Proc. Natl. Acad. Sci. USA
Vol. 82, pp. 1155–1159, February 1985
Developmental Biology

One gene and one pseudogene for the cytokeratin endo A

(trophectoderm/embryonal carcinoma cells/repetitive sequences)

MARC VASSEUR, PHILIPPE DUPREY, PHILIPPE BRÛLET, AND FRANÇOIS JACOB

Unité de Génétique Cellulaire du Collège de France et de l'Institut Pasteur, 25, rue du Dr. Roux, 75724 Paris Cedex 15, France

Contributed by François Jacob, September 27, 1984

ABSTRACT **The recombinant cDNA RecXVI, which hybridizes to the endo A mRNA, detects two specific bands on a Southern blot of *Eco*RI-digested genomic mouse DNA. We have screened a mouse genomic library with this cDNA and isolated these two sequences. Endo A is encoded by a 7.5-kilobase gene and by a 1.6-kilobase pseudogene devoid of introns. A repetitive sequence belonging to the B2 family is located in the third intron of the gene. We have observed that transcription of B2 sequences and of the endo A gene are mutually exclusive.**

The first detectable morphological differentiation of mouse embryonic cells takes place during blastocyst formation, in the course of which two different types of cells, the trophectoderm and the inner cell mass cells, become differentiated (1). This differentiation is accompanied by modifications in the pattern of protein synthesis, revealing a difference in the regulation of gene expression (2). Among these differences, some intermediate filament proteins are synthesized in the trophectoderm but not in inner cell mass cells (3, 4). Analysis of the onset of synthesis of these proteins might provide an insight into molecular events directly related to the cell commitment program.

We describe here the isolation and characterization of the genes encoding endo A (5), also referred to as cytokeratin A (6). Endo A appears during blastocyst formation and has been identified in trophoblast but not in inner cell mass cells (7). It is therefore a valuable marker of the modifications in gene expression during the first binary choice made by embryonic cells. Later in development, endo A is still expressed in a tissue-specific manner; it is expressed in visceral and parietal endoderm and in epithelial cells derived from mesoderm and endoderm (i.e., in liver and kidney) but not in fibroblasts, myoblast, neural tissues, or keratinocytes (8). *In vitro*, endo A is found in trophoblastoma cell line TDM-1, but not in embryonal carcinoma cells.

A cDNA clone, RecXVI, was isolated (9) from a cDNA library prepared from TDM-1. This cDNA hybridizes to a specific 18S mRNA encoding endo A. We have used this cDNA to isolate the endo A genes to analyze the regulation signals involved in the expression of this protein during the first embryonic differentiation. We describe here the structure of the genes and discuss the possible regulatory role of a repetitive sequence that is included in one of them.

MATERIALS AND METHODS

Cells. Teratocarcinoma cell lines F9 and PCC3 (10) were cultured under standard conditions. TDM-1 is a trophoblastoma cell line (10).

Isolation of Mouse Genomic Clones. A genomic library of BALB/c mouse DNA restricted with *Eco*RI and cloned into phage λ Charon 4A was plated to a density of 15,000 phages per 10-cm plate and screened as described (11) using the cDNA RecXVI.

Electron Microscopy of Heteroduplexes. Heteroduplexes were prepared by renaturing the DNAs in 50% formamide/0.58 M NaCl/50 mM Pipes, pH 6.8/1 mM EDTA at 64°C for 1 hr. After fixation with glyoxal for 2 hr at 12°C, a solution was prepared for electron microscopy by the formamide/cytochrome monolayer spreading procedure (12). Magnification was ×16,000 and micrographs of molecules were measured and compared by computer analysis. R loops were formed for 90 min under the same conditions except that the temperature was lowered to 25°C over a 3-hr period.

Nuclease S1 Mapping. RNA was extracted from the various strains (13) and poly(A)$^+$ fractions were selected on an oligo(dT)-cellulose column. The desired probe was cloned into an M13 mp8 vector and uniformly labeled to high specific activity with ^{32}P by primed DNA synthesis. The DNA was then restricted with *Eco*RI, which cleaves the polylinker of M13 mp8 but not the inserts. The labeled single-stranded probe was isolated by electrophoresis on 6% polyacrylamide gels and recovered by elution with 0.5 M NH_4OAc/10 mM $MgCl_2$/1 mM EDTA/0.1% $NaDodSO_4$ at 37°C. Hybridizations to the RNA were carried out according to ref. 14. After S1 nuclease digestion of the DNA·RNA hybrids, the protected fragments were electrophoresed on 8 M urea/6% polyacrylamide gels (15).

DNA Sequence. DNA sequences were determined by the dideoxy chain termination method of Sanger *et al.* (16).

RESULTS

Cloning of the Endo A Gene. On a Southern blot of *Eco*RI-digested mouse DNA, the cDNA RecXVI hybridized to specific DNA fragments of 2.3 and 2.5 kilobases (kb). We screened a mouse genomic library with the RecXVI probe and found six positive clones out of 4 × 10^5. The pattern of restriction showed that those clones fall into two categories. The first type (referred to as α1) contains the 2.5-kb *Eco*RI fragment, the second (referred to as α2) contains the 2.3-kb fragment.

To understand the relationships between these two types of clones, we performed cross-hybridization experiments between α1 and α2 DNA. We found that, whereas sequences contained in α1 hybridize only to the 2.3-kb band of α2, α2 sequences are able to hybridize to four bands of *Eco*RI-digested α1 DNA. Subcloning of the two types of recombinant confirmed that the sequences of the 2.3-kb fragment of α2 hybridize to sequences dispersed along 8 kb in α1. The mouse genome therefore seems to contain two endo A genes in which sequence organization is different.

The sequence organization of α1 and α2 was compared by electron microscopic analysis of heteroduplexed formed between the two clones. The α1 gene was excised from the λ vector by digestion with *Hin*dIII. This enzyme generated a 14-kb fragment that totally encompasses the sequences able

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviations: kb, kilobase(s); bp, base pair(s).

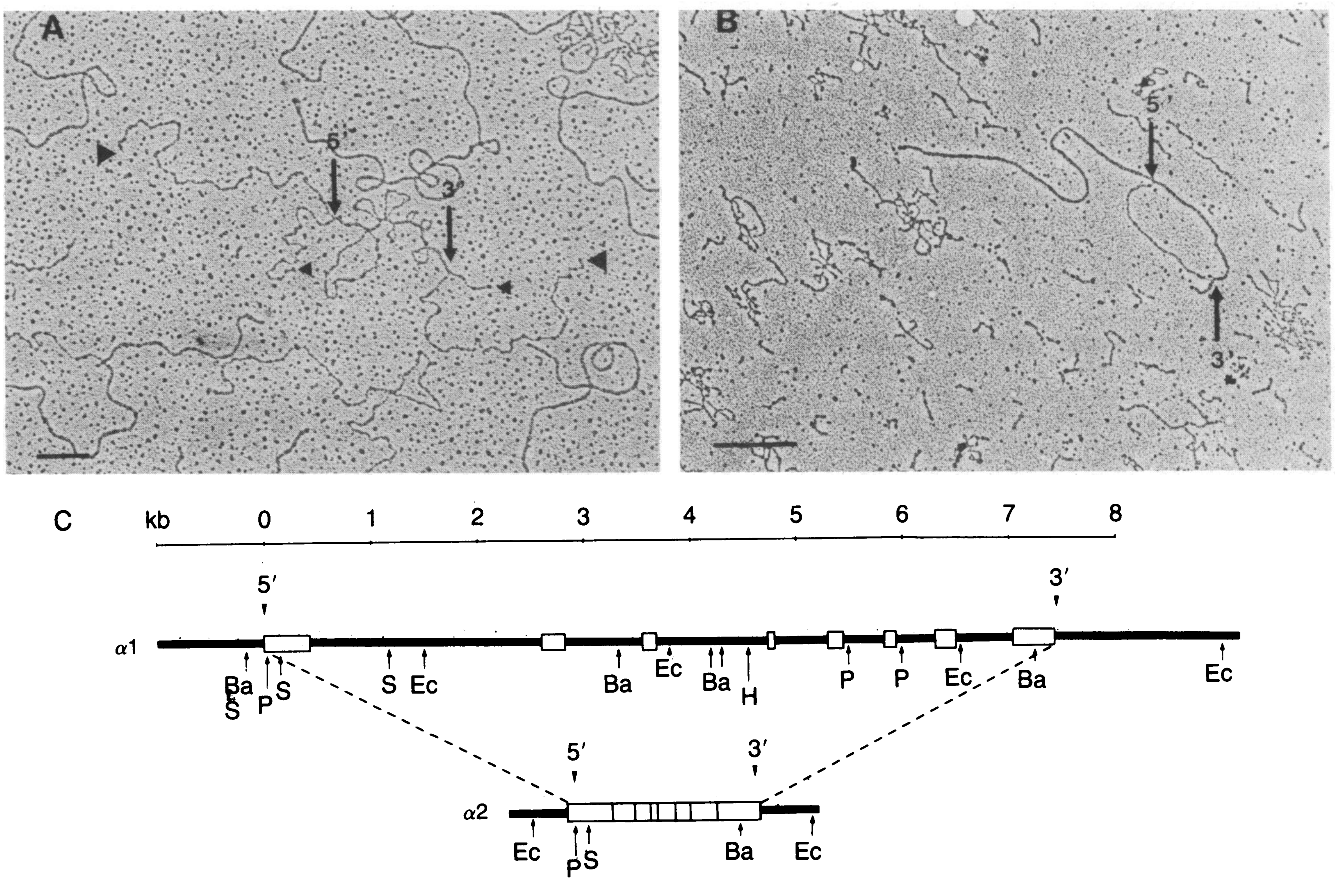

FIG. 1. Heteroduplex analysis of α1 and α2. (*A*) Electron micrograph of the heteroduplex formed between α1, excised from the λ genomic clone by *Hin*dIII, and the 2.3-kb band of α2, cloned in the *Eco*RI site of pBR322. To orient the molecule, the α2-containing plasmid was linearized with *Hin*dIII, which does not digest α2 and cuts pBR322 at a single point close to the *Eco*RI site. The long tail that does not hybridize with α1 is thus pBR322. Arrows indicate the ends of α1 (large triangles) and α2 (small triangles). (*B*) Electron micrograph of the heteroduplex formed between α2, prepared as described above, and poly(A)$^+$ RNA extracted from TDM-1 cells. (Bar = 1 kb.) (*C*) Schematic map of α1 and α2 genes, established from results of electron micrographs of DNA·DNA and DNA·RNA heteroduplexes, restriction nuclease analysis, and nuclease S1 mapping. Ba, *Bam*HI; S, *Sma* I; P, *Pvu* II; Ec, *Eco*RI; H, *Hin*dII.

to hybridize to α2. This fragment was hybridized to the 2.3-kb fragment of α2 subcloned in pBR322. A micrograph of a typical heteroduplex is shown in Fig. 1A. The hybridization of α1 to α2 was interrupted by seven loops of various sizes. The total length of the hybridized sequences was 1650 ± 120 base pairs (bp) as calculated from measurement of more than 30 molecules. The total length of the single-stranded loops was 5700 ± 400 bp. To find out whether all the single-stranded loops belong to the α1 gene or not, we measured the total length of α2 nonhybridized sequences plus the double-stranded region. The average length obtained without taking into account any loop was 6500 ± 320 bp, which is in good agreement with the calculated size of α2-containing pBR322 (6600 bp). It seems therefore that α2 does not contain any of the loops observed in the α1·α2 hybrid. This conclusion is supported by DNA·RNA hybridization experiments performed between α2 and poly(A)$^+$ RNA extracted from TDM-1 cells (Fig. 1*B*). Only one uninterrupted R loop was observed. Its total length was 1680 ± 110 bp, which is identical to the size of the double-stranded part of the α1·α2 heteroduplex molecule. Nuclease S1 mapping analysis confirmed all these results and particularly the size of the coding regions of α1 and α2, which were estimated, by this technique, to be 1640 ± 25 bp (data not shown). This size corresponds to the length of the endo A mRNA as measured by denaturing gel electrophoresis (see below).

The simplest interpretation of these results is to assume that the α1 gene encoding endo A is a total of 7.5 kb long and is composed of eight coding segments separated by seven intervening sequences. The other gene, α2, may be considered a pseudogene derived from a cDNA copy of endo A mRNA (map in Fig. 1). This copy seems to be full length and well-conserved, since restriction sites mapped in the α1 gene are localized in the corresponding regions of the α2 gene (Fig. 1). At the 5′ end, two sites, *Pvu* II and *Sma* I, are present in both genes, separated by the same segment of 115 bp. At the 3′ end, a *Bam*HI site is found not only in α1 and α2 but also at a similar position in the cDNA RecXVI.

Analysis of the 5′ End Region of α1. We analyzed the 5′ end of the α1 gene by nuclease S1 mapping and sequencing. S1 mapping was carried out by using the *Sma* I/*Sma* I 295-bp fragment (Figs. 1 and 2) localized at the 5′ end. This fragment was cloned into M13 mp8 to prepare single-stranded ^{32}P-labeled probes. The cap site of endo A mRNA extracted from TDM-1 cells was localized 20 bp upstream from the *Pvu* II site, with an accuracy of ±4 bp. As an internal standard control, the *Pvu* II/*Sma* I fragment was used as a single-stranded probe in the same conditions.

When the RNA used for mapping was extracted from F9 cells, no signal was observed, even with this sensitive technique. This supports previous observations of a transcriptional block of endo A expression in F9 cells (9).

We sequenced the *Sma* I/*Sma* I fragment (Fig. 3*A*). This region presents the typical features of the 5′ end of a eukaryotic gene (17). The cap site of endo A mRNA, as measured by nuclease S1 mapping, is 20–24 bp downstream from a "TATAA" box and 78–82 bp upstream from the

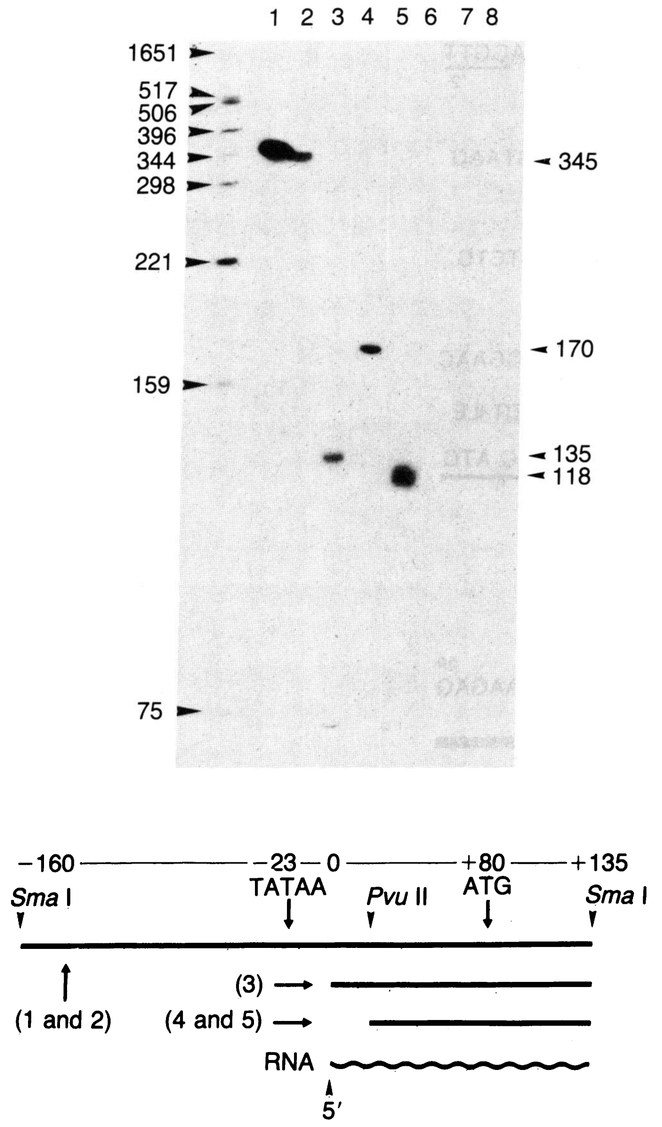

FIG. 2. Nuclease S1 mapping analysis. The structures of the undigested and digested probes are schematized below the autoradiograph. Numbers in parentheses indicate lanes in the gel. Lanes 1 and 2: *Sma* I/*Sma* I probe not hybridized with RNA and not treated with S1. Lane 3: *Sma* I/*Sma* I probe hybridized with 1 μg of poly(A)$^+$ RNA extracted from TDM-1 cells. The protected segment is 135 bp long. Lane 4: *Sma* I/*Pvu* II probe not hybridized and not digested. Lane 5: same probe hybridized with 1 μg of poly(A)$^+$ TDM-1 RNA and digested with S1. Lane 6: *Sma* I/*Sma* I probe hybridized with 2 μg of F9 RNA. Lane 7: *Pvu* I/*Sma* I probe hybridized with 10 μg of F9 RNA. Lane 8: probe *Sma* I/*Sma* I not hybridized, treated with S1. Differences between the theoretical lengths of the undigested probes and the sizes observed on gels are due to the M13 mp8 polylinker and primer sequences, which are 50 bp long. These sequences are removed during S1 digestion.

initiation codon ATG. The region localized upstream from the TATA box presents some peculiar structural features. A 12-bp palindroma is localized at −142 and another 12-bp palindroma, although potentially less stable, at −86. Two inverted 7-bp repeats are found in the −140 to −110 region and a 7-bp sequence is repeated in the same orientation at −90 and −60. The region containing these sequences is particularly (G + C) rich.

Presence of a Repetitive Element in the α1 Gene. When the total α1 gene was used as a probe on a Southern blot of mouse genomic DNA, the specific bands were masked by a dense smear indicating the presence of a repetitive element. This repetitive sequence was localized in the third intron of the gene, between the *Bam*HI and the *Eco*RI sites (Fig. 1). We sequenced this region and found that the repetitive element belongs to the B2 family (18) (Fig. 3*B*). Discrepancies between the sequence obtained and the B2 consensus sequence (18) are limited to 14 nucleotides out of 190, which is in the range of deviations observed among known B2 elements. All the B2 characteristics are observed, including direct repeats on each side, promoter and terminator signals for RNA polymerase III, and an (A + T)-rich region followed by an oligo(dA) stretch of 20 nucleotides. Direct repeats of 12 bp (at position 10 and position 214) and of 9 bp (at position 34 and position 228), respectively, are flanking the B2 element of α1. These two repeats share a 6-bp sequence. It is intriguing to note that a typical TATA box can be found at nucleotide 174, just downstream from a 20-bp sequence of purine-pyrimidine alternation. Such a structure may be used as a promoter element for RNA polymerase II (19).

We compared the relative levels of transcription of B2 elements and endo A in RNA from various types of cells and tissues. We probed RNA blots with two types of probes—(*i*) the α2 pseudogene, which contains all the coding sequences of endo A (Fig. 4*A*), and (*ii*) a subclone of α1 containing the B2 sequence and the fourth and fifth exons of endo A (Fig. 4*B*)—and found endo A to be expressed as a 1.6-kb RNA in TDM-1, liver, and kidney cells but not in F9, 3T3, or L cells. A small (0.6 kb) RNA hybridizing to B2 was detected at high levels in F9 and at lower levels in L cells. In F9 cells, some higher molecular weight RNA (1.6, 3.2, and 6 kb) also contained B2 elements. One of these transcripts (the 3.2-kb transcript) was also found in TDM-1 cells (Fig. 4*B*). All these results were confirmed by hybridization of the same types of RNA, carried out separately with either B2 or endo A purified probes.

We observed that expression of the 1.6-kb endo A mRNA seemed to exclude expression of the B2 0.6-kb RNA. In all cell lines tested, we found that either B2 or endo A was transcribed, or neither was transcribed (i.e., in 3T3), but we never found coexpression of both B2 small RNA and endo A in the same cell type.

DISCUSSION

We have isolated and characterized the genes encoding endo A to set up an experimental model for studying gene regulation both in embryonal carcinoma cells and during the first differentiation of the mouse embryo. We found that endo A is encoded by a 7.5-kb gene, α1, that contains seven introns and by a 1.65-kb pseudogene, α2, devoid of introns. The definition of α2 as a pseudogene is supported by several types of evidence: (*i*) hybridization of α1 and α2 gives a typical pattern of intervening sequences in the α1 gene; (*ii*) α2 is devoid of introns; (*iii*) α2 presents a restriction pattern identical to those of α1 in the coding region but different in the flanking sequences, at both the 3′ and the 5′ ends; and (*iv*) nuclease S1 mapping of the 5′ end of endo A mRNA indicates only one capping site (i.e., only one type of transcript), compatible with the 5′ sequence of α1. It is thus very unlikely that α2 is transcribed. α2 is probably a full-length reverse copy of a processed endo A mRNA transcribed from α1. Such mechanisms of reverse copying are now a well-established phenomenon and different types of intronless nontranscribed pseudogenes have been described, including p53 (20), Ki-*ras* (21), and the immunoglobulins (22).

We have analyzed the 5′ end of the α1 gene to define the capping site of the mRNA. The sequences localized at the 5′ end of a gene are generally assumed to be responsible for the regulation of gene expression (17). The 5′ flanking sequence of α1 contains the consensus signals for RNA polymerase II initiation and, upstream, (G + C)-rich symmetrical struc-

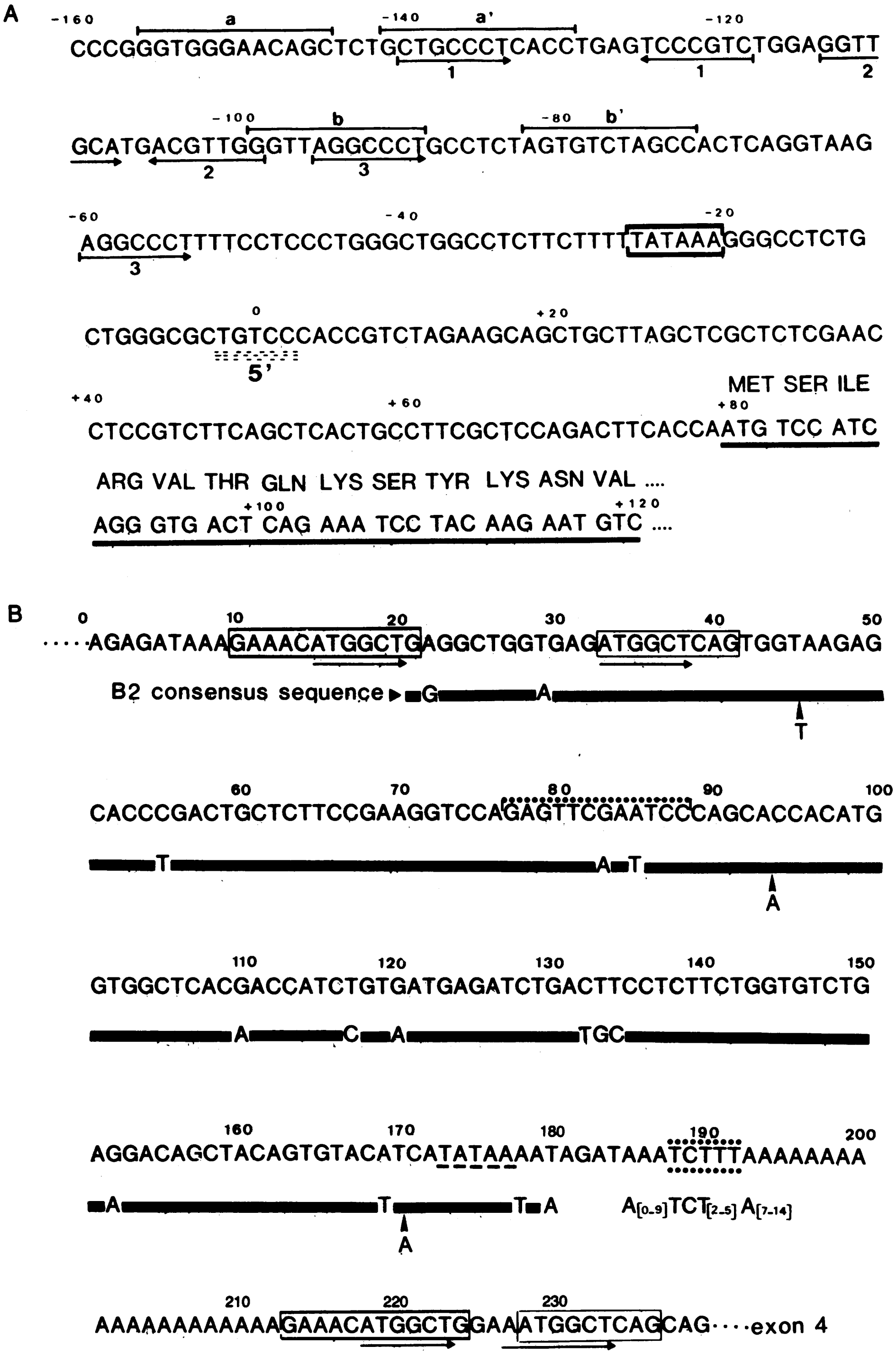
A
-160
a
-140
a'
-120
CCCGGGTGGGAACAGCTCTGCTGCCCTCACCTGAGTCCCGTCTGGAGGTT
1
1
2
-100
b
-80
b'
GCATGACGTTGGGTTAGGCCCTGCCTCTAGTGTCTAGCCACTCAGGTAAG
2
3
-60
-40
-20
AGGCCCTTTTCCTCCCTGGGCTGGCCTCTTCTTTTTATAAAGGGCCTCTG
3
0
+20
CTGGGCGCTGTCCCACCGTCTAGAAGCAGCTGCTTAGCTCGCTCTCGAAC
5'
MET SER ILE
+40
+60
+80
CTCCGTCTTCAGCTCACTGCCTTCGCTCCAGACTTCACCAATG TCC ATC
ARG VAL THR GLN LYS SER TYR LYS ASN VAL ....
+100
+120
AGG GTG ACT CAG AAA TCC TAC AAG AAT GTC ....
B
0
10
20
30
40
50
·····AGAGATAAAGAAACATGGCTGAGGCTGGTGAGATGGCTCAGTGGTAAGAG
B2 consensus sequence
G
A
T
60
70
80
90
100
CACCCGACTGCTCTTCCGAAGGTCCAGAGTTCGAATCCCAGCACCACATG
T
A
T
A
110
120
130
140
150
GTGGCTCACGACCATCTGTGATGAGATCTGACTTCCTCTTCTGGTGTCTG
A
C
A
TGC
160
170
180
190
200
AGGACAGCTACAGTGTACATCATATAAAATAGATAAATCTTTAAAAAAAA
A
T
T
A
A
A[0-9]TCT[2-5]A[7-14]
210
220
230
AAAAAAAAAAAAGAAACATGGCTGGAAATGGCTCAGCAG····exon 4

FIG. 3. Partial nucleotide sequence of the 5′ end of the third intron of α1. (*A*) Sequence of the 5′ end of α1. The 5′ capping region is schematized by dotted lines, and the TATA sequence is boxed. Bars a–a′ and b–b′ indicate palindromic structures. Arrows underlaying the sequence indicate direct (nos. 1 and 2) and inverted (no. 3) repeats. The coding portion of the sequence is underlined and translated. (*B*) Partial sequence of the third intron of α1. Nucleotide 1 is 220 bp from the *Eco*RI site mapped at the beginning of the third intron and nucleotide 239 is 32 bp from the *Bam*HI site (see Fig. 1). Direct repeats flanking the B2 element are boxed. Arrows indicate the sequence shared by the two repeats. Initiator and terminator sites for RNA polymerase III are indicated by dots.

tures. Such (G + C)-rich structures have been described as regulatory regions in the mouse metallothionein gene (23), in human and rabbit globin genes (24), and in *Drosophila* hsp70 genes (25).

In the third intron of α1, we have detected a B2 sequence. B2 is a family of interspersed repetitive elements, which are scattered throughout the mouse genome (26). B2 sequences have been found associated with single-copy genes (27). These elements are transcribed and the RNA has been found in both the nucleus and the cytoplasm, mainly as double-stranded molecules (26). In the cytoplasm, the transcripts are mostly detected as small (0.6 kb) poly(A)$^+$ RNA. We have compared the level of expression of B2 and endo A in various cell types and observed that the transcription of endo A and of the B2 family is exclusive. The question is to determine whether or not this observation merely represents a coincidence. Do such repetitive elements have a role in the regulation of gene expression? The observation that repeat sequence expression is associated, in various systems (28, 29), with developmental changes does not demonstrate directly any function but suggests, at least, a correlation with the variation in gene expression. The hypothesis that a set of unlinked genes could be coordinately regulated by means of *cis* interactions with a common regulatory molecule retains an attractive simplicity. This idea requires that some form of homologous sequences are linked to genes that have to be coregulated. The short interspersed repetitive sequences might represent such regulatory elements. A short repetitive "ID sequence" has recently been found to be present in genes transcribed only in rat brain (30). In cells in which

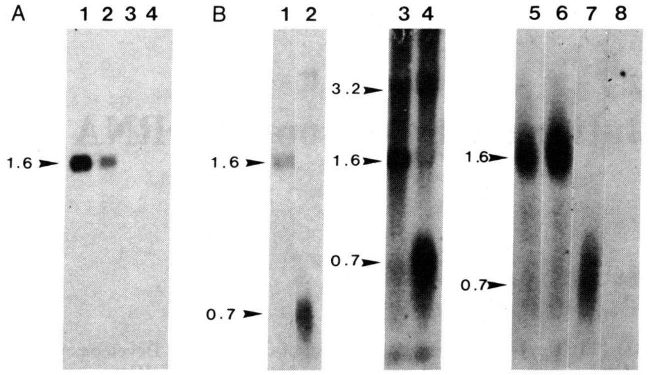

FIG. 4. RNA blot of RNA. (*A*) The probe used was the pseudogene α2. Lanes 1 and 2: 6 and 2 μg of TDM-1 RNA; lanes 3 and 4: 6 and 2 μg of F9 RNA. (*B*) The probe used contained both B2 and endo A sequences. Lanes 1 and 2: 1 and 20 μg of TDM-1 RNA; lanes 2 and 4: 1 and 20 μg of F9 RNA; lanes 5, 6, 7, and 8: 2 μg of RNA from kidney, liver, L cells, and 3T3 cells, respectively. In *A*, samples were run on a 1% agarose gel. In *B*, agarose gels were 1% except for samples in lanes 3 and 4, which were run on a 1.8% gel.

these brain-specific genes are expressed, the ID elements themselves are also transcribed into small poly(A)$^+$ RNA by RNA polymerase III. These results have suggested a model in which polymerase III transcription of ID sequences located in introns of brain-specific genes would activate these genes (31). No molecular evidence is available to support this hypothesis. The observation of the mutually exclusive expression of B2 and endo A suggests something very different.

The isolation of the endo A gene provides a system for analysis of the regulation of a protein that is expressed in the trophectoderm but not in the inner cell mass cells of the blastocyst. In embryonal carcinoma cells, the activity of this gene is regulated at the transcriptional level. Since the expression of endo A in embryonal carcinoma cells (9) as well as in the early embryo (7) seems to be induced by a topological event (i.e., the compaction/polarization process) (32), this system may provide an insight into molecular mechanisms directly related to a morphogenetical phenomenon.

We are very grateful to Odile Croissant for instructing one of us (M.V.) in DNA electron microscopy techniques and overseeing the heteroduplex preparation and analysis. We are indebted to M. Darmon and H. Condamine for enlightening discussions. This work was supported by the Centre National de la Recherche Scientifique (Grant LA 269), the Fondation pour la Recherche Médicale, the Ligue Nationale contre le Cancer, the Ministère de la Recherche et de l'Industrie (Grant 82V1388), and the Fondation André Meyer.

1. Johnson, M. H., Chabraborty, J., Handyside, A. H., Willison, K. & Stern, P. (1979) *J. Embryol. Exp. Morphol.* **54,** 241–261.
2. Van Blerkom, J., Barton, S. C. & Johnson, M. H. (1976) *Nature (London)* **259,** 319–321.
3. Jackson, B. W., Grund, C., Schmid, E., Bûrki, K., Franke, W. W. & Illmensee, K. (1980) *Differentiation* **17,** 161–179.
4. Brûlet, P., Babinet, C., Kemler, R. & Jacob, F. (1980) *Proc. Natl. Acad. Sci. USA* **77,** 4113–4117.
5. Oshima, R. G. (1982) *J. Biol. Chem.* **257,** 3414–3421.
6. Jackson, B. W., Grund, C., Winter, S., Franke, W. W. & Illmensee, K. (1981) *Differentiation* **20,** 203–216.
7. Oshima, R. G., Howe, W. E., Tabor, J. M. & Trevor, K. (1983) in *Teratocarcinoma Stem Cells*, eds. Silver, L. M., Martin, G. R. & Strickland, S. (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY), Vol. 10, pp. 51–61.
8. Kemler, R., Brûlet, P., Schnebelen, M. T., Gaillard, J. & Jacob, F. (1981) *J. Embryol. Exp. Morphol.* **64,** 45–60.
9. Brûlet, P. & Jacob, F. (1982) *Proc. Natl. Acad. Sci. USA* **79,** 2328–2332.
10. Nicolas, J. F., Jakob, H. & Jacob, F. (1981) in *Functionally Differentiated Cell Lines*, ed. Sato, G. (Liss, New York), pp. 185–210.
11. Benton, W. & Davis, R. W. (1977) *Science* **196,** 180–182.
12. Kaback, D. B., Angerer, L. M. & Davidson, N. (1979) *Nucleic Acids Res.* **6,** 2499–2517.
13. Auffray, C. & Rougeon, F. (1980) *Eur. J. Biochem.* **107,** 303–314.
14. Nicolas, J.-F. & Berg, P. (1983) in *Teratocarcinoma Stem Cells*, eds. Silver, L. M., Martin, G. R. & Strickland, S. (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY), Vol. 10, pp. 469–485.
15. Sanger, F. & Coulson, A. R. (1978) *FEBS Lett.* **87,** 107–110.
16. Sanger, F., Nicklen, S. & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. USA* **74,** 5463–5467.
17. Yaniv, M. (1984) *Biol. Cell.* **50,** 203–216.
18. Krayev, A. S., Markusheva, T. V., Kramerov, D. A., Ryskov, A. P., Skryabin, K. G., Bayev, A. A. & Georgiev, G. P. (1982) *Nucleic Acids Res.* **10,** 7461–7475.
19. Lescure, B. & Arcangioli, B. (1984) *EMBO J.* **3,** 1067–1073.
20. Zakut-Houri, R., Oren, M., Bienz, B., Lavie, V., Hazum, S. & Givol, D. (1983) *Nature (London)* **306,** 594–597.
21. McGrath, J. P., Capon, D. J., Smith, D. H., Chen, E. Y., Seeburg, P. H., Goeddel, D. V. & Levinson, A. D. (1983) *Nature (London)* **304,** 501–506.
22. Hollis, G. F., Hieter, P. A., McBride, O. W., Swan, D. & Leder, P. (1981) *Nature (London)* **296,** 321–325.
23. Glanville, M., Durnam, D. M. & Palmiter, R. D. (1981) *Nature (London)* **292,** 267–269.
24. Mellon, P., Parker, V., Gluzman, Y. & Maniatis, T. (1981) *Cell* **27,** 279–288.
25. Karch, F., Török, I. & Tissière, A. (1981) *J. Mol. Biol.* **148,** 219–230.
26. Kramerov, D. A., Grigoryan, A. A. & Georgiev, G. P. (1979) *Nucleic Acids Res.* **6,** 697–713.
27. Page, G. S., Smith, S. & Goodman, H. (1981) *Nucleic Acids Res.* **9,** 2087–2104.
28. Davidson, E. H. & Posakony, J. W. (1982) *Nature (London)* **297,** 633–635.
29. Zuker, C. & Lodish, H. F. (1981) *Proc. Natl. Acad. Sci. USA* **78,** 5386–5390.
30. Milner, R. J., Bloom, F. E., Lai, C., Lerner, R. A. & Sutcliffe, J. G. (1984) *Proc. Natl. Acad. Sci. USA* **81,** 713–717.
31. Sutcliffe, J. G., Milner, R. J., Gottesfeld, J. M. & Lerner, R. A. (1984) *Nature (London)* **308,** 237–241.
32. Johnson, M. H. (1981) *Biol. Rev.* **56,** 463–498.