Volume 322, number 2, 89–94 FEBS 12399 May 1993

Exon/intron organization of the gene encoding the mouse epithelin/granulin precursor (acrogranin)

Tadashi Baba[a], Hiromi Nemoto[a], Ken Watanabe[a], Yuji Arai[a] and George L. Gerton[b]

[a]*Institute of Applied Biochemistry, University of Tsukuba, Ibaraki 305, Japan* and [b]*Department of Obstetrics and Gynecology, University of Pennsylvania School of Medicine, Philadelphia, PA 19104-6080, USA*

Received 24 February 1993; revised version received 18 March 1993

Mouse genomic clones encoding the epithelin/granulin gene and its 5′- and 3′-flanking regions have been isolated and sequenced. This gene was found to be a single-copy gene, and contained 13 exons interrupted by 12 introns. Eight out of the 12 introns are classified as phase 0, and are located within the central part of each of the tandem repeats in the amino acid sequence of the epithelin/granulin precursor. The first intron is unique because of the interruption of the 5′-untranslated region and its fairly large size (approximately 2.4 kbp). Consensus sequences for several of the potential regulatory elements are present in the 5′-flanking sequence, including a common CCAAT sequence.

Epithelin; Granulin; Acrogranin; Genomic DNA; Gene organization; Growth-modulating peptide

1. INTRODUCTION

The acrosome is a sperm-specific organelle overlying the anterior part of the sperm head. To examine acrosome biogenesis during spermatogenesis, we have identified several genes coding for acrosomal proteins [1-3], and used these genes as markers of acrosome formation. Acrogranin is a 67 kDa glycoprotein which was co-purified through the initial steps of the purification of proacrosin from acid extracts of guinea pig testes [4]. Immunohistochemical studies using an antiserum against acrogranin revealed the localization of this protein in proacrosomal granules of meiotic spermatocytes, as well as in acrosomes of round spermatids [4]. This protein was not detectable by immunohistochemistry in Sertoli cells, Leydig cells, or early-stage germ cells, such as spermatogonia and early spermatocyte [3,4].

Molecular cloning of guinea pig and mouse acrogranins [3] reveals that this protein is a cysteine-rich molecule made up of 7 tandem repeat sequences. Surprisingly, the sequences of several of these repeats are highly homologous to those of rat epithelins [5] and human granulins [6], putative growth-modulating peptides. The cDNA sequence of the mouse epithelin precursor [7] is, in fact, identical to that of mouse acrogranin [3]. Since the structural and functional features of the epithelins and granulins are considered to be identical [7,8], acrogranin is the same protein as the epithelin/granulin precursor. Furthermore, Northern blot analysis indicates that the acrogranin (epithelin/granulin) gene is ubiquitously expressed in all tissues, including testis [3,7,8]. Therefore, the male germ cells appear to be unable to process the epithelin/granulin precursor into the active peptides because the precursor molecule is present throughout spermatogenesis and undergoes only partial proteolysis in epididymal sperm [3,4]. Recently, the structural organization and chromosomal location of the human granulin gene have been reported by Bhandari and Bateman [9]; the protein-coding region of the human granulin gene comprises 12 exons, and the granulin gene is assigned to human chromosome 17, however, the true first exon and intron of the human gene have not yet been identified.

It is of great interest to examine the correlation between the tandem repetitive structure of the epithelin/granulin precursor containing oppositely functional components [5,7] and the arrangement of the protein-coding regions in the gene. Moreover, to facilitate further studies of the epithelins and granulins, the gene structure needs to be established. In this paper, we describe the exon/intron organization of the mouse epithelin/granulin gene. The sequence of the putative promoter region is also reported.

Correspondence address: T. Baba, Institute of Applied Biochemistry, University of Tsukuba, Tsukuba Science City, Ibaraki 305, Japan. Fax: (81) (298) 53-6632.

The nucleotide sequence data reported in this paper will appear in the DDBJ, EMBL, and GenBank Nucleotide Sequence Databases.

2. EXPERIMENTAL

2.1. *Materials*

A mouse genomic library in Charon28, pUCSV3cat, and pUCSV0cat were kind gifts of Dr. Akiyoshi Fukamizu at the Institute of Applied Biochemistry, University of Tsukuba. These two plasmids were derived from pUC19 and pSV2-CAT [10]. pUCSV3cat contains

the chloramphenicol acetyltransferase (CAT) gene driven by the simian virus 40 (SV40) early promoter and SV40 enhancer, whereas the two SV40-derived sequences are missing in pUCSV0cat [10]. The radioisotope, [α-^{32}P]dCTP (3,000 Ci/mmol), was purchased from Bresatec Ltd., Australia. Restriction endonucleases and modifying enzymes were purchased from Nippon Gene (Toyama, Japan). All other reagents were of the highest purity available.

2.2. *Screening of a mouse genomic library*

The mouse genomic library was screened by the plaque hybridization method [11]. Plaque lifts were prehybridized at 42°C in 5 × SSPE (1 × SSPE = 10 mM sodium phosphate, pH 7.7, 0.18 M NaCl, and 1 mM EDTA), 0.02% Ficoll 400, 0.02% polyvinylpyrrolidone, 0.02% BSA, and 0.1% SDS, followed by hybridization at 60°C overnight in the prehybridization buffer containing denatured salmon testis DNA (0.1 mg/ml) and ^{32}P-labeled probe. The filters were washed in 2 × SSC (1 × SSC = 15 mM sodium citrate, pH 7.0, and 0.15 M NaCl) at room temperature for 10 min, in 2 × SSC containing 0.1% SDS at 60°C for 10 min, and in 2 × SSC at room temperature for 10 min, prior to autoradiography at −80°C. Positive clones were plaque purified, and the DNA fragments were subcloned into the appropriate sites of pUC19 or pUC119 for further characterization.

2.3. *Transfection and CAT assays*

HeLa cells were grown in Dulbecco's modification of Eagle's medium (low glucose, Difco) with 10% fetal calf serum in 60-mm dishes. Plasmid DNA (3 μg) was introduced into HeLa cells by the calcium phosphate co-precipitation procedure [12], using a CellPhect transfection kit (Pharmacia LKB Biotechnology). Cells were harvested 48 h after the DNA transfection. The cell lysates were obtained by 3–5 cycles of freeze–thawing in a cell-extraction buffer of 250 mM Tris-HCl, pH 8.0, and 5 mM EDTA. After centrifugation, the supernatant was heated at 60°C for 10 min to inactivate contaminating deacetylase. CAT assays [13] were carried out at 37°C for 2 h in a mixture (0.15 ml) containing 140 mM Tris-HCl, pH 7.8, 4 mM acetyl coenzyme A, 0.2 μCi of [^{14}C]chloramphenicol (DuPont/New England Nuclear, 0.05 mCi/ml, 50–60 mCi/mmol), and 40 μg of protein from the above extract. The samples were analyzed on 0.25-mm Polygram plates (Macherey-Nagel), using a 19:1 chloroform/methanol mixture as a running solvent. The chromatograms were subjected to autoradiography at −80°C. This experiment was repeated three times to assure reproducibility.

2.4. *Analytical procedures*

Nucleotide sequence analysis was carried out by the dideoxy chain-termination method [14], using a commercial kit (Sequenase Version 2.0) from US Biochemicals. Southern blot analysis was carried out as described previously [15]. Computer-aided analysis of nucleotide and protein sequences was carried out using a GENETYX program (Software Development Co., Tokyo).

3. RESULTS

Two positive clones, termed λMAG8 and λMAG11, were identified by screening of approximately 1×10^6 plaques from a mouse genomic library, using a cDNA fragment of pMG1 [3] encoding the mouse epithelin/granulin precursor as a probe. The phage DNAs were purified and analyzed by restriction enzyme digestion followed by Southern blot hybridization, using different parts of the mouse epithelin/granulin precursor cDNA sequence as probes. These two clones overlapped and contained two *Eco*RI–*Eco*RI fragments of 4.4 and 7.4 kbp, which hybridized to the *Eco*RI–*Xba*I and *Bam*HI–*Eco*RI fragments carrying the 5′ and 3′ sections of the cDNA sequence, respectively (Fig. 1). Thus, the *Eco*RI–*Eco*RI fragments were subjected to sequence analysis. Southern blot analysis showed that only two *Eco*RI fragments (4.4 and 7.4 kbp) from mouse genomic DNA hybridized to the 2.1-kbp probe prepared from the isolated cDNA clone, pMG1 [3] (data not shown). This result indicates that the epithelin/granulin precursor is a single gene product.

The organization of the mouse epithelin/granulin gene together with the mRNA, and the genomic DNA sequence including 5′- and 3′-flanking sequences are shown in Figs. 1 and 2, respectively. This gene is approximately 6.3 kbp in length and consists of 13 exons separated by 12 introns. There are three nucleotide substitutions between the mouse genomic and cDNA sequences: the substitutions of G, C, and G at nucleotides 609, 1,229, and 1,231 in the cDNA sequence to C, G, and C (see Fig. 2 and [3]). The substitution at nucleotide 609 makes no change in the amino acid sequence, while the Ser–Ala sequence at residues 385–386 [3] is changed into a Cys–Pro sequence by the nucleotide substitutions at nucleotides 1,229 and 1,231. Thus, the amino acid sequence of the mouse epithelin/granulin precursor obtained by us [3] is identical to that reported by Plowman et al. [7], except for an Arg at residue 333. Our cDNA sequence [3] contains only 17 nucleotides in the first exon (exon 1), so that the translation initiator codon,

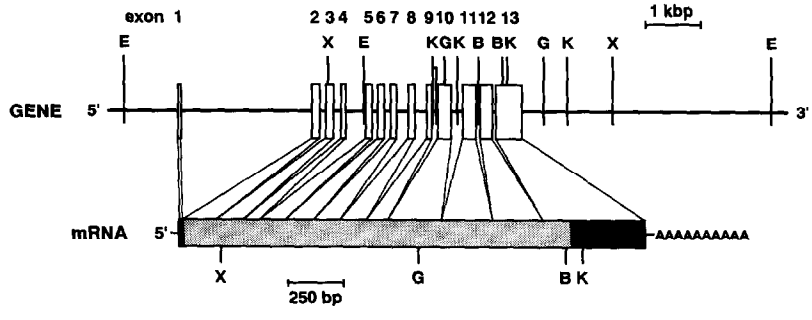

Fig. 1. Organization of the mouse epithelin/granulin gene. The gene contains 13 exons which are numbered at the top (exons 1–13) and indicated by open boxes. The mRNA structure is also represented; the 5′- and 3′-untranslated regions are shown by black boxes. The protein-coding region is indicated by a dotted box. Restriction enzyme sites are shown as follows: B, *Bam*HI; E, *Eco*RI; G, *Bgl*II; K, *Kpn*I; X, *Xba*I.

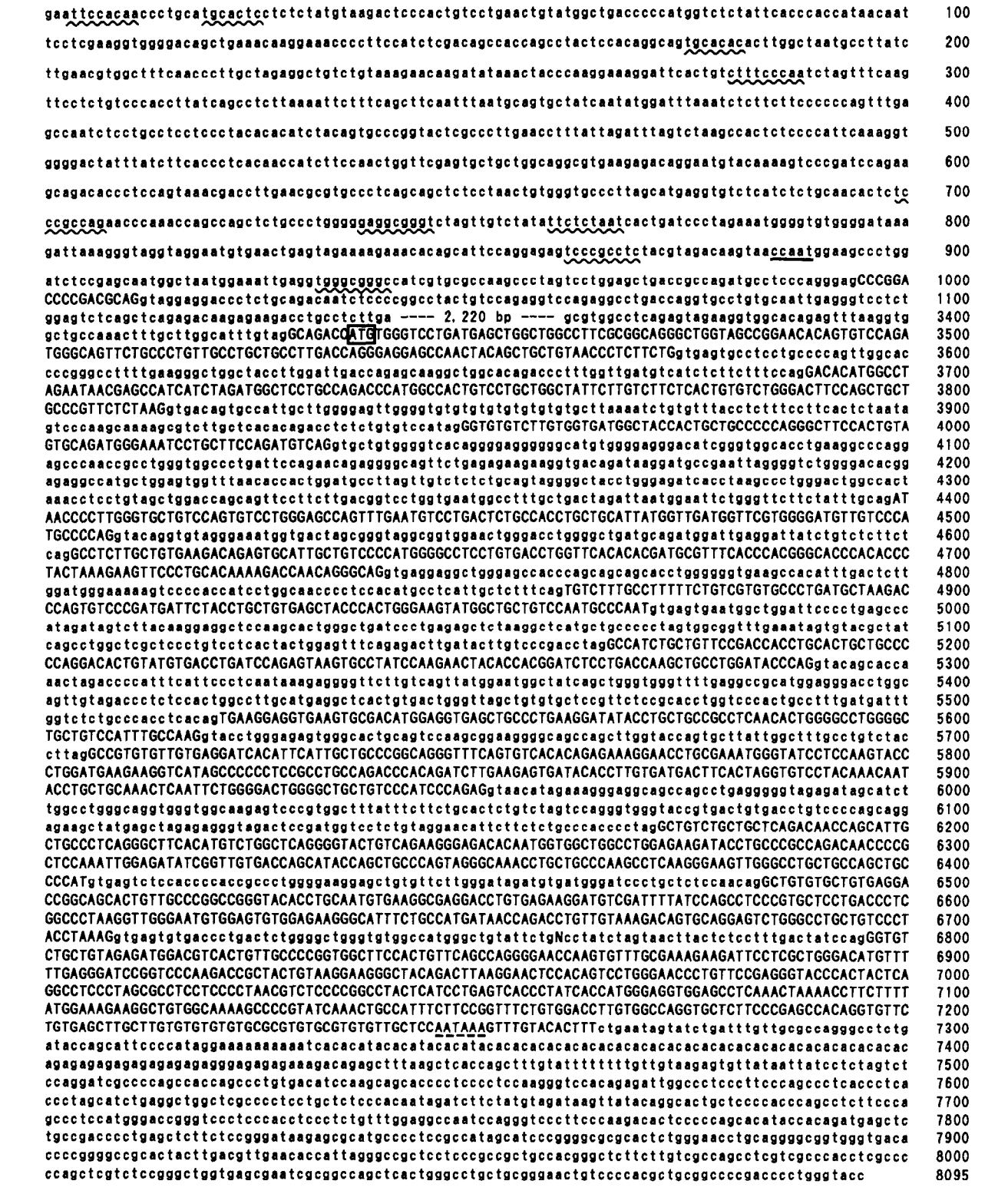

Fig. 2. DNA sequence of the mouse epithelin/granulin gene. The sequences encoded by exons are given by capital letters. A possible CCAAT sequence is underlined. The translation initiation codon ATG and polyadenylation signal sequence AATAAA are, respectively, boxed and underlined with a broken line. The consensus sequences for several potential regulatory elements are underlined with wavy lines: NF-IL6, 5′-ATTCCACAA-3′ (reverse orientation), 5′-CTTTCCCAA-3′ (reverse orientation), and 5′-TTCTCTAAT-3′ at nucleotides 3–11, 280–288, and 759–767, respectively; MTF-1, 5′-TGCACTC-3′ and 5′-TGCACAC-3′ at nucleotides 19–25 and 175–181; Sp-1, 5′-TCCCGCCAG-3′ (reverse orientation), 5′-GAGGCGGGT-3′, 5′-TCCCGCCTC-3′ (reverse orientation), and 5′-TGGGCGGGC-3′ at nucleotides 699-707, 737–745, 861–869, and 932–940, respectively.

Table I

Nucleotide sequences of splice junctions in the mouse epithelin/granulin gene

Exon			Intron				Exon		Intron phase class
Term	Size (nt)[a]	Sequence	Donor	Term	Size (nt)[a]	Acceptor	Sequence	Term	
1	17	CCCGACGCAG	gtaggagg	A	2417	atttgtag	GCAGACCATG Met	2	
2	145	CCCTCTTCTG nProLeuLeu	gtgagtgc	B	115	ctttccag	GACACATGGC AspThrTrpP	3	0
3	126	GTTCTCTAAG oPheSerLys	gtgacagt	C	134	gtccatag	GGTGTGTCTT GlyValSerC	4	0
4	85	CAGATGTCAG GlnMetSerA	gtgctgtg	D	365	atttgcag	ATAACCCCTT spAsnProLe	5	1
5	110	CATGCCCCAG oMetProGln	gtacaggt	E	95	cttctcag	GCCTCTTGCT AlaSerCysC	6	0
6	136	AACAGGGCAG AsnArgAlaV	gtgaggag	F	121	tctttcag	TGTCTTTGCC alSerLeuPr	7	1
7	110	AATGCCCAAT oMetProAsn	gtgagtga	G	195	cgacctag	GCCATCTGCT AlaIleCysC	8	0
8	124	GGATACCCAG GlyTyrProV	gtacagca	H	231	cctcacag	TGAAGGAGGT alLysGluVa	9	1
9	98	ATTTGCCAAG oPheAlaLys	gtacctgg	I	87	taccttag	GCCGTGTGTT AlaValCysC	10	0
10	246	CATCCCAGAG oIleProGlu	gtaacata	J	220	acccctag	GCTGTCTGCT AlaValCysC	11	0
11	234	GCTGCCCCAT nLeuProHis	gtgagtct	K	78	tccaacag	GCTGTGTGCT AlaValCysC	12	0
12	225	CTACCTAAAG oTyrLeuLys	gtgagtgt	L	87	ctatccag	GGTGTCTGCT GlyValCysC	13	0
13	469	TGTACACTTT							

[a] Nucleotide.

The sequences of exons and introns are indicated by capital and small letters, respectively.

ATG, is encoded in the second exon (exon 2). A putative poly(A) tail of the mouse epithelin/granulin precursor cDNA [3] is added 13 nucleotides downstream of the consensus signal sequence, AATAAA. The G/T-cluster sequence, which is important for 3′ processing of polyadenylated mRNAs [16], is located at nucleotides 7,280–7,286 (Fig. 2). It should be noted that there are CA- and GA-repetitive sequences further downstream of the G/T-cluster sequence.

Although the transcription initiation site of the epithelin/granulin gene is not clear at present, a consensus CCAAT sequence was located in the putative promoter region of the gene, 105 nucleotides upstream from the 5′ end of the known cDNA sequence (Fig. 2). There was no typical TATA sequence in the promoter region. To examine whether the 5′-flanking sequence of the epithelin/granulin gene functions as a promoter, CAT assays were carried out using HeLa cells (Fig. 3). The CAT activity was detectable in cells transfected by each of the six plasmids containing the sequences with different lengths from the *Eco*RI–*Pst*I fragment at nucleotides 1–1,029 (Fig. 2). The activity in pmAG12 was obvious, but lower than those in other plasmids. By computer analysis, several putative *cis*-acting sequences involved in the regulation of transcription are identified, including the consensus sequences for the Sp1-binding site, the MTF-binding site (MTF-1), and the IL-1/IL-6-responsive element (NF-IL6) [17]. In addition, an AT-rich sequence containing nine repeats of ATTA, which is similar to a recognition signal for selective degradation of the mRNAs for some lymphokines, cytokines, and proto-oncogenes [18], is present at nucleotides 2,105–2,177 in the first intron (not shown in Fig. 2).

The sizes of exons and introns, and intron phase classes are summarized in Table I. All sequences at the exon/intron boundaries are consistent with the consensus GT/AG sequence at the donor and acceptor sites of RNA splicing [19]. The sizes of introns range from 78 to 365 nucleotides, except for the first intron (intron A) with a size of 2,417 nucleotides. The largest exon (exon 13) encodes the carboxy-terminal region of the epithelin/granulin precursor and the 3′-untranslated region. All phase classes of introns are in phase 0, except for introns A, D, F, and H. The phase classes of the introns D, F, and H are in phase 1. As shown in Fig. 4, the phase 0 introns are all located within the central part of each of the tandem repeats in the amino acid sequence of the epithelin/granulin precursor. Two out of three phase 1 introns (introns F and H) are present in the regions preceding the amino termini of repeats #3 and

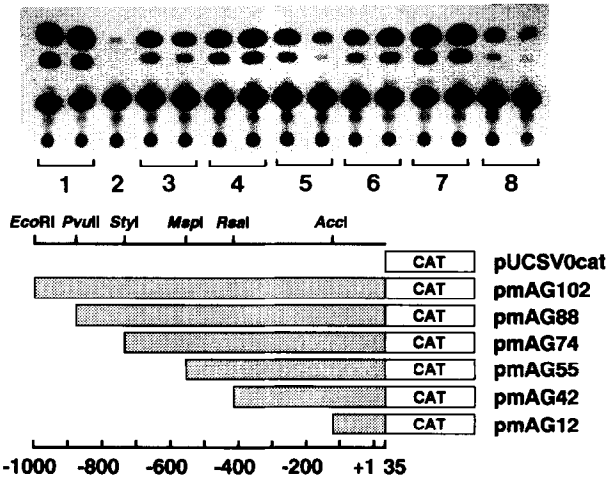

Fig. 3. CAT activity of extracts from HeLa cells transfected by plasmids of mouse epithelin/granulin-CAT chimeric gene. An *Eco*RI–*Pst*I DNA fragment at nucleotides 1–1,029 carrying the 5′-flanking region, first exon (exon 1), and a part of first intron (intron A) of the epithelin/granulin gene, was introduced into the *Eco*RI–*Pst*I site of pUC19. Six different fragments were produced by digestion with *Hin*dIII and restriction enzymes indicated. After addition of an *Hin*dIII linker, the DNA fragments were inserted into the *Hin*dIII site of pUCSV0cat (lane 2, see [10]). pUCSV3cat (lane 1) contains the SV40 early promoter and enhancer sequences. The plasmids, including pmAG102 (lane 3), pmAG88 (lane 4), pmAG74 (lane 5), pmAG55 (lane 6), pmAG42 (lane 7), and pmAG12 (lane 8), were transfected into HeLa cells. The CAT activity exhibited by extracts of transfected cells was detected by autoradiography. Duplicate assays show the results of independent transfections. The scale at the bottom indicates nucleotide position in base pairs from the 5′ end of the cDNA sequence reported by us [3].

#4 which correspond to epithelins 2 and 1, respectively. The exon/ intron organization of the mouse epithelin/granulin gene within the protein-coding region is identical to that of the human gene described by Bhandari and Bateman [9].

4. DISCUSSION

A mouse genomic fragment containing the epithelin/granulin gene was characterized, including approximately 1 kbp of the 5′-flanking region. Sequence analysis shows that this gene is separated into 13 exons interrupted by 12 introns (Figs. 1 and 2). The first intron is present within the region corresponding to the 5′-untranslated region of the mRNA. The presence of a first intron between the first and second exons has been also reported for other genes, including the corticotropin/β-lipotropin precursor [20,21], interleukin 1α and -β [22,23], and tumor necrosis factor genes [24,25]. Since these gene products are the precursors of bioactive peptides or proteins, as in the case for the epithelin/granulin gene, it is interesting to suppose that these first introns play an important role in the regulation of gene transcription negatively and/or positively. Moreover, consensus sequences for regulatory elements, including the Sp1-binding site, MTF-1, and NF-IL6, are located in the 5′-flanking region of the mouse epithelin/granulin gene (Fig. 2). An AT-rich sequence similar to the recognition signal for selective degradation of mRNA [18] is also present in the first intron, however, it is not certain that these sequences are regulationally functional. Thus, further experiments are necessary to elucidate the regulatory mechanism of the epithelin/granulin gene.

There are nine AG sequences in the region between a consensus CCAAT sequence at nucleotides 885–889 and the 5′ end of the cDNA sequence reported by us [3] (Fig. 2). It is thus possible to consider the presence of additional exon(s) in the 5′-flanking region of the mouse

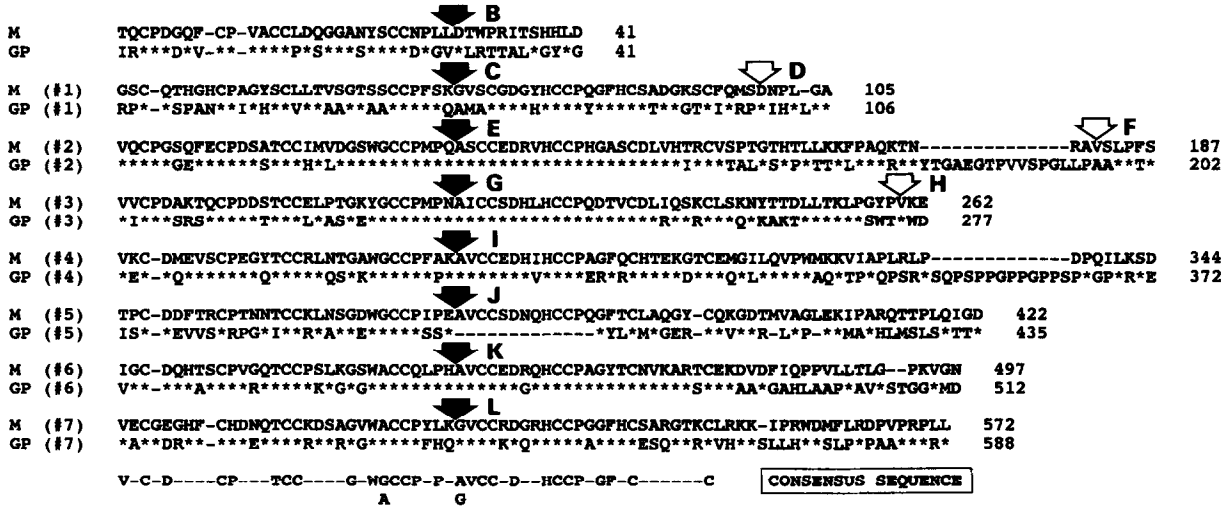

Fig. 4. Location of introns in the tandem repetitive structure of the mouse epithelin/granulin precursor. The positions of the phase 0 (introns B, C, E, G, I, J, K and L) and phase 1 introns (introns D, F and H, see Table I) are shown in the amino acid sequence of the mouse epithelin/granulin precursor by closed and open arrows, respectively. Since the sequence of the epithelin/granulin precursor is highly conserved among human, rat, and mouse species (see [3,7,8]), the mouse sequence (M) is aligned only with the guinea pig sequence (GP). The consensus sequence for the seven tandem repeats (repeats #1–7) is indicated at the bottom.

epithelin/granulin gene described here, however, CAT assay of the 5'-flanking region demonstrates that at least a 121-bp sequence at nucleotides 874–994, including the CCAAT sequence, is able to function as a promoter sequence (Fig. 3). Therefore, the presence of an additional exon(s) is unlikely, or if it exists, it would be a small one. The size of the combined length of all 13 exons is close to that of the gene transcript determined by Northern blot analysis [3].

Seven tandem repeats in the protein-coding region of the epithelin/granulin precursor are separated by introns which are located within the central parts of the repeats (Table I and Fig. 4). Since the phase classes of these introns are all in phase 0, the exons of the epithelin/granulin gene may have been duplicated during evolution and assembled into the gene by recombination, according to the exon shuffling model [26]. The introns D, F, and H, which are present at positions close to the amino-termini of repeats #2, #3, and #4, corresponding to granulin F, epithelin 2, and epithelin 1, respectively [7,8], belong to different phase classes (phase 1). Therefore, these three introns had been located at the corresponding positions before the epithelin/granulin gene was assembled.

As described by Bhandari et al. [8], the functions of the epithelins and granulins are still speculative; only epithelins 1 and 2 (granulin A) inhibit proliferation of an human epithelial tumor cell line, A431 [5,8]. The functions of granulins B–G have not yet been found. Nevertheless, the nucleotide sequence of the epithelin/granulin gene makes it possible to examine the regulation of this gene expression.

Acknowledgements: We acknowledge the technical assistance of Mr. Shintaroh Ohnishi. This study was partly supported by grants from the Ministry of Education, Science, and Culture of Japan to T.B. (#04760056), and from the National Institutes of Health (HD-22899) and the Mellon Foundation to G.L.G.

REFERENCES

[1] Baba, T., Kashiwabara, S., Watanabe, K., Itoh, H., Michikawa, Y., Kimura, K., Takada, M., Fukamizu, A. and Arai, Y. (1989) J. Biol. Chem. 264, 11920–11927.
[2] Kashiwabara, S., Arai, Y., Kodaira, K. and Baba, T. (1990) Biochem. Biophys. Res. Commun. 173, 240–245.
[3] Baba, T., Hoff, H.B., Nemoto, H., Lee, H., Orth, J., Arai, Y. and Gerton, G.L. (1993) Mol. Reprod. Dev. 34, 233–243.
[4] Anakwe, O.O. and Gerton, G.L. (1990) Biol. Reprod. 42, 317–328.
[5] Shoyab, M., McDonald, V.L., Byles, C., Todaro, G.J. and Plowman, G.D. (1990) Proc. Natl. Acad. Sci. USA 87, 7912–7916.
[6] Bateman, A., Belcourt, D., Bennett, H., Lazure, C. and Solomon, S. (1990) Biochem. Biophys. Res. Commun. 173, 1161–1168.
[7] Plowman, G.D., Green, J.M., Neubauer, M.G., Buckley, S.D., McDonald, V.L., Torado, G.J. and Shoyab, M. (1992) J. Biol. Chem. 267, 13073–13078.
[8] Bhandari, V., Palfree, R.G.E. and Bateman, A. (1992) Proc. Natl. Acad. Sci. USA 89, 1715–1719.
[9] Bhandari, V. and Bateman, A. (1992) Biochem. Biophys. Res. Commun. 188, 57–63.
[10] Fukamizu, A., Kanimoto, K., Uehara, S., Sheo, M.S., Handa, S., Sagara, M., Takahashi, S., Imai, T. and Murakami, K. (1991) Biomed. Biochim. Acta 50, 659–663.
[11] Benton, W.D. and Davis, R.W. (1977) Science 196, 180–182.
[12] Graham, F.L. and Van der Eb, A.J. (1973) Virology 52, 456–467.
[13] Gorman, C.M., Moffat, L.F. and Howard, B.H. (1982) Mol. Cell. Biol. 2, 1044–1051.
[14] Sanger, F., Nicklen, S. and Coulson, A.R. (1977) Proc. Natl. Acad. Sci. USA 74, 5463–5467.
[15] Selden, R.F. (1987) in: Current Protocols in Molecular Biology (Ausubel, F.M., Brent, R., Kingston, R.E., Moore, D.D., Seidman, J.G., Smith, J.A. and Struhl, K., eds.) pp. 2.9.1–2.9.10, Wiley, New York.
[16] Birnstiel, M.L., Busslinger, M. and Strup, K. (1985) Cell 41, 349–359.
[17] Faisst, S. and Meyer, S. (1992) Nucleic Acids Res. 20, 3–26.
[18] Shaw, G. and Kamen, R. (1986) Cell 46, 659–667.
[19] Breathnach, R. and Chambon, P. (1981) Annu. Rev. Biochem. 50, 349–383.
[20] Nakanishi, S., Teranishi, Y., Noda, M., Notake, M., Watanabe, Y., Kakidani, H., Jingami, H. and Numa, S. (1980) Nature 287, 752–755.
[21] Nakanishi, S., Teranishi, Y., Watanabe, Y., Notake, M., Noda, M., Kakidani, H., Jingami, H. and Numa, S. (1981) Eur. J. Biochem. 115, 429–438.
[22] Furutani, Y., Notake, M., Fukui, T., Ohue, M., Nomura, H., Yamada, M. and Nakamura, S. (1986) Nucleic Acids Res. 14, 3167–3179.
[23] Clark, B.D., Collins, K.L., Gandy, M.S., Webb, A.C. and Auron, P.E. (1986) Nucleic Acids Res. 14, 7897–7914.
[24] Nedwin, G.E., Naylor, S.L., Sakaguchi, A.Y., Smith, D., Jarrett-Nedwin, J., Pennica, D., Goeddel, D.V. and Gray, P.W. (1985) Nucleic Acids Res. 13, 6361–6373.
[25] Nedospasov, S.A., Hirt, B., Shakhov, A.N., Dobrynin, V.N., Kawashima, E., Accolla, R.S. and Jongeneel, C.V. (1986) Nucleic Acids Res. 14, 7713–7725.
[26] Patthy, L. (1987) FEBS Lett. 214, 1–7.