  A  S  Y  L  D  K  V  K  S  L  E  T  E  N  R  R
2641 GACTGGAGAGCAAAATCCGGGAACATCTGGAGAAGAAGGGGCCCCAGGGCGTCAGAGACTGGGGCCACTACTTCAAGATCATCGAAGACCTGAGGGCTCAGGTAAGGGGCCTAAGGGGGG
 100  L  E  S  K  I  R  E  H  L  E  K  K  G  P  Q  G  V  R  D  W  G  H  Y  F  K  I  I  E  D  L  R  A  Q  └→INTRON 1
2761 AGGGGCAGTAGCTCCGGGCTGTTCTGTCTATCCAGTGTGCCCACACCCCCAAACTTCAATAGAGTGTGGGCTGTTGACTCCCTCCACTCTTTCCACATTCACCACATCCCCCCACACACA
2881 CTCTATGGGACGGAGTTAGGAATACTCTGGACTCTCACCCACAACCTGTAGGAATTCCAAGCGTAGTGGTTACACACGCGAGCCGGCTTCCCCCAAGCAGTGTAGTGGGAAGAGAACAGG
3001 ATTGGAAAAAAAATAAAATAAAAAGGTGTGTGTGGGGGAATCATAAATGAACTCAGACACTCACTTCTTTTCCAGTAGGAACAGTTAGAGGGTTAAGCGGATGTGGCTAAGGCTGAGTC
3121 ATCTAGGAGTAAACAGGAGCTTACCTGTAGGAGGGGCCAGCGGGACAAGTGGGGGGGGGGCGCGTCCCAGGACGGGGGGGGGGGGGGGGGTTAGCAGGTGCACCTGGAGAAGATGCC
3241 AGGAGAGGATCAGAGAAATCTTGTTGGAAGCTGCTCTTTTGTAAGCAATCCAGATAGTGCAGGCCCTGGCAGAGGGGCAAGGAATCCAGGAAGGGAGAGGAGGCTCAGGTGGGAAGGGAA
3361 GACAGAAATACACGCCTCTGACTGTTCCTGGGACTGGGATGGATTCACTGAAACACAAGACGTGTTTTCTCATCCCCCACTTCTCCACACACACACACACACACACACACACACACACAC
3481 CCTCATGTGACCCGGAACTCTCAAAGCGACTCTCTCCACTTTAATCCACCCAGATCTTTGCGAATTCTGTGGACAATGCCCGCATCGTCTTGCAGATCGACAATGCCCGCCTTGCCGCCG
 133                                                  INTRON 1←┘I  F  A  N  S  V  D  N  A  R  I  V  L  Q  I  D  N  A  R  L  A  A  D
3601 ATGACTTTAGAGTCAAGTAAGTCCGGGGCTGCAGCTGGAGCTAAGAATGGCCTACTCCCAGAACTGTATTAGAGTAGGGTTCCTTCCTTTCCGAAGTTTATTCAGGTTAGGGTTCATTGA
 156  D  F  R  V  K └→INTRON 2
3721 CTACTGAGTTAGCTCTTCATCTGTGCAGCCCAGTTTAAGCAAACGCTCCAAGGCCAAGAATCGTGCTAAGGGAATGACTCCGTGACTGTTTATGGTGGGCACGGGGCTAGTCTATGGAGG
3841 TGGCACTGGTTGAGAAGGTTCAGGCCCTCTGATCACCTCCACTCCTTCAGGTATGAGACAGAACTAGCCATGCGCCAGTCTGTGGAGAGCGACATCCATGGACTCCGCAAGGTGGTAGAT
 161                                                  INTRON 2←┘Y  E  T  E  L  A  M  R  Q  S  V  E  S  D  I  H  G  L  R  K  V  V  D
3961 GACACCAACATCACAAGGCTGCAGCTGGAGACAGAAATCGAGGCACTCAAGGAAGAACTTCTGTTCATGAAGAAGAATCATGAAGAGGTAAGCCGGGCCACTGACCAGGCCTAGGAGCTA
 184 D  T  N  I  T  R  L  Q  L  E  T  E  I  E  A  L  K  E  E  L  L  F  M  K  K  N  H  E  E └→INTRON 3
4081 AGGGCCAAGACTAGGCGGGCCAGGCTGCCCCCAAAGCAGCCTAGCATGGAAGCAGAAGCCCCGGCTGTTAGCCTGTTTTGATTAAGAGCAGAGCCATTGGCTCACACATGTTCTATGGCT
4201 CACCTAAGTTACACTATTGAGAGTCAAGATAACATTAAGCCAGTTAAGGCCTGGCTTGTGCATCACTACCTCCTGGAAAGCTGAGACAGGAGGATTGAGTTCAATGCCATCCTGTCTTAG
4321 TCTAGTAAAACAGAGATTGTGTGAGTAAGACCCTGAGTTGGAGGCAAAGGTGGGGACTGCCAGTCTTAGGAGCTGGTGCTCATGTGGGCCTTCTGGAAGAAGCTTAAGGTAAGATCTGTG
4441 GAAGGAAAAAGCAGCTACAACGAGGAGAGGATATCACCAGCGAGTGGCAGGTTCAGGCTGTCAAGGAGTGGCATTCCCAACCAGCTGGAGCAGAAGCCCAGTCTTAGGCTTGCCAAGGGT
4561 TTCCTCTGCTCTCTTCTGCCTTTACACTCAACATCACCTTCTAGAAAAGGCCATCAGTAATGAAAGGCTGGGCTGAACTGACTTTAAGTGTCACCTGCCACCCCCCTTTATTTCTGCCTT
                                                                                                                          INTRON 3
4681 CTTTAGGAAGTCCAAGGTCTGGAAGCCCAGATTGCCAGCTCTGGATTGACTGTGGAAGTGGATGCCCCCAAATCTCAGGACCTCAGCAAGATCATGGCGGACATCCGCGCCCAGTATGAA
 213 ←┘E  V  Q  G  L  E  A  Q  I  A  S  S  G  L  T  V  E  V  D  A  P  K  S  Q  D  L  S  K  I  M  A  D  I  R  A  Q  Y  E
4801 GCGCTGGCTCAGAAGAACCGCGAGGAACTGGACAAGTACTGGTCTCAGCAGGTACAAACTAGGAAAGGCTAGCTGCCCGCCCCCACTCCGCTTAAGGTAAGCTCTGTGTGAAGGGGAGGT
 251 A  L  A  Q  K  N  R  E  E  L  D  K  Y  W  S  Q  Q └→INTRON 4
4921 GCCAAGGACAGCAGGTCATGACTGACCCTCCTCACCCTTCAGATTGAGGAGAGTACCACAGTTGTCACCACCAAGTCTGCCGAAATCAGGGACGCTGAGACCACACTCACGGAGCTGAGA
 268                                  INTRON 4←┘I  E  E  S  T  T  V  V  T  T  K  S  A  E  I  R  D  A  E  T  T  L  T  E  L  R
5041 CGCACCCTCCAGACCTTGGAGATTGACTTGGACTCCATGAAAAACCAGGTGAGCATCCTCCAGTCACCTGCTGCAGTTCTGCTAGGCCCAGACCTCAGACACTTGGTGGCCCCAGCCCCA
 294 R  T  L  Q  T  L  E  I  D  L  D  S  M  K  N  Q └→INTRON 5
5161 ACTGTCTCACCCCTAAGCCTGCTCCACACCCCCATTCCCAATCAAGCGATTCCCTAAGATGAATCACTTCTATTTGATTTTTTTTTCCATTGCTCCCACACTGTAGCCTAGCATTTGATA
5281 GACAAGAGGCATCAAGACCCTTTCAAAGAAGGGTTGAGCAGGCTAAGAGTAGAGACTCACTCTCCCATCTCTGTGTCCTGCAGAACATCAACTTGGAGAACAGCCTCGGGGATGTGGAGG
 310                                                                                  INTRON 5←┘N  I  N  L  E  N  S  L  G  D  V  E  A
5401 CCCGATACAAGGCACAGATGGAGCAGCTCAATGGGGTCCTTCTGCATCTGGAGTCAGAGCTGGCACAAACTCGGGCAGAGGGCCAGCGCCAGGCCCAGGAATATGAAGCCCTCTTGAACA
 323  R  Y  K  A  Q  M  E  Q  L  N  G  V  L  L  H  L  E  S  E  L  A  Q  T  R  A  E  G  Q  R  Q  A  Q  E  Y  E  A  L  L  N  I
5521 TCAAGGTGAAGCTTGAGGCAGAGATTGCCACCTACCGCCGCTTGCTGGAGGATGGAGAAGATTTCAGGTGAGTGGTACCCTGCTAACCACACAAGTGAGGGGCTTTTCTCCACTCCCGGA
 363  K  V  K  L  E  A  E  I  A  T  Y  R  R  L  L  E  D  G  E  D  F  S └→INTRON 6
5641 GCTTTGGGAAGTGCCAAGGTCTGTTTGTTGGTAACCACTCACCAGTGGCCTCAAGTGAGTGAGTTATGGGCTCCCTCCCGTGAGAAAAGAATGATTACATATTATCCCTGGCTGTCCATG
5761 GGATGGCACCGAGTGCCTCTTTTCTGGAAGAAATGGCTGAGGAGACTTTGGGGGGGTAAGGAAGGGCCTCAGGGAACCCTTCTTCCACATACTCTGTCTCCCCCTCTATTCTGGGCTGTTT
5881 ATAACTAAGGCTTGGTCATCTGTTACAGTCTCAACGATGCCCTGGACTCCAGCAACTCCATGCAAACTGTGCAGAAGACAACTACCCGTAAGATCGTGGATGGCAGAGTGGTGTCCGAGA
 385                  INTRON 6←┘L  N  D  A  L  D  S  S  N  S  M  Q  T  V  Q  K  T  T  T  R  K  I  V  D  G  R  V  V  S  E  T
6001 CTAATGACACCAGAGTTCTGAGGCACTGAGGCAGAGAAGGAGGGAACCCCTGGGAACTGAGGGACCAATAAAAGTTGAGAGCTCACTGGACATCACTTTGTGTCTTCCTTGGCAGTCTTC
 416  N  D  T  R  V  L  R  H
6121 CTGCTTGCGCACACAAACATTAGGCAGGAGTCTCTTGACCCTAGAGTTTCACACGACTGGCCTGTGGGAGAGGAAAAGTCTGGGTTCTTGAGACATCTAGGAGACGAAGTAGGAGATATG
6241 ACTGTCCTAGGTGGAGGTGAAAGTGTAATAAAATTAGAACCTCCCATTCAAGAGACCTGGTGAAGAAGCCAGTGGTGGTGATGTACACACACATGGTTTAGGTTGAAGCAGAAGGATCGTGA
6361 GTTCAAGGCTGGTCTGGGCTAGATAGCAAGTTCCAGGGCAACCTGCATCACATGCAAATATCCTGCTAAAACAAGAGCTGGGTATCGGTTGTGCGCTTCTAAAATCTCTGAGCTTGGTAG
6481 TGGCGAGAGGATCAAGAATTTGAGGTGATCATCTGCTACATAGCAAATCTGAAACCAGCCTGGGCTACATGAGACCTTGTCTCATAAGGGAGTCTAGGAAGGGACTTGGGGAAGCCAAGC
6601 ATGTTGACACAGGCCTATAATCCCAGCACCTGGGAAAGAAACAGAAGCAGGAACATCTTGAGTTCAAAGCCAGTCTACAGCACCCTCTCTGAAGAAAAAAAAAGAAAGAGAGAAAAGAGA
6721 AAAAAAGAAAAAATGTGTAATTCTCCAGCAGAGTCCAGAGAAAACTCGCCCCCTGTTCTACTACTTTTTCCTTTGCCTAAATCTTAGTTTTTCCCACCCAAAGATTCTCAAACCCCATCC
6841 CACACTCACTACCCAGAAACCTGTGCTCCCCCTGGGCTCTGCTCTGACACTCAGTGAAGACACACTCTTCAGAGGGGTGTGGCCAGAAGTGAGGAGACATGATTAGAGAACATGAAGCAAT
6961 GTACATGGTGGATCTCAGGTAGGCACCAGGGTGGGCACCATGCAGGTAAAAATAGGAGCCCCATGGCACCATTCTGGGCTTTGTTTTTGGAGCTATAGGACAAAGCAATGGAAACAACACA
7081 TGTCTGAAATAACATGACCAGGCATGGTGGCATACACCTGCAACAGGAGGCAGAGGCAGGCGGATCTCTGTGAGTTTGAAACCAGCCTGGTCTACATAGCCAGTTCATGACAGTCAATGC
7201 TACATGGAGAGGTCCTGTCTTTACTACTAATAGTAGTAAATAATATGATTTAAAATAAAAATAACAATCTGTTTCAATTGGGGGGATAATGTATCTTTATTAGTTTCAAGAATTATAGAA
7321 GTTAATATATTGTCCAAAAGGAAAGAGGGAAATAAAATTGAAAACAAATCTCTGTTCCCCCAAAGTACTTTATCCCAGTTGAATACTTCGTTAGGACAGGCATATACAAATATTTTATTT
7441 TACTTCATAGATGTAGGGTTATACTATAAGTTCTGCACCTTAATTGTTTCAATTCTACATCCTTCACTTAACAATGTATCTTGAAATGGGGGATGTGGTGGCTCAGGTCTGTAATCCCAG
7561 CCCTCTGTAGGCTAAAGTGGGAGGTAGGATTACTCAGAGTTCAGGACCAGTGTGGTTACAAAGTAAAGCCTTCACGTCACTCATTTGAACAGCTGCTTGCAGCTATAGAATGGAGTACAA
7681 AGTTACTTAACCCAAATACTCTAGTTCCTAGAGATGTCTTGGTCTCTAGCAGTTTTAATTTAGCAAAAGTCTGCAGCTGATATCTATGTGTACTTTTGCACTCCATGAGCATGTCTGTGG
7801 AGTAATTTAGTTAAAAAATCGTACCATTTAAACACACATGTGAAGGACTTACTTTAGGCCAGGCCCTGTGCTAAGCTT
```

Fig.6

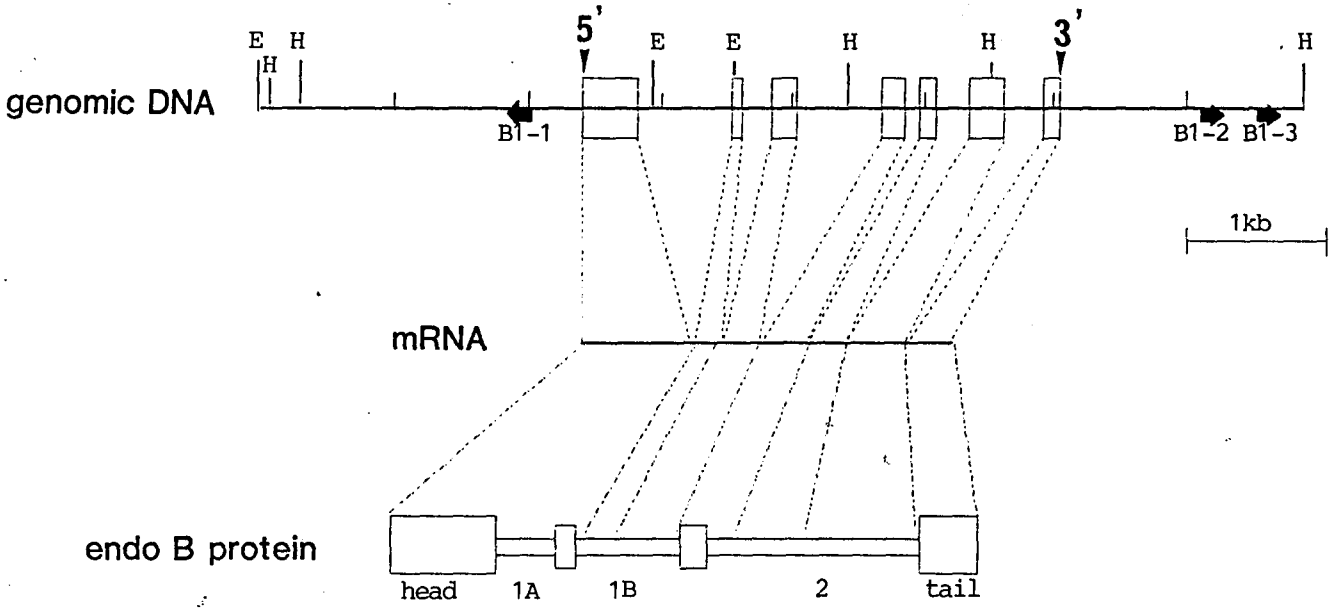

genomic DNA
E H
H
5'
E
E
H
H
3'
H
B1-1
B1-2
B1-3
1kb
mRNA
endo B protein
head
1A
1B
2
tail

Fig.7

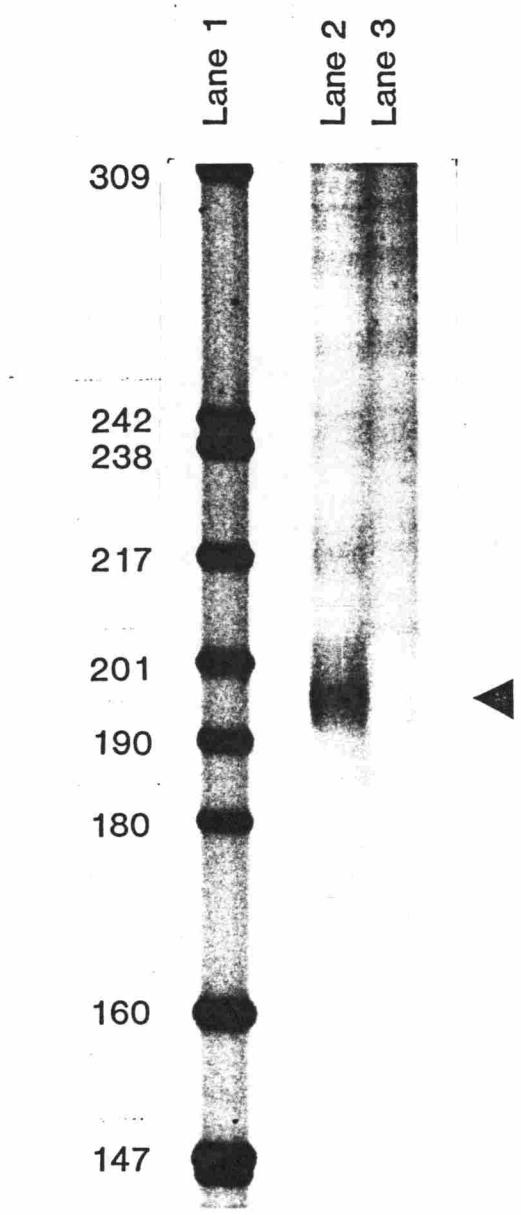


Fig.8

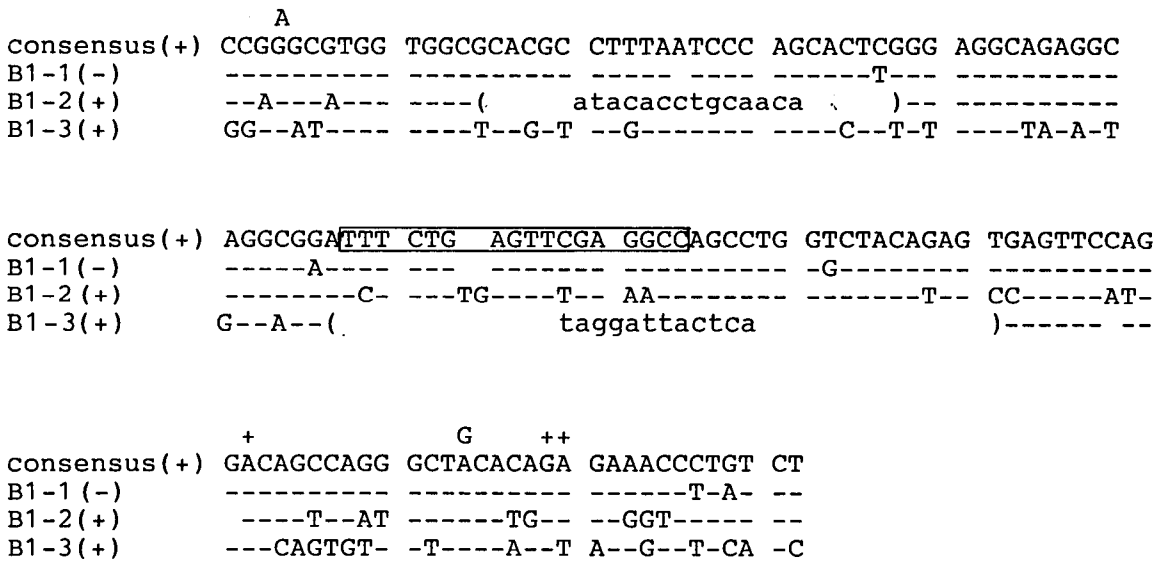

A
consensus(+) CCGGGCGTGG TGGCGCACGC CTTTAATCCC AGCACTCGGG AGGCAGAGGC
B1-1(-) ---------- ---------- ----- ---- ------T--- ----------
B1-2(+) --A---A--- ----(atacacctgcaaca)-- ----------
B1-3(+) GG--AT---- ----T--G-T --G------- ----C--T-T ----TA-A-T
consensus(+) AGGCGGATTT CTG AGTTCGA GGCCAGCCTG GTCTACAGAG TGAGTTCCAG
B1-1(-) -----A---- --- ------- ---------- -G-------- ----------
B1-2(+) --------C- ---TG----T-- AA-------- -------T-- CC-----AT-
B1-3(+) G--A--(taggattactca)------ --
+ G ++
consensus(+) GACAGCCAGG GCTACACAGA GAAACCCTGT CT
B1-1(-) ---------- ---------- ------T-A- --
B1-2(+) ----T--AT ------TG-- --GGT----- --
B1-3(+) ---CAGTGT- -T----A--T A--G--T-CA -C

Fig.9

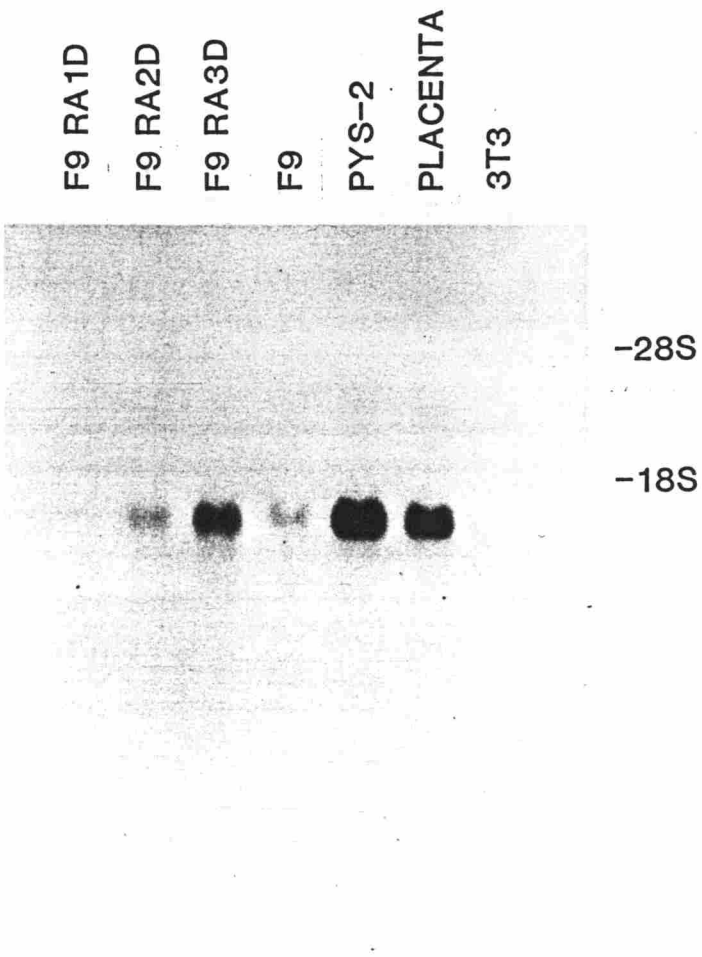


Fig.10

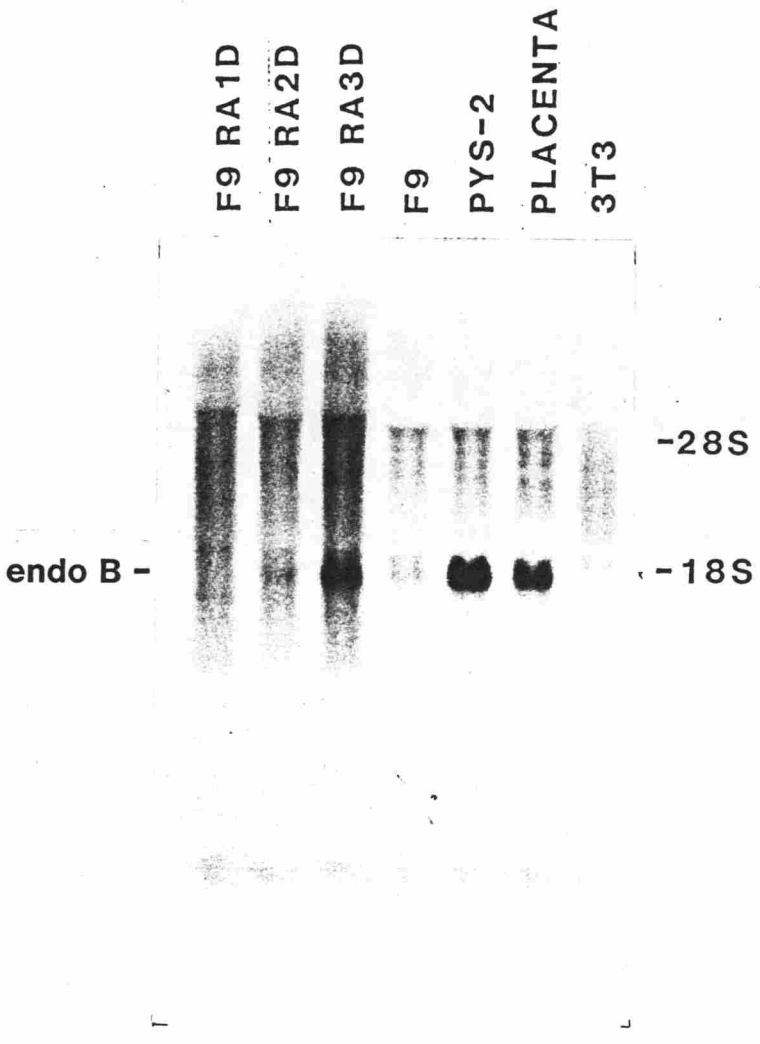


Fig.11

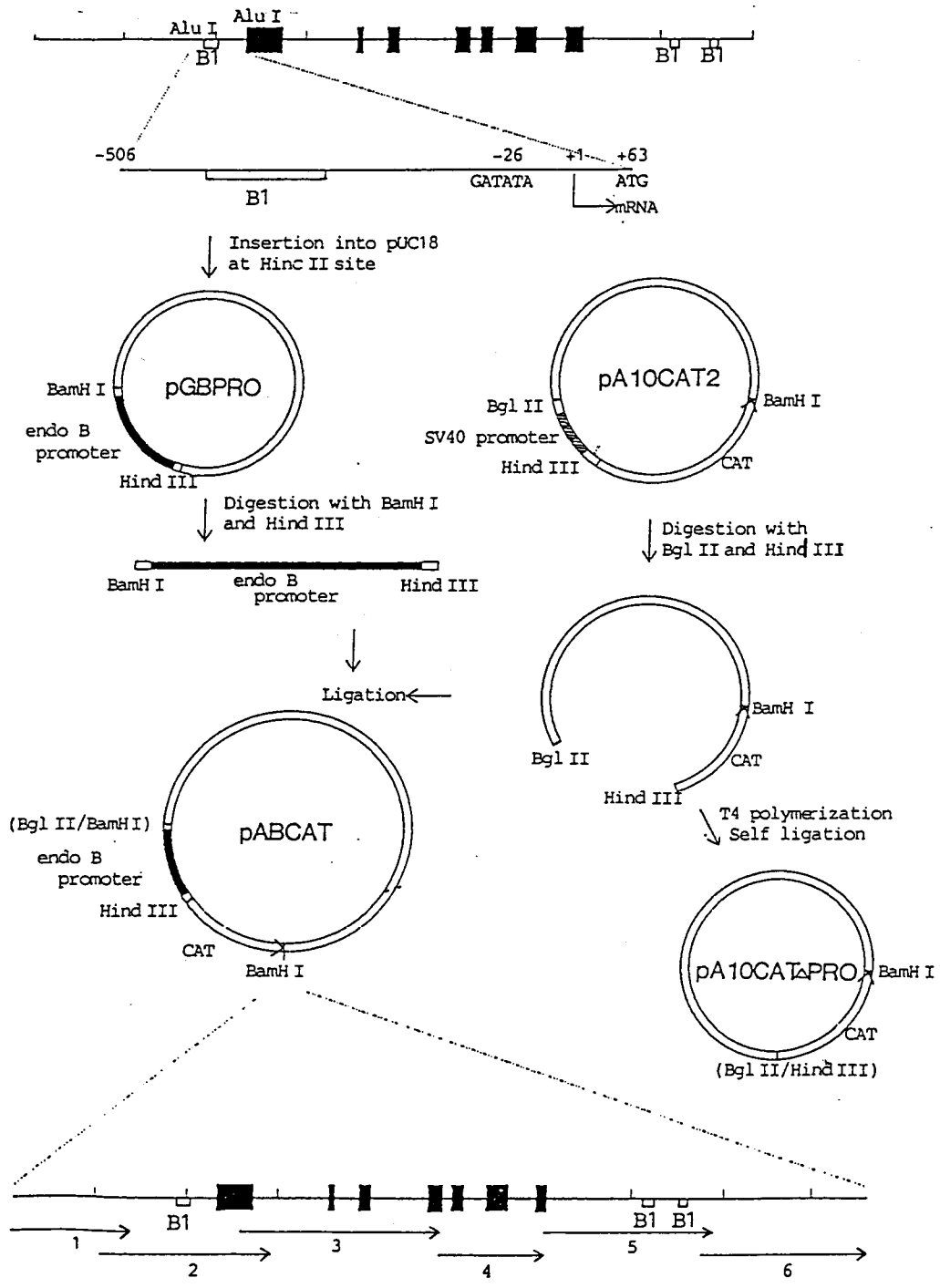
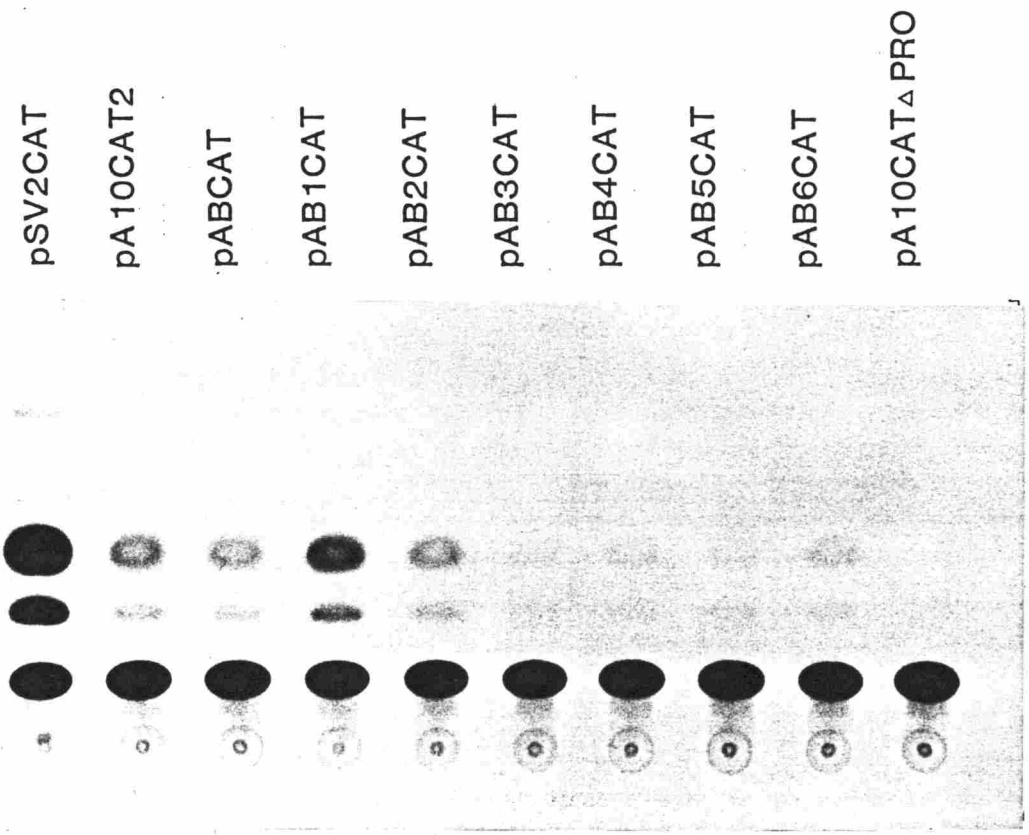

Alu I
Alu I
B1
B1 B1
-506
-26
+1
+63
GATATA
ATG
B1
mRNA
Insertion into pUC18
at Hinc II site
pGBPRO
BamH I
endo B
promoter
Hind III
pA10CAT2
Bgl II
BamH I
SV40 promoter
Hind III
CAT
Digestion with BamH I
and Hind III
Digestion with
Bgl II and Hind III
BamH I
endo B
promoter
Hind III
Ligation
BamH I
Bgl II
CAT
Hind III
T4 polymerization
Self ligation
(Bgl II/BamH I)
pABCAT
endo B
promoter
Hind III
CAT
BamH I
pA10CATΔPRO
BamH I
CAT
(Bgl II/Hind III)
B1
B1 B1
1
2
3
4
5
6

Fig.12

PYS-2


pSV2CAT
pA10CAT2
pABCAT
pAB1CAT
pAB2CAT
pAB3CAT
pAB4CAT
pAB5CAT
pAB6CAT
pA10CATΔPRO

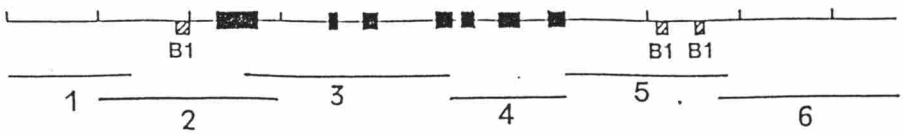

B1
B1
B1
1
2
3
4
5
6

Fig.13

PLACENTA

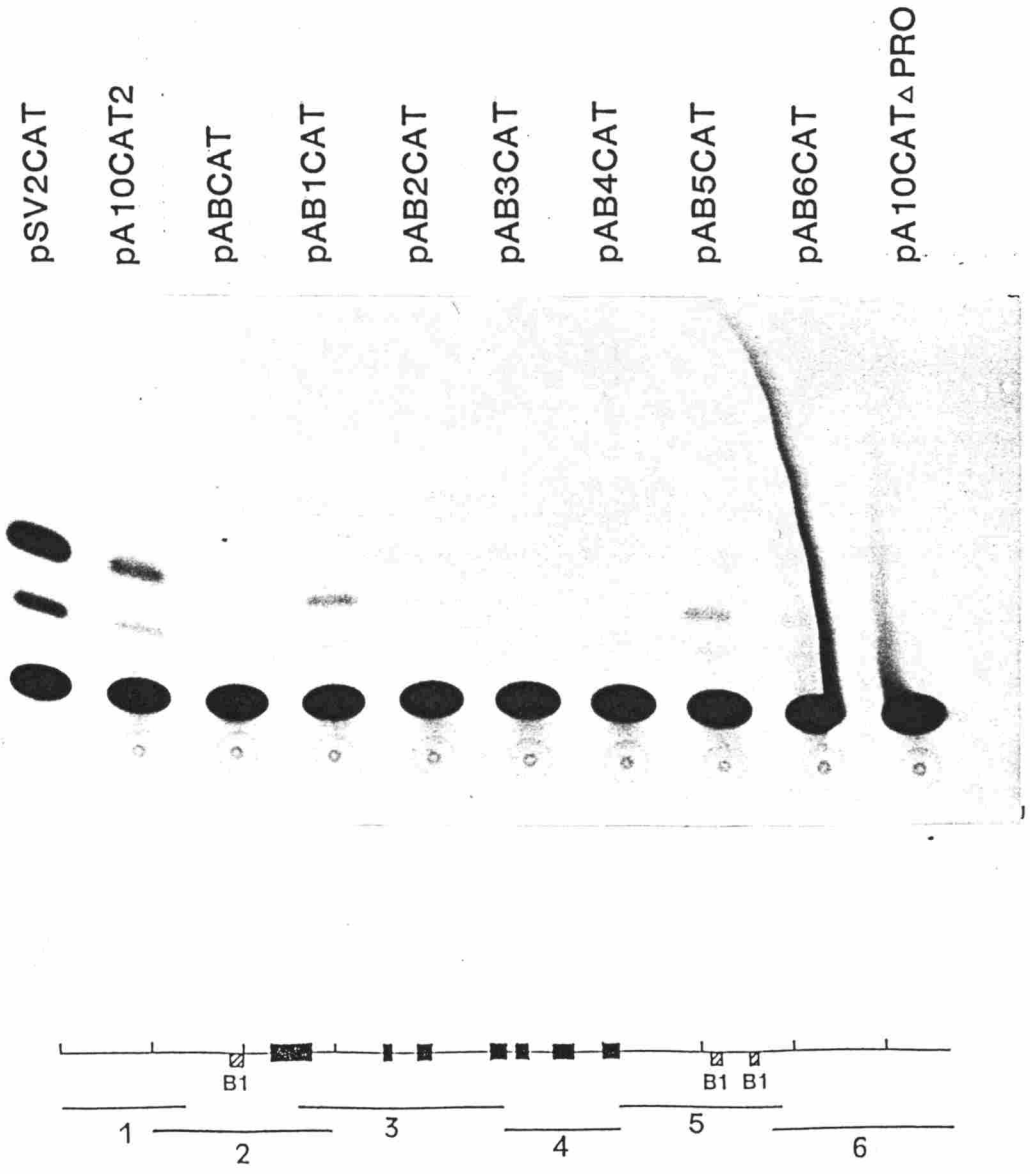


Fig.14

F9

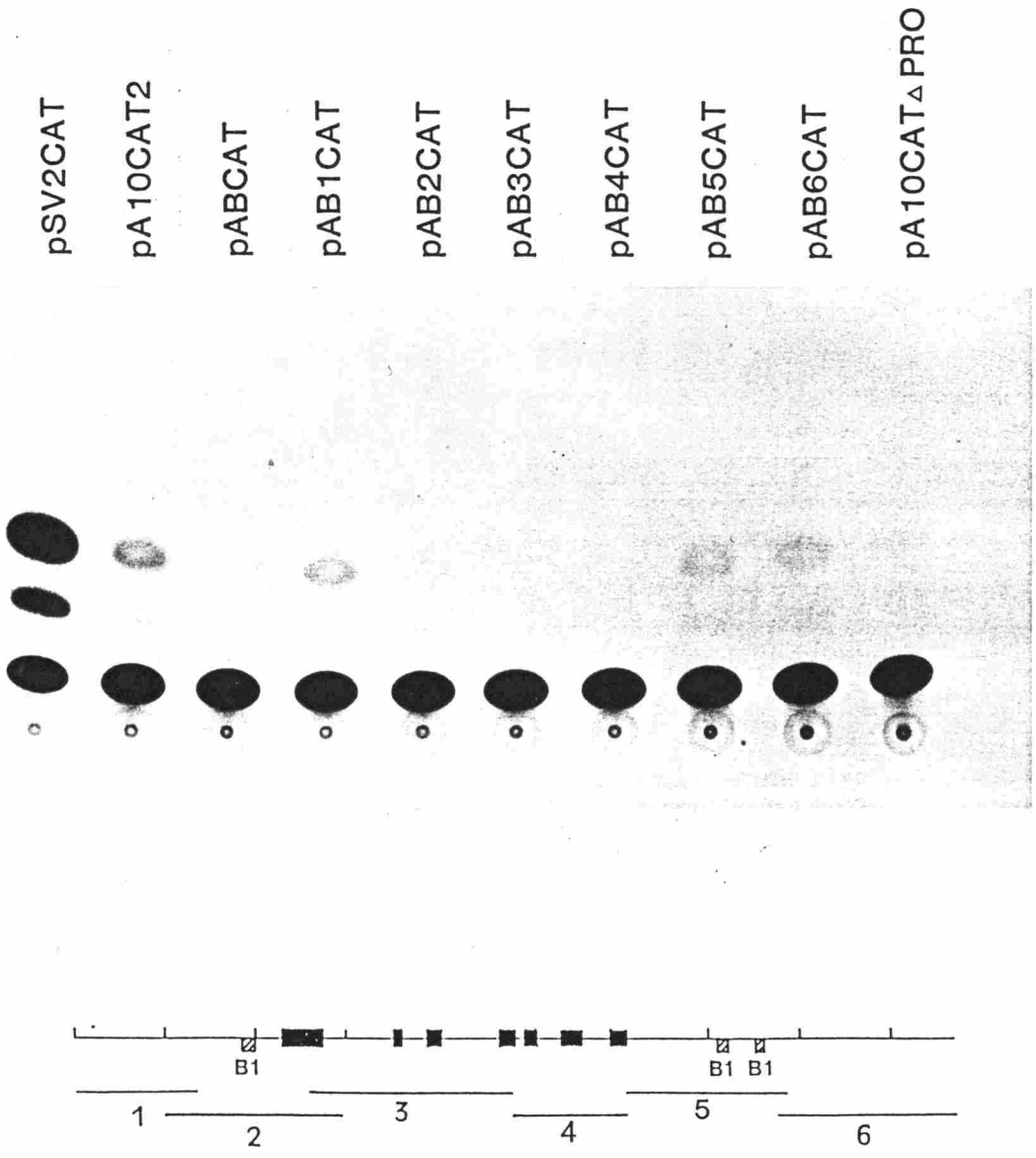

pSV2CAT
pA10CAT2
pABCAT
pAB1CAT
pAB2CAT
pAB3CAT
pAB4CAT
pAB5CAT
pAB6CAT
pA10CATΔPRO
B1
B1
B1
1
2
3
4
5
6

Fig.15

3T3

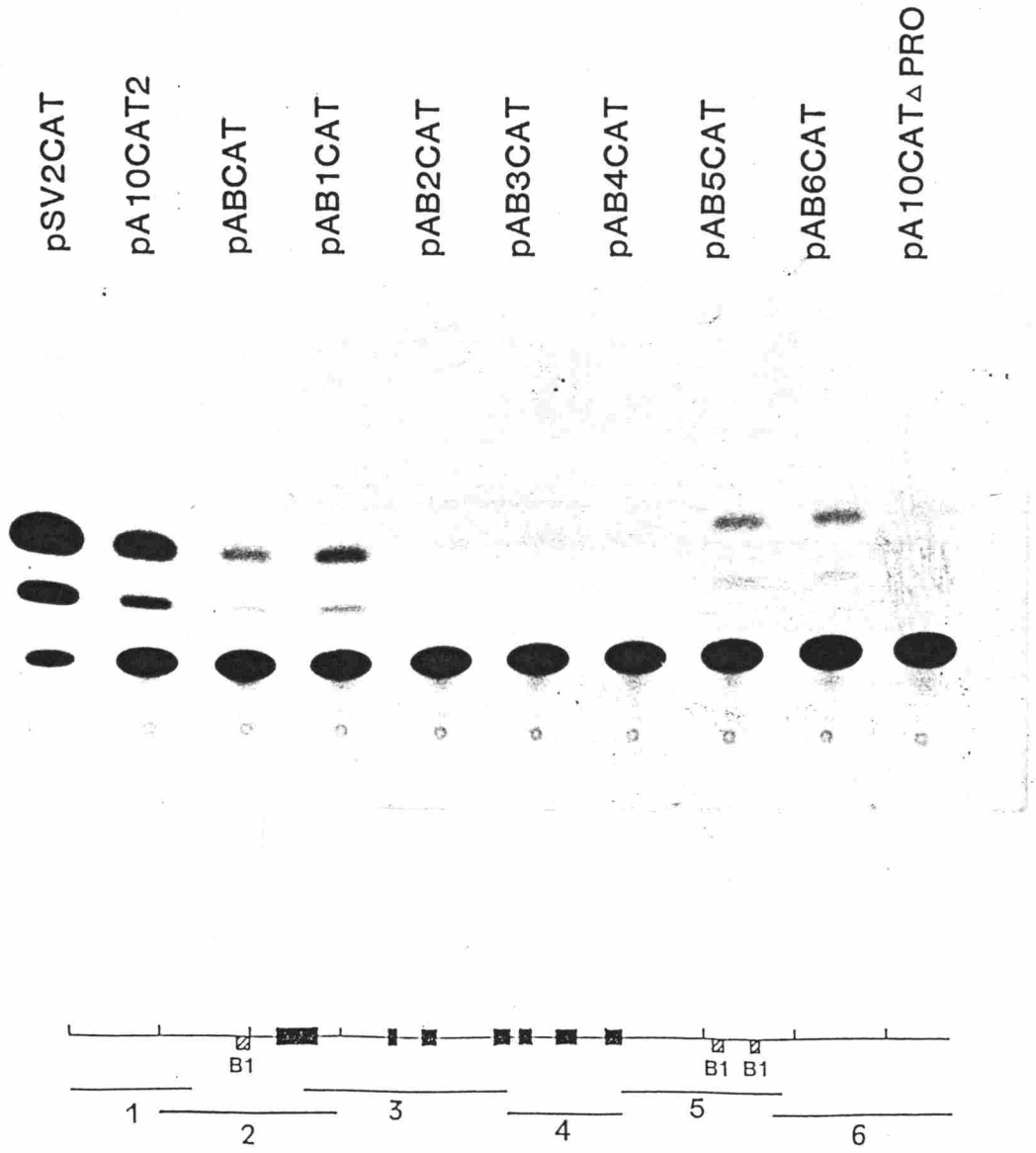

pSV2CAT
pA10CAT2
pABCAT
pAB1CAT
pAB2CAT
pAB3CAT
pAB4CAT
pAB5CAT
pAB6CAT
pA10CATΔPRO
B1
B1 B1
1
2
3
4
5
6

Fig.16

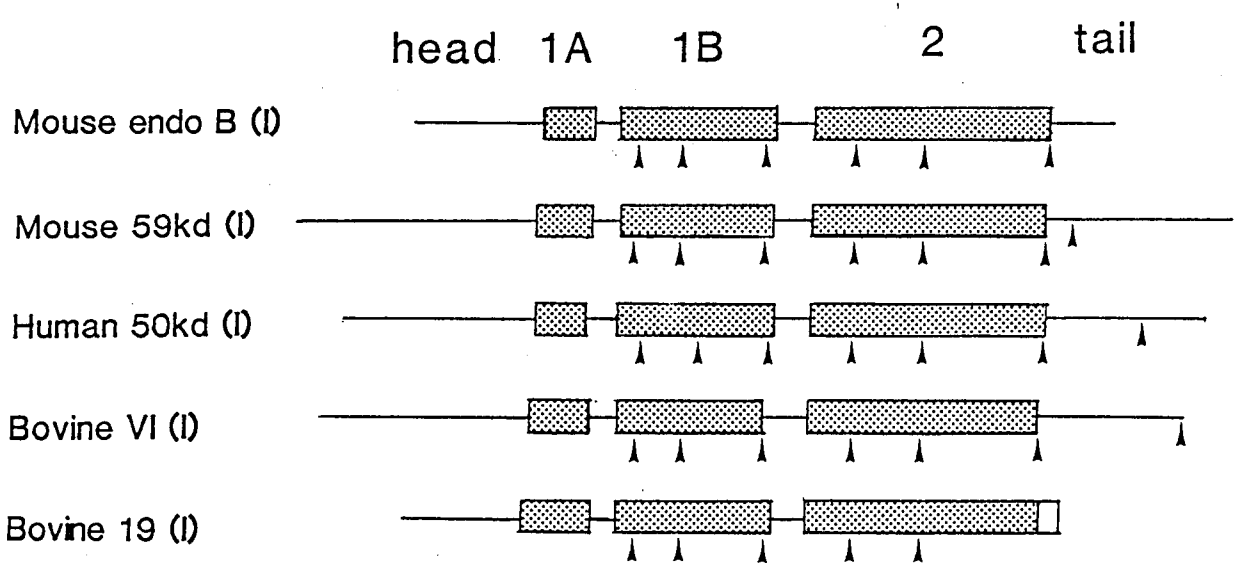

head
1A
1B
2
tail
Mouse endo B (I)
Mouse 59kd (I)
Human 50kd (I)
Bovine VI (I)
Bovine 19 (I)

Fig.17

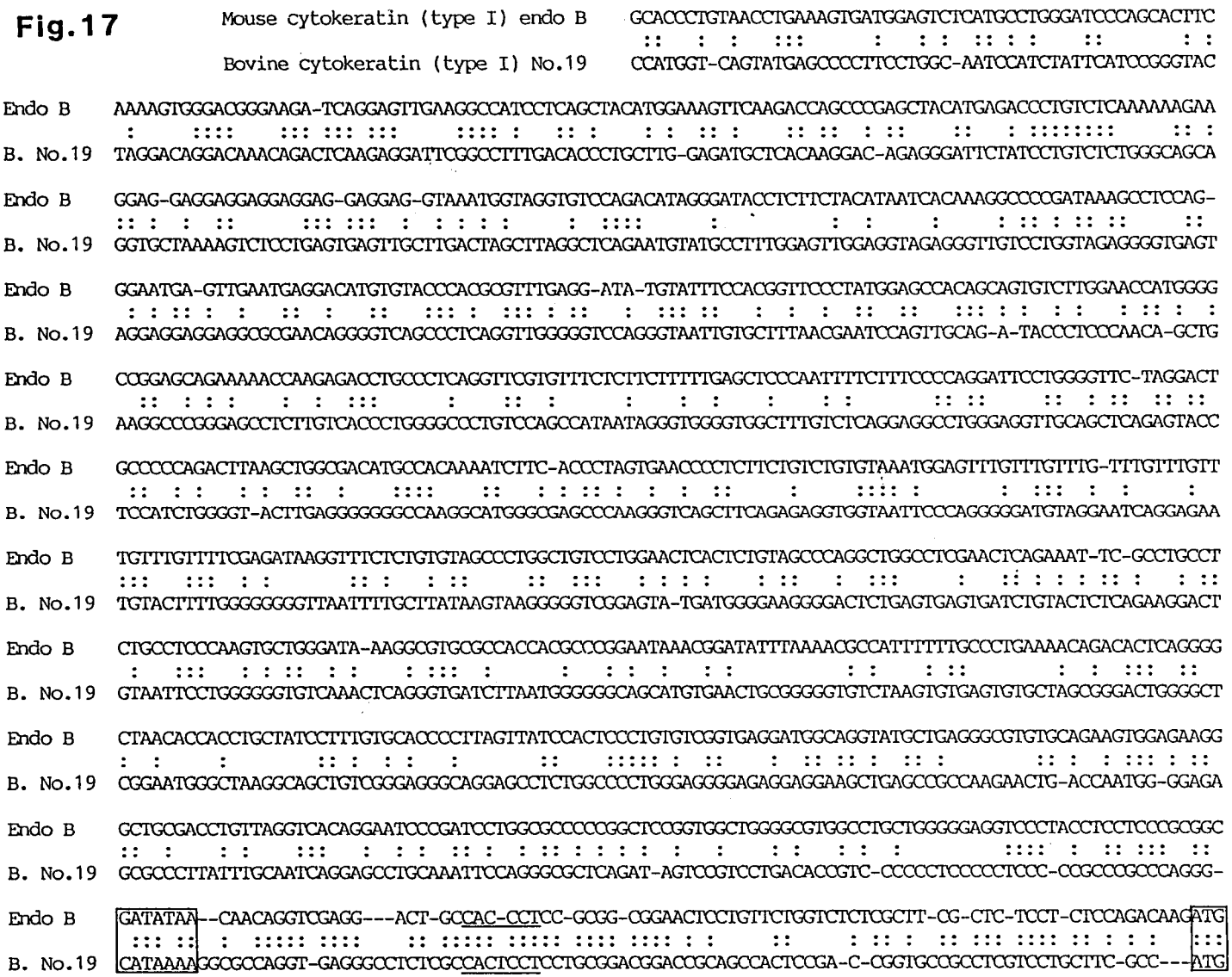

Mouse cytokeratin (type I) endo B GCACCCTGTAACCTGAAAGTGATGGAGTCTCATGCCTGGGATCCCAGCACTTC
Bovine cytokeratin (type I) No.19 CCATGGT-CAGTATGAGCCCCTTCCTGGC-AATCCATCTATTCATCCGGGTAC
Endo B AAAAGTGGGACGGGAAGA-TCAGGAGTTGAAGGCCATCCTCAGCTACATGGAAAGTTCAAGACCAGCCCGAGCTACATGAGACCCTGTCTCAAAAAAGAA
B. No.19 TAGGACAGGACAAACAGACTCAAGAGGATTCGGCCTTTGACACCCTGCTTG-GAGATGCTCACAAGGAC-AGAGGGATTCTATCCTGTCTCTGGGCAGCA
Endo B GGAG-GAGGAGGAGGAGGAG-GAGGAG-GTAAATGGTAGGTGTCCAGACATAGGGATACCTCTTCTACATAATCACAAAGGCCCCGATAAAGCCTCCAG-
B. No.19 GGTGCTAAAAGTCTCCTGAGTGAGTTGCTTGACTAGCTTAGGCTCAGAATGTATGCCTTTGGAGTTGGAGGTAGAGGGTTGTCCTGGTAGAGGGGTGAGT
Endo B GGAATGA-GTTGAATGAGGACATGTGTACCCACGCGTTTGAGG-ATA-TGTATTTCCACGGTTCCCTATGGAGCCACAGCAGTGTCTTGGAACCATGGGG
B. No.19 AGGAGGAGGAGGCGCGAACAGGGGTCAGCCCTCAGGTTGGGGGTCCAGGGTAATTGTGCTTTAACGAATCCAGTTGCAG-A-TACCCTCCCAACA-GCTG
Endo B CCGGAGCAGAAAAACCAAGAGACCTGCCCTCAGGTTCGTGTTTCTCTTCTTTTTGAGCTCCCAATTTTCTTTCCCCAGGATTCCTGGGGTTC-TAGGACT
B. No.19 AAGGCCCGGGAGCCTCTTGTCACCCTGGGGCCCTGTCCAGCCATAATAGGGTGGGGTGGCTTTGTCTCAGGAGGCCTGGGAGGTTGCAGCTCAGAGTACC
Endo B GCCCCCAGACTTAAGCTGGCGACATGCCACAAAATCTTC-ACCCTAGTGAACCCCTCTTCTGTCTGTGTAAATGGAGTTTGTTTGTTTG-TTTGTTTGTT
B. No.19 TCCATCTGGGGT-ACTTGAGGGGGGGCCAAGGCATGGGCGAGCCCAAGGGTCAGCTTCAGAGAGGTGGTAATTCCCAGGGGGATGTAGGAATCAGGAGAA
Endo B TGTTTGTTTTCGAGATAAGGTTTCTCTGTGTAGCCCTGGCTGTCCTGGAACTCACTCTGTAGCCCAGGCTGGCCTCGAACTCAGAAAT-TC-GCCTGCCT
B. No.19 TGTACTTTTGGGGGGGGTTAATTTTGCTTATAAGTAAGGGGGTCGGAGTA-TGATGGGGAAGGGGACTCTGAGTGAGTGATCTGTACTCTCAGAAGGACT
Endo B CTGCCTCCCAAGTGCTGGGATA-AAGGCGTGCGCCACCACGCCCGGAATAAACGGATATTTAAAACGCCATTTTTTGCCCTGAAAACAGACACTCAGGGG
B. No.19 GTAATTCCTGGGGGGGTGTCAAACTCAGGGTGATCTTAATGGGGGGCAGCATGTGAACTGCGGGGGTGTCTAAGTGTGAGTGTGCTAGCGGGACTGGGGCT
Endo B CTAACACCACCTGCTATCCTTTGTGCACCCCTTAGTTATCCACTCCCTGTGTCGGTGAGGATGGCAGGTATGCTGAGGGCGTGTGCAGAAGTGGAGAAGG
B. No.19 CGGAATGGGCTAAGGCAGCTGTCGGGAGGGCAGGAGCCTCTGGCCCCTGGGAGGGGAGAGGAGGAAGCTGAGCCGCCAAGAACTG-ACCAATGG-GGAGA
Endo B GCTGCGACCTGTTAGGTCACAGGAATCCCGATCCTGGCGCCCCCGGCTCCGGTGGCTGGGGCGTGGCCTGCTGGGGGAGGTCCCTACCTCCTCCCGCGGC
B. No.19 GCGCCCTTATTTGCAATCAGGAGCCTGCAAATTCCAGGGCGCTCAGAT-AGTCCGTCCTGACACCGTC-CCCCCTCCCCCTCCC-CCGCCCGCCCAGGG-
Endo B GATATAA--CAACAGGTCGAGG---ACT-GCCAC-CCTCC-GCGG-CGGAACTCCTGTTCTGGTCTCTCGCTT-CG-CTC-TCCT-CTCCAGACAAGATG
B. No.19 CATAAAAGGCGCCAGGT-GAGGGCCTCTCGCCACTCCTCCTGCGGACGGACCGCAGCCACTCCGA-C-CGGTGCCGCCTCGTCCTGCTTC-GCC---ATG

Fig.18

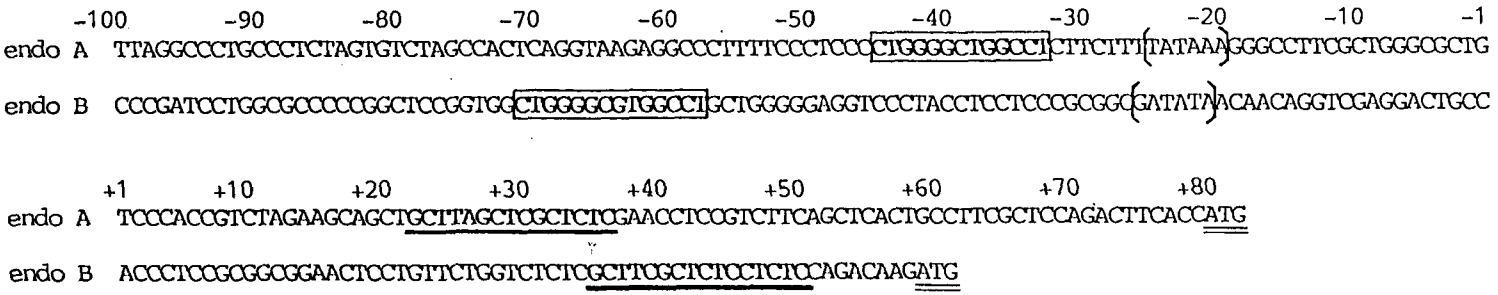

-100 -90 -80 -70 -60 -50 -40 -30 -20 -10 -1
endo A TTAGGCCCTGCCCTCTAGTGTCTAGCCACTCAGGTAAGAGGCCCTTTTCCCTCCCCTGGGGCTGGCCTCTTCTTTATAAAGGGCCTTCGCTGGGCGCTG
endo B CCCGATCCTGGCGCCCCCGGCTCCGGTGGCTGGGGGCGTGGCCTGCTGGGGGAGGTCCCTACCTCCTCCCGCGGCGATATAACAACAGGTCGAGGACTGCC
+1 +10 +20 +30 +40 +50 +60 +70 +80
endo A TCCCACCGTCTAGAAGCAGCTGCTTAGCTCGCTCTCGAACCTCCGTCTTCAGCTCACTGCCTTCGCTCCAGACTTCACCATG
endo B ACCCTCCGCGGCGGAACTCCTGTTCTGGTCTCTCGCTTCGCTCTCCTCTCCAGACAAGATG

Table 1. Sequences Homologous to the E1A Enhancer Core found in the 5' Upstream Region on the Endo B Gene

Adenovirus 5 E1A Enhancer(-200)	GGAAGTGAAA
Adenovirus 5 E1A Enhancer(-300)	GGAAGTGACA
Adenovirus 7 E1A Enhancer(-200)	GGAAGTGAAT
Polyoma Enhancer	GGAAGTGACT
Human HSP70 Promoter	AGAAGGGAAA
endo B (31-40)	AGAACTGAAA
endo B (215-224)	AGAAGTGAAA
endo B (6253-6262)	GGAGGTGAAA
endo B (7336-7345)	AAAAG-GAAA