J. Biochem 128, 83–92 (2000)

Four Types of Calpastatin Isoforms with Distinct Amino-Terminal Sequences Are Specified by Alternative First Exons and Differentially Expressed in Mouse Tissues[1]

Jiro Takano, Masanori Watanabe, Kiyotaka Hitomi, and Masatoshi Maki[2]

Laboratory of Molecular and Cellular Regulation, Department of Applied Molecular Biosciences, Graduate School of Bioagricultural Sciences, Nagoya University, Furo-cho, Chikusa-ku, Nagoya 464-8601

Received March 21, 2000, accepted April 26, 2000

Calpastatin, a specific inhibitor of calpain, consists of a unique N-terminal domain (domain L) and four repetitive protease-inhibitor domains (domains 1–4). The isolated cDNAs from various mammalian species have conspicuous differences in the regions encoding the N-terminal sequences and can be classified into four types. Mouse and bovine calpastatins (Type I and Type II, respectively), which also differ from each other in the uttermost N-terminal regions, possess longer domain L sequences than those of rabbit, pig, and human inhibitors (Type III). A sequence of a shorter isoform, registered in a DNA data bank, starts from a part of domain 2 with a different N-terminal sequence (Type IV). To clarify the source of this molecular diversity, we investigated the entire exon-intron organization of the mouse calpastatin gene. The previously obtained mouse calpastatin cDNA is encoded by as many as 31 exons including the first exon, designated 1xa. Three additional exons specifying the N-terminal sequences of the other types (designated exons 1xb, 1u, and 14t, respectively) were identified in the mouse genomic DNA sequence. While the mRNAs for Types I and III were expressed at high levels in liver, the Type II mRNA was abundant in heart and skeletal muscle and expressed at lower levels in liver, brain and testis. The Type IV mRNA was specifically expressed in testis among the tissues examined. These results suggest that the calpastatin isoforms possessing different N-terminal sequences are generated by alternative transcription initiation from their own promoters and skipping of the mutually exclusive exons.

Key words: calpastatin, exon, isoform, multiple promoters, protease inhibitor.

Calpastatin is an endogenous inhibitory protein acting specifically on calpains, the calcium-dependent cysteine proteinases present in animal cells (*1*). The calpain-calpastatin system is implicated in various physiological and pathological calcium-dependent processes such as differentiation of myoblasts and adipocytes (*2*, *3*), actin reorganization (*4*), cell cycle (*5*), signal transduction (*6*), apoptosis (*7*, *8*), cataract (*9*), *etc.* (see Refs *10–12* for reviews). A calpastatin molecule contains four mutually homologous inhibitory regions of about 140 amino acid residues (domains 1–4) and a unique but functionally unkown region on its N-terminal side (domain L) (*13*). Conserved residues among the four repetitive domains are not randomly distributed but are clustered in three restricted regions designated A, B, and C, respectively, within each domain. Structure-function analyses have revealed that region B is the inhibitory center of calpastatin and regions A and C potentiate its inhibitory activity through interacting with the calcium-binding domains of calpain subunits (*14–16*). Function of domain L remains unknown.

Plausible full-length calpastatin cDNAs have been isolated from various mammals since the first report on the rabbit inhibitor more than a decade ago (*17–22*). The isolated rabbit calpastatin cDNA possessed an in-frame stop codon upstream of the assigned translation initiation codon. Regardless of the absence of such in-frame stop codons, the translation initiation Met residues have been assigned accordingly for the calpastatin cDNAs from other mammalian species. Recently, however, novel N-terminal regions (designated XL region in domain L) translatable from the upstream ATGs have been found in longer cDNA clones of the bovine and mouse inhibitors (*23*, *24*). The uttermost N-terminal region of the mouse sequence was, however, completely different from that of the bovine sequence (*24*). Surprisingly, an open reading frame of a human EST clone agreed well with the mouse sequence (*24*). Thus, the divergence of the N-terminal sequences between mouse/human and bovine calpastatins seemed not merely due to the species difference, and the discrepant region was predicted to correspond to a yet unidentified exon of the calpastatin gene.

The genomic DNA structures of the calpastatin genes

[1] This work was supported by a Grant-in-Aid for Scientific Research on Priority Areas, No 08278102, from the Ministry of Education, Science, Sports and Culture of Japan (to M.M.).

[2] To whom correspondence should be addressed Phone +81-52-789-4088, Fax. +81-52-789-5542, E-mail: mmaki@agr.nagoya-u.ac.jp

Abbreviations: Acc. No., accession numbers of the GenBank/EMBL/DDBJ databases, BAC, bacterial artificial chromosome, bp, base pairs; EST, expressed sequence tags; GAPDH, glyceraldehyde 3-phosphate dehydrogenase; nt, nucleotide, PAGE, polyacrylamide gel electrophoresis, RT-PCR, reverse-transcription and polymerase chain reaction.

have still been only partially analysed (*23*, *25*). To clarify the origin of the molecular diversity in the reported N-terminal sequences of calpastatins among mammals, we investigated the entire exon-intron organization of the mouse calpastatin gene and analysed the calpastatin mRNA isoforms by RT-PCR. Here we show that mouse differentially expresses four calpastatin isoforms with distinct N-terminal sequences which are specified by alternative first exons.

EXPERIMENTAL PROCEDURES

Reagents—Restriction endonucleases, *Ex Taq* DNA polymerase, a kit for DNA ligation, λ DNA and a DNA size marker of ΦX174 DNA/*Hin*cII digests and most other molecular biological reagents were purchased from Takara (Kyoto). Other reagents were purchased from Nacalai Tesque (Kyoto) or Wako Pure Chemicals (Osaka).

Isolation of the Genomic DNA Clone—A genomic DNA library of mouse strain 129/SVJ, cloned into a bacterial artificial chromosome (BAC) vector (*26*), was custom-screened by the PCR method using a pair of mouse calpastatin-specific primers (5′-AGTCAGGGGGAGGGTCTTGT-3′ and 5′-GGGAAAACGCAGCGAAAT-3′) at Genome System (St. Louis, Missouri, USA). One positive clone was provided, and BAC plasmid DNA was isolated essentially according to the supplied instruction manual which was based on the alkaline/SDS method (*27*). Contaminating RNA and other impurities were removed by the gel filtration method using Sepharose CL-2B (Amersham Pharmacia Biotech, Uppsala, Sweden) in 10 mM Tris-HCl, pH 8.0, 1 mM EDTA, and 0.15 M NaCl.

DNA Sequencing—Nucleotide sequences were determined with an automated fluorescent sequencer, ABI PRISM 310 (PE Applied Biosystems), using a BigDye™ terminator cycle sequencing ready reaction kit (PE Applied Biosystems). Oligonucleotide primers were custom-synthesized by Rikaken (Nagoya).

RT-PCR—Total RNA was isolated from mouse tissues by the acid guanidine thiocyanate phenol/chloroform method (*28*). Reverse transcription was carried out in 20 μl of the following reaction mixture: 50 mM Tris-HCl (pH 8.3), 100 mM KCl, 10 mM $MgCl_2$, 10 mM dithiothreitol, 0.5 mM each dNTP, 4 units of Rous associated virus 2 (RAV-2) reverse transcriptase (Takara), 40 μg/ml RNA, 10 μg/ml oligo-$dT_{12\text{-}18}$, and 40 units of RNasin. The reaction mixtures were incubated at 42°C for 30 min and 50°C for 30 min, then heated to 95°C for 5 min and centrifuged at 14,000 rpm for 5 min. The resultant supernatants were used as cDNA templates for PCR DNA amplification. PCR was performed with a programmed thermal cycler GeneAmp PCR System 2400 (PERKIN-ELMER) using 0.63 unit of *Ex Taq* in a 25-μl reaction mixture containing the supplied buffer, 100 μM each dNTP, 0.2–2 μl of the reverse-transcribed products, and 1 μM each of primer sets unless otherwise stated. Expression of the glyceraldehyde 3-phosphate dehydrogenase (GAPDH) mRNA was monitored as a ubiquitously expressing gene using a pair of primers encoding nt residue No. 565–586 (upper primer) and nt residue No 997–1017 (lower primer), respectively, of the mouse GAPDH cDNA sequence (Acc. No. M32599).

PCR products (5 μl) were electrophoresed on either 5% polyacrylamide gels or 1% agarose gels. Gels were stained with ethidium bromide (1 μg/ml), then DNA fragments were visualized under shortwave (254 nm) UV illumination and photographed. For direct sequencing of the short DNA PCR products, DNA fragments were excised from the polyacrylamide gels under longwave (365 nm) UV illumination and eluted with 0.5 M ammonium acetate, 10 mM magnesium acetate, 1 mM EDTA, and 0.1% SDS after crushing the gels. For recovery of the long DNA fragments, agarose gels were crushed and treated with phenol followed by phenol/chloroform extraction and precipitation with ethanol as described elsewhere.

RESULTS

Calpastatin Isoforms—As shown in Fig. 1, we classified four types of mammalian calpastatins with distinct N-terminal sequences which were deduced from the plausible full-length cDNAs reported to date in published journals or DNA data banks (GenBank/EMBL/DDBJ): Type I, mouse myoblast (*24*); Type II, bovine heart (*23*); Type III, rabbit lung and heart, composite (*17*), pig heart (*18*), human heart and liver, composite (*19*), bovine skeletal muscle (*21*), rat liver (*20*) and brain (*22*), sheep heart (Acc. No. U66320); Type IV, human testis (Acc. No. U58996). Isoforms with deletions in the internal regions generated by alternative splicing have also been reported (*22*, *24*, *25*, *29*), but they have either Type I or Type III N-terminal sequences (not shown).

Exon-Intron Organization of the Mouse Calpastatin Gene—Exon-intron junctions of the mouse calpastatin gene were determined by nucleotide sequencing of the isolated BAC genomic DNA clone using calpastatin-specific oligonucleotide primers. The lengths of introns were estimated either by sequencing of the entire region or by PCR of the BAC DNA followed by agarose gel electrophoresis. The previously isolated mouse calpastatin cDNA (Type I) is encoded by 31 exons from the first exon (designated exon 1xa)

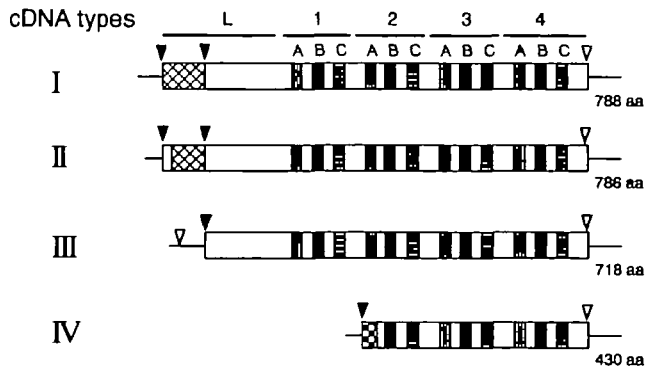


Fig 1 **Schematic structures of the four types of mammalian calpastatins with distinct N-terminal sequences.** Amino acid sequences are deduced from the isolated cDNAs (Acc. Nos. registered in GenBank are indicated in the parentheses): Type I, from mouse myoblast (AB026997); Type II, from bovine heart (AF-159246), Type III, from rabbit lung and heart, composite (M16476); Type IV, from human testis (U58996). Three highly conserved regions within each repetitive domain are marked with vertical stripes (subdomains A), closed boxes (subdomains B), and horizontal stripes (subdomains C), respectively A novel N-terminal sequence in domain L (XL region) is cross-hatched Translation initiation codons and stop codons are indicated by closed and open arrowheads, respectively

to the last exon (designated exon 29) dispersed in the range of about 60 kb (Fig. 2). The positions of the introns are well conserved among the four calpain inhibitory domains, and each repetitive domain is divided into 5 exons (Table I). Domain 4 has an additional intron in the C-terminal region. All exon-intron boundaries conform to the GT-AG rule (*30*).

Exons 1xa and 1xb—Cong *et al.* (*23*) reported a partial DNA sequence of the bovine calpastatin gene around the promoter region including the first exon. The corresponding region of the mouse calpastatin gene was compared with the bovine sequence. As shown in Fig. 3A, a significant nucleotide sequence similarity was observed between the N-terminus–encoding exon (Type I) of the mouse calpastatin gene (exon 1xa) and the region 0.8 kb upstream of the transcription initiation site of the bovine calpastatin gene (–812 nt to –744 nt). Alignment of the nucleotide sequences revealed high conservation of the sequence, particularly in the protein-coding region (identity of 22 out of 25 amino acid residues) (Fig. 3B, upper panel). On the other hand, a sequence similar to the bovine N-terminal region (Type II) was not readily found by the dot matrix method. A careful DNA sequence comparison, however, revealed a significant similarity in a short segment of the mouse genomic DNA sequence downstream of exon 1xa, and the predicted amino acid sequence was completely conserved (Fig. 3B, lower panel). RT-PCR of the Type II mRNA from mouse heart produced a specific DNA band in PAGE as described below (Fig. 5B, 264 bp), and direct sequencing of the amplified DNA revealed a splicing donor site of exon 1xb.

Identification of a cDNA Containing Exon 1u—To establish that the Type III mRNA isoform is also present in mouse tissues, semi-nested RT-PCR was performed using a fixed upper primer hCSexUS2, which was a consensus sequence derived from the rabbit and pig calpastatin cDNAs as well as a human EST AA330457 cDNA clone (5′-CTAGGAATGCAGACCTCC-3′, underlined residues are different from the mouse sequence which was determined later), and a lower primer designated 853L22[3], derived from a complementary sequence of the mouse cDNA (Fig. 4). An aliquot was taken from the first reaction mixture and used for the following reaction using an inner lower primer (677L17). The nested PCR was repeated using fur-

[3] When the primers used for RT-PCR analyses are derived from the mouse calpastatin cDNA sequence registered in GenBank/EMBL/DDBJ (Acc. No. AB026997), the individual sequences are not described but expressed as follows: nucleotide residue No., upper (U, sense) or lower (L, antisense), length of the primer. For instance, 883L22 indicates a 22-mer lower primer (nt residue No. 883–904).

TABLE I. **Summary of the splicing junctions in the mouse calpastatin gene.** The first exon of each type of calpastatins (1xa, 1xb, 1u, and 14t) lacks the splicing acceptor site, whereas the common last exon, exon 29, lacks the splicing donor site. The conserved dinucleotides in the intron sequences, ag and gt at the splicing acceptor sites (SA) and the splicing donor sites (SD), respectively, are represented in boldface letters. The lengths of the introns were estimated either by PCR of the isolated BAC clone or by nucleotide sequencing (indicated by asterisks). Values in the parentheses are uncertain due to appearance of nonspecific bands. The partial calpastatin genomic DNA sequences are registered in DDBJ (Acc. Nos. AB044310–AB044334).

Calpastatin Type	Exon number	Splicing acceptor site (SA)		Protein coding length	Splicing donor site (SD)		Intron length
				bp			kb
Type I	1xa	---	-	75	GAG	gtagaggtag	4.3
	1y	atatttttag	CAT	63	GAG	gtaatttctg	3.8
domain	1x	tgtttctaag	AAA	72	AAG	gtcagtcaat	3.4
	2	tcctttctag	GTG	60	AAG	gtaaagaaac	3.9
L	3	cttctattag	GCA	57	CAG	gtgatgtagt	*0.6
	4	cccactttag	GCT	39	GAG	gtaagtaggt	1.5
	5	gaaattacag	CAA	54	GAG	gtgagtgata	*0.5
	6	tgcacgtcag	CCC	114	AAG	gtgagcttga	1.3
	7	ctgctctcag	ACA	90	AAG	gtgagcatac	3.4
	8	ctccctctag	AAA	69	AAG	gtatggagcc	1.2
	9	ctatttctag	TCA	99	TTG	gtatgtgttg	1.5
	10	tttcttccag	GAC	81	GAA	gtaagttatc	*0.6
1	11	tttggttcag	AAA	39	CCT	gtgagttgaa	0.8
	12	ttttctctag	AAA	87	GAG	gtattgattt	*0.1
	13	tctttttag	AAG	90	GAG	gtacttaaat	1.6
	14	tttttaacag	GAG	102	GAG	gtaatgactt	3.3
	15	ttttcatcag	GCA	84	AAG	gtaagcgctc	1.3
2	16	ttatttacag	GAT	48	AAG	gttggagaat	2.0
	17	cttcttacag	AGC	90	AAG	gtagaagtat	1.2
	18	cttttctcag	AAA	96	TTG	gtgagtgatg	1.8
	19	gctcttgcag	AAG	102	AAG	gtaatggcag	*0.5
	20	ttttatttag	GAG	84	AAG	gtaacaggct	3.2
3	21	ctctgaatag	GAT	51	GCA	gtaagcaatc	*0.2
	22	cttcttgcag	CCC	72	CTT	gtaagtatac	(6.9)
	23	ttaataaaag	GAA	99	GAG	gtaagcaaca	(3.0)
	24	taccttttag	CAG	105	AAG	gtaaaagatc	1.0
	25	atgtttgtag	GAG	93	AAG	gtaaactaat	1.0
4	26	tgtttttaag	GAC	36	GAG	gtaaatacta	*0.5
	27	acatgttaag	AAA	93	AAG	gtacctacac	2.0
	28	attgctttag	GGG	69	AAG	gtaaacatgg	3.5
	29	tcccccttag	AAG	51	-	---	-
Type II	1xb	---	-	30	ACG	gtaaatagaa	3.7
Type III	1u	---	-	-	TGA	gtgagtactg	1.1
Type IV	14t-L	---	-	165	GAG	gtgcacatta	2.1
	14t-S	---	-	120	CAT	gtaatatgct	2.1

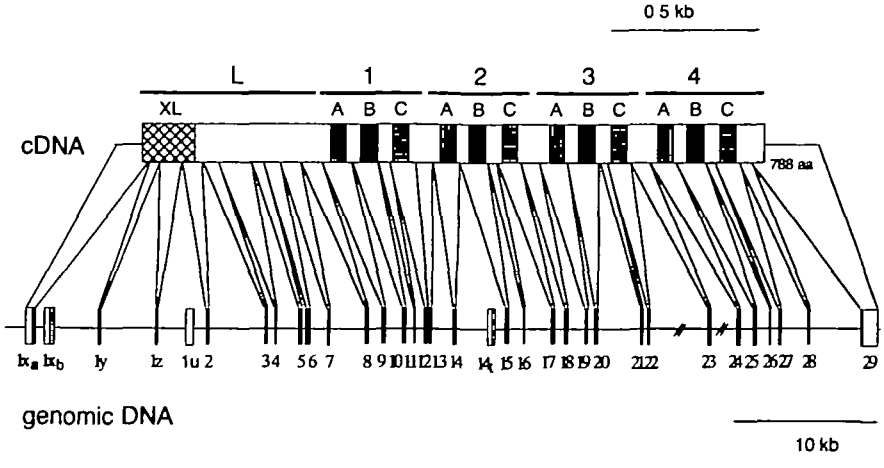
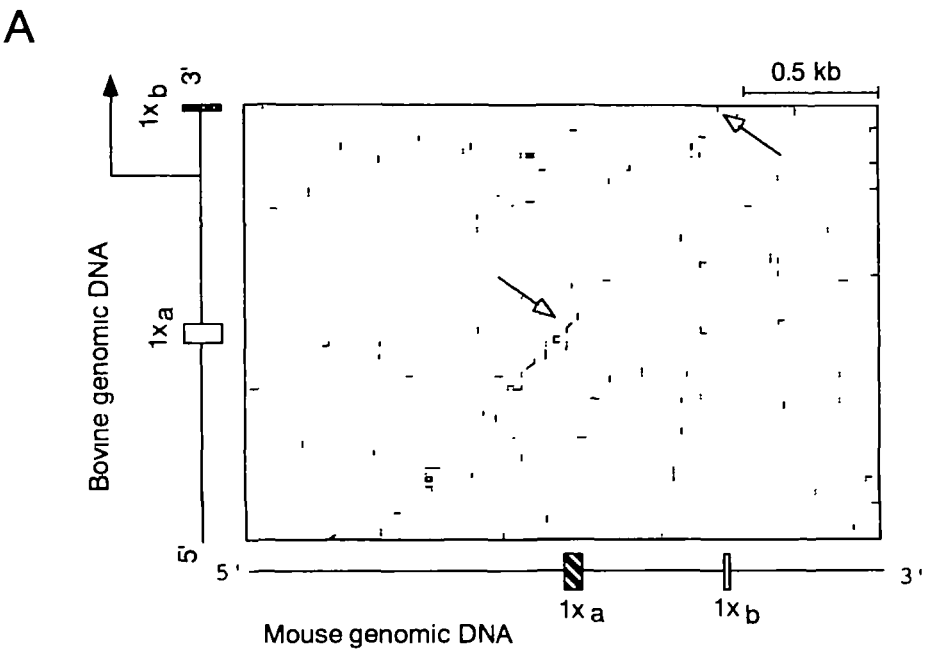


Fig 2. **Exon–intron organization of the mouse calpastatin gene.** Exons, not drawn to the scale, in the mouse calpastatin gene are indicated by boxes, where filled areas represent translated regions. Double slashes indicate introns whose lengths are not certain.



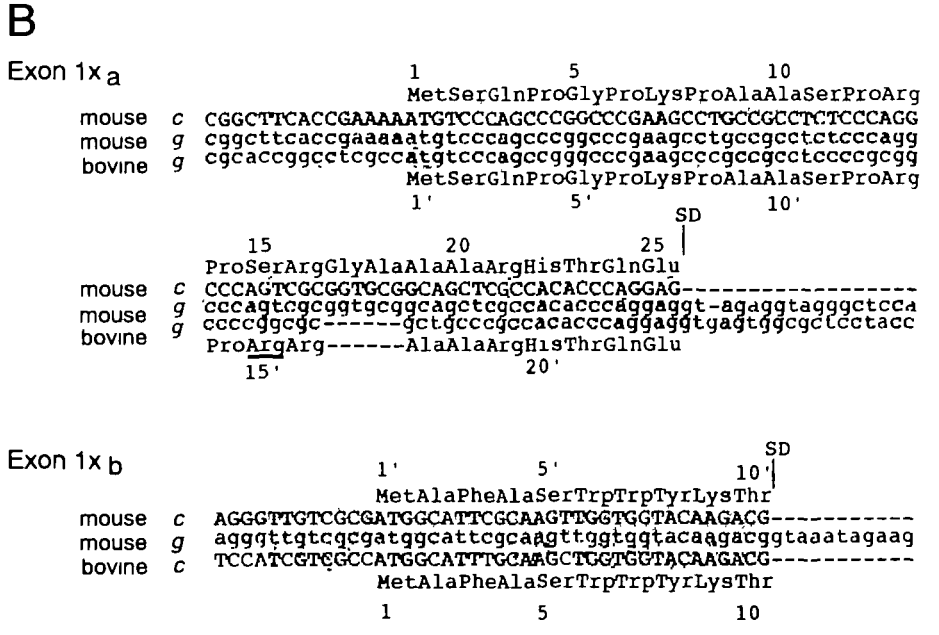


Fig. 3. **Comparison of the DNA sequences corresponding to exons 1xa and 1xb between the mouse and bovine calpastatin genes.** A. Dot matrix analysis of the mouse and bovine genomic DNA sequences containing exons 1xa and 1xb. A dot indicates a segment where 7 out of 9 nucleotide residues are identical between the two species. Open arrows indicate regions corresponding to exons 1xa and 1xb. B. Alignment of the nucleotide and amino acid sequences of regions containing exons 1xa and 1xb. Genomic and cDNA sequences are indicated by g and c, respectively. Nucleotide residues of the genomic DNA sequences are expressed by lowercase letters. Identical nucleotide residues in the mouse and bovine calpastatin genes are stippled. Amino acid substitution in the predicted exon 1xa of the bovine calpastatin gene is underlined. SD indicates a splicing donor site.

ther inner primers (589L23 and 494L18) (Fig. 4A). Direct sequencing of a 100-bp DNA fragment from the final RT-PCR product of the 4th reaction (closed arrowhead) revealed a sequence similar to those for the human, pig, and rabbit cDNAs (Fig. 4B). Then, specific primers were designed for the genomic DNA sequencing. Comparison of the nucleotide sequences of the obtained cDNA and the corresponding region (designated exon 1u) of the mouse genomic DNA revealed a splicing junction. An in-frame upstream stop codon was also found in the mouse sequence. The faint band of about 140 bp in the 3rd RT-PCR reaction (Fig. 4A, open arrowhead) agreed with the expected size of the RT-PCR product derived from the calpastatin mRNA of an alternatively spliced exon-3 deletion form, which is predominant in mouse heart (*24*).

RT-PCR Analysis of the Calpastatin mRNAs for Types I, II, and III—RT-PCR was performed to examine the difference of the expression levels among the calpastatin isoforms in various mouse tissues including heart, skeletal muscle, liver, brain, and testis (Fig. 5). Compared with the similar expression levels of the glyceraldehyde 3-phosphate dehydrogenase (GAPDH) mRNA (panel D) among tissues, those of the Type I and Type III mRNAs in liver were higher than in other tissues examined (panels A and C). In contrast, the Type II mRNA was expressed at lower levels in liver, brain, and testis than in heart and skeletal muscle (panel B).

Testis-Specific Isoform—A sequence similar to the N-terminal region of the human testis calpastatin isoform (Type IV) was found between exons 14 and 15 (Fig. 6A). RT-PCR

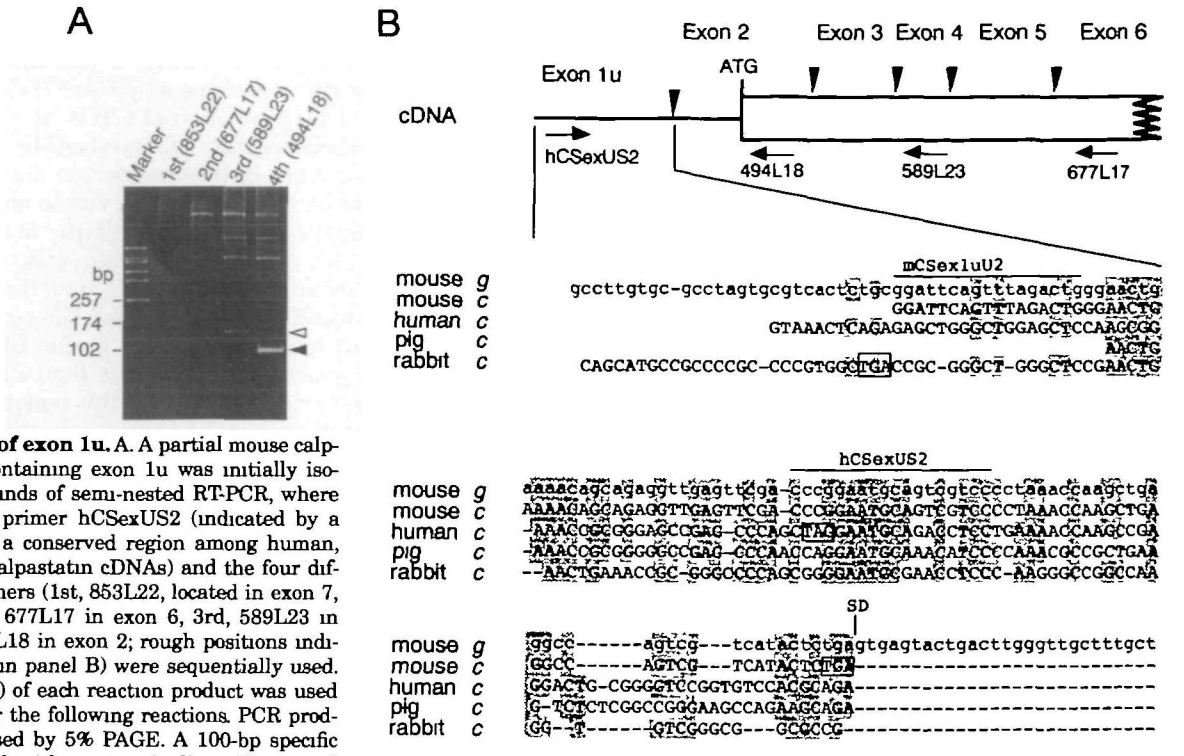


Fig. 4. **Analysis of exon 1u.** A. A partial mouse calpastatin cDNA containing exon 1u was initially isolated by four rounds of semi-nested RT-PCR, where the fixed upper primer hCSexUS2 (indicated by a line in panel B, a conserved region among human, pig, and rabbit calpastatin cDNAs) and the four different lower primers (1st, 853L22, located in exon 7, not shown; 2nd, 677L17 in exon 6, 3rd, 589L23 in exon 4, 4th, 494L18 in exon 2; rough positions indicated by arrows in panel B) were sequentially used. An aliquot (0.1%) of each reaction product was used as a template for the following reactions. PCR products were analysed by 5% PAGE. A 100-bp specific DNA band after the 4th reaction (indicated by arrow) was extracted from the gel and subjected to DNA sequencing. Then, specific primers were designed from the determined cDNA sequence for further determination of the genomic DNA sequence using the BAC plasmid DNA. Marker, pUC119/*Hae*III digests. B: Comparison of the DNA sequences corresponding to the exon 1u regions of the mouse calpastatin gene and mammalian calpastatin cDNAs. Genomic and cDNA sequences are indicated by g and c, respectively. Nucleotide residues of the genomic DNA sequences are expressed by lowercase letters. Identical nucleotide residues in the mouse, human, pig, and rabbit calpastatin cDNAs are stippled. SD indicates a splicing donor site. In-frame upstream stop codons of Type III calpastatins are boxed. Positions of introns are indicated by arrowheads.

was performed to detect the mouse Type IV calpastatin mRNA using an upper primer (14tU3) located in the exon designated 14t and a lower primer (1549L18) located in exon 16. Doublet bands were detected in mouse testis but not in the other tissues examined (Fig. 6B). Direct sequencing of DNA from each excised band revealed a difference in the splicing donor sites. The amino acid sequence deduced from the short form of this exon (14t-S, 40 amino acid residues) was similar to the human sequence registered in Genbank (Acc. No. U58996; 70% residue identity in the corresponding region). Similar sequences for the long form (14t-L, 55 amino acid residues) have not been found in the human EST database. RT-PCR analysis using primers located in exons 12 and 16 gave an expected single band of 350 bp in all tissues including testis, indicating that the conventional types of the calpastatin mRNAs (Types I, II, and III) do not contain exon 14t as alternatively spliced mRNA isoforms (Fig. 6C).

DISCUSSION

We analysed the mouse calpastatin gene and determined the exon–intron junctions. By comparing the mouse genomic DNA with the reported four types of calpastatin cDNAs from mouse and other mammals, we identified 34 exons including 5 exons upstream of exon 2 and 1 more exon between exons 14 and 15 (Fig. 2). Our proposed model to explain the molecular diversity in the N-terminal sequences is summarized in Fig. 7. Exons 1xa, 1xb, 1u, and 14t are the first exons, which are transcribed from different promoters and specify the four types of calpastatins. Exon 1u does not contain an in-frame ATG codon for the calpastatin-coding sequence, but it possesses an in-frame stop codon (Fig. 4). Thus, the transcript from exon 1u encodes Type III calpastatin, which uses a translation initiation codon located in exon 2 as originally reported for the rabbit calpastatin cDNA (*17*).

The calpastatin mRNAs for the isoforms derived from exons 1xa (Type I), 1xb (Type II), and 1u (Type III) were detected by RT-PCR analysis in various mouse tissues regardless of differences in their expression levels (Fig. 5). On the other hand, Type IV was specifically expressed in testis (Fig. 6B). The generation of the molecular diversities by the proposed model may not be unique to mouse. Indeed, a search of the human EST database for calpastatin-like sequences reveals partial cDNAs for Type I (AI017990, fetal lung/testis/B-cell mixed library), Type III (AB026049, heart; AA330457, embryo; AA362843, ovary), and Type IV (AL044822 and AL044436, testis). Moreover, the bovine calpastatin gene has been shown to possess a region corresponding to exon 1xa of the mouse gene, and the predicted amino acid sequence is well conserved with the mouse sequence (Fig. 3).

At present, we do not have direct evidence that a promoter exists immediately upstream of exon 1u. Transcripts containing the sequence of exon 1u driven from either exon

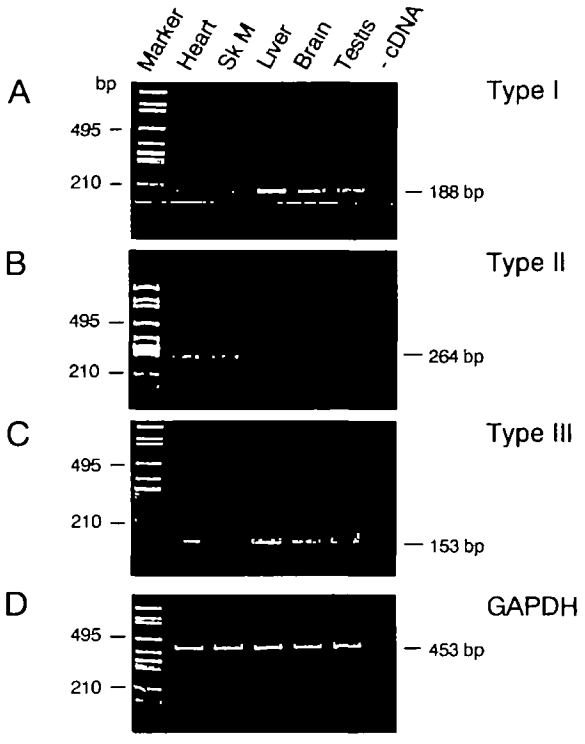


Fig 5. **Detection of the calpastatin mRNA isoforms in the adult mouse tissues by RT-PCR analyses.** Primers were selected to detect the mRNAs for Type I (A), Type II (B), Type III (C) calpastatins or that for GAPDH (D) as a control of ubiquitous expression PCR products were analysed by 5% PAGE PCR without a cDNA template was performed as a negative control Upper primers used are located in exons 1xa (A, 242U18), 1xb (B, 5′-TTCTGCGTTCTC-GAGAGG-3′) and 1u (C, hCSexUS2, Fig 4B), respectively, for calpastatins Lower primers used are located in exon 1z for analysis of the Type I mRNA (413L17) and exon 2 for analyses of the Type II and Type III mRNAs (494L18) PCR conditions used are as follows denaturation at 94°C for 30 s in all cases and (A), annealing at 54°C for 20 s, extension at 72°C for 20 s, 35 cycles; (B), annealing at 54°C for 30 s, extension at 72°C for 30 s, 45 cycles; (C), annealing at 48°C for 30 s, extension at 72°C for 30 s, 36 cycles, (D), annealing at 50°C for 30 s, extension at 72°C for 30 s, 22 cycles, respectively. DNA size marker, ΦX174 *Hin*cII digests, Sk M, skeletal muscle.

1xa or exon 1xb would also generate a Type III calpastatin isoform by using ATG in exon 2 as an internal translation initiation codon. However, the following observations make this unlikely. RT-PCR analysis of the Type II calpastatin mRNA using a pair of primers located in exons 1xb and 2, respectively, gave a single band and the size of the product agreed with the expected size (264 bp) of the amplified DNA consisting of exons 1xb, 1y, 1z, and 2 (Fig. 5B). No extra bands containing exon 1u were detected. The 264-bp PCR product did not contain the exon 1u sequence as further confirmed by nucleotide sequencing (data not shown). These results indicate that exon 1u does not have a functional splicing acceptor site, and that it is skipped upon splicing of the precursor mRNA transcribed from upstream promoters. In other words, exon 1u is the alternative first exon for Type III calpastatin.

Cong *et al.* (*23*) reported cAMP-dependent transactivation of the bovine calpastatin gene whose promoter was located upstream of exon 1 (exon 1xb according to our nomenclature). They identified a sequence GTCA (–76 nt upstream of the transcription initiation site) which was important for the cAMP responsiveness and corresponded to the half site of the full CRE (a consensus palindromic cAMP-responsive *cis*-element: TGACGTCA) (*31*). They demonstrated that mutation of GTCA at –76 nt to ATCT completely abolished the dibutyryl-cAMP mediated promoter enhancement in the transfection assay. Comparison of the nucleotide sequences of the mouse and bovine genomic DNAs did not show a high similarity in the corresponding regions, but a short similar sequence GTGCGGTGT-CAGCCGG (identical residues are underlined) containing GTCA was found. Cong *et al.* also showed that deletion from –994 nt to –671 nt in the bovine calpastatin gene reduced the promoter activity more than 7-fold. Compared with the mouse calpastatin gene, this region may also contain a potential promoter which drives transcription from exon 1xa (Fig. 3). The differential expression patterns of the Types I, II, and III mRNAs among the mouse tissues examined suggest presence of different transcriptional regulatory elements upstream of the respective promoters (Fig. 5).

In contrast to the rather ubiquitous expression of the other types, the Type IV calpastatin mRNA is specifically expressed in testis (Figs. 5 and 6). The regulation of the alternative transcription initiation from exon 14t remains to be clarified. Switching of promoters from somatic sites to testis-specific sites is known in some genes. The transcription factor GATA-1, a fundamental regulator of genes in haematopoietic cell lineages, is also expressed in testis, and the 5′-untranslated region of the testis mRNA is different from that of the erythroid mRNA (*32*). An alternative testis promoter located 5′ to the erythroid promoter is used, and the transcript is alternatively spliced to skip the erythroid first exon. On the other hand, two isozymes of phosphoglycerate kinases (PGK-1, somatic type; PGK-2, testis-specific type) are encoded by different genes (*33*). The PGK-2 gene is suggested to be activated during the spermatogenic pathway through involvement of negative and positive *cis*-elements upstream of the promoter region (*34*, *35*). The nucleotide sequence between exons 14 and 14t of the mouse calpastatin gene does not contain sequences consistent with those *cis*-elements. Potential *cis*-elements for transcriptional regulation of the testis-specific calpastatin promoter should be identifiable by transfection assay with various plasmid DNA constructs containing suitable reporter genes fused with the upstream sequence of exon 14t. Immunohistochemical analyses to identify the cells expressing Type IV calpastatin and screening of suitable testis-derived cultured cells, however, would be prerequisite to such a study on the testis-specific calpastatin promoter.

In our preliminary experiments using a polyclonal anti-calpastatin antibody, two major bands (70 and 110 kDa) were observed by the immunoblot analysis of the mouse testis calpastatins (data not shown). Since tissue type calpastatins are known to migrate abnormally in SDS-PAGE as 100–120 kDa proteins (*18*), the 70-kDa band may represent the testis-specific isoform. The molecular masses of calpastatins observed by the immunoblot analysis, however, should be carefully interpreted. Calpastatins isolated from human, pig, and rabbit erythrocytes also lack domains L and 1, and migrate as 68–70 kDa proteins in SDS-PAGE (*36–38*). Since rabbit erythrocyte calpastatin has been shown to retain subdomain 2A (*39*), the erythrocyte type is

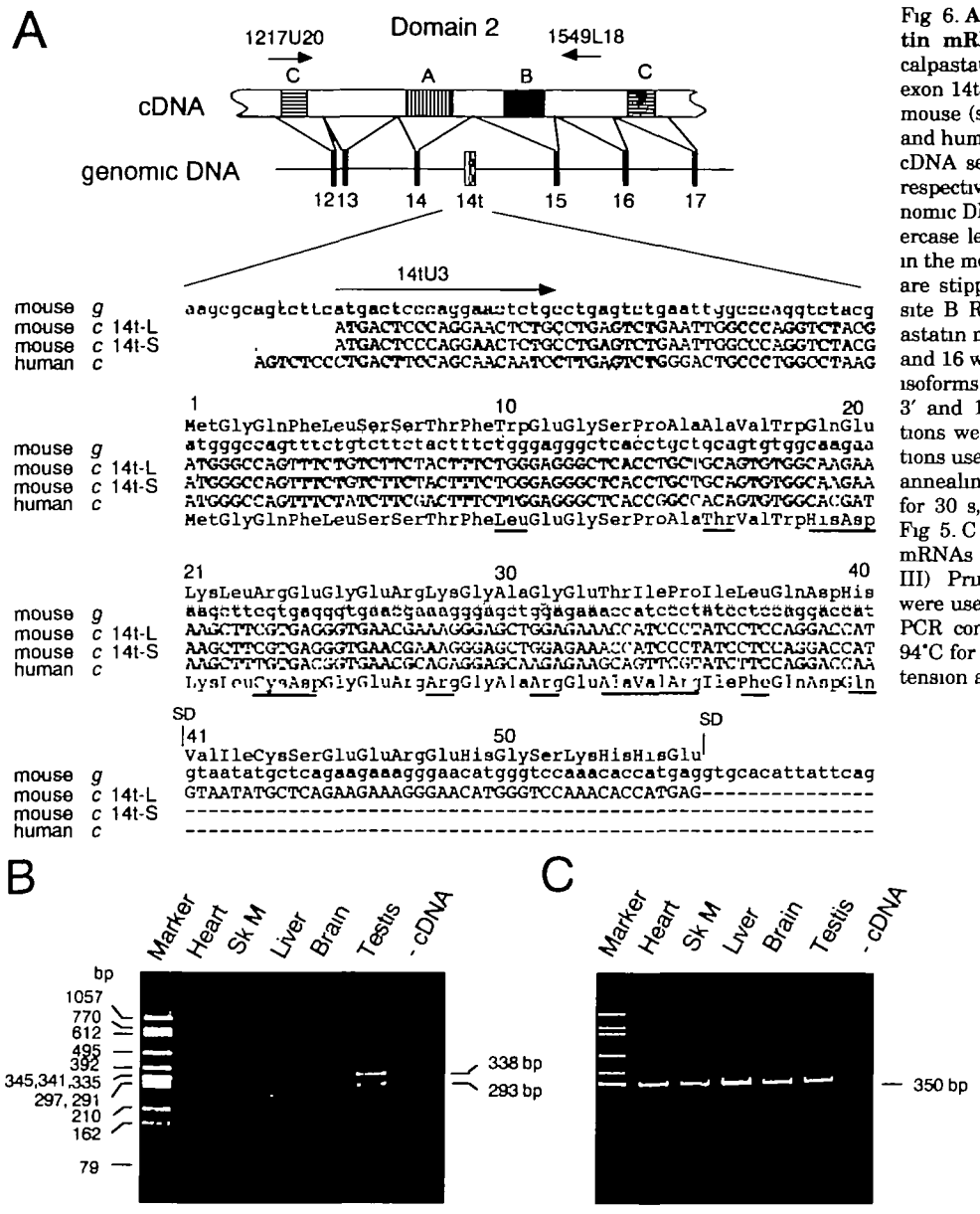


Fig 6. **Analysis of the Type IV calpastatin mRNA.** A. Comparison of the mouse calpastatin genomic DNA sequence in the exon 14t region with the cDNA sequences of mouse (short form, 14t-S; long form, 14t-L) and human testis calpastatins. Genomic and cDNA sequences are indicated by g and c, respectively. Nucleotide residues of the genomic DNA sequences are expressed by lowercase letters. Identical nucleotide residues in the mouse and human calpastatin cDNAs are stippled. SD indicates a splicing donor site. B. RT-PCR analysis of the Type IV calpastatin mRNA. Primers located in exons 14t and 16 were used to detect the testis specific isoforms (5′-ATGACTCCCAGGAACTCTGC-3′ and 1549L18) where primer concentrations were reduced to 0.25 μM. PCR conditions used are denaturation at 94°C for 30 s, annealing at 52°C for 30 s, extension at 72°C for 30 s, 29 cycles. Lanes are as shown in Fig. 5. C. RT-PCR analysis of the calpastatin mRNAs without exon 14t (Types I, II, and III). Primers located in exons 12 and 16 were used (1217U18 and 1549L18) at 1 μM. PCR conditions used are: denaturation at 94°C for 30 s, annealing at 54°C for 30 s, extension at 72°C for 30 s, 40 cycles.

different from the testis-specific type (Type IV). Interestingly, rat erythrocyte calpastatin gave exclusively a 100-kDa band regardless of the similar preparation of the sample (*38*). The unheated cytosol from human erythrocytes also gave higher molecular weight bands of calpastatins (*40*). Thus, the erythrocyte type must be generated by a different mechanism, most probably by post-translational processing. It remains to be established in what kind of testis cells the specific isoform (Type IV) is expressed. Although Rojas *et al.* demonstrated a 68-kDa calpastatin protein in the cytosolic fraction of human spermatozoa by immunoblot analysis (*41*), their and also our immunoblot analyses using polyclonal calpastatin antibodies may not have distinguished the testis-specific isoforms from proteolytically processed calpastatin fragments. Immunohistochemical studies using specific antibodies against the N-terminal region encoded by exon 14t would solve the problem.

Differences in physiological functions among the calpastatin isoforms remain to be established. Interestingly, the N-terminal sequence following the potential translation initiation methionine of Type IV calpastatin (2-GQFLS) agrees well with the consensus sequence for myristoylation at Gly (*42*). This modification may allow the isoform to change its intracellular localization to membranes. Calpains are known to be activated on membranes (*43*), and the calpain-calpastatin system has been implicated in calcium-dependent membrane fusion events such as myotube formation (*2*, *44*, *45*) and erythrocyte ghost fusion (*46*). Type IV calpastatin, if present in sperm, might be related to the regulation of acrosome membrane fusion or sperm-egg fusion. Recently, De Tullio *et al.* reported changes in intracellular localization of calpastatin from aggregated structures near nuclear invaginations to widespread distribution in the cytosol during calpain activation induced with calcium ionophore (*47*). The N-terminal regions encoded by exons 1xa and 1y contain potential phosphorylation sites

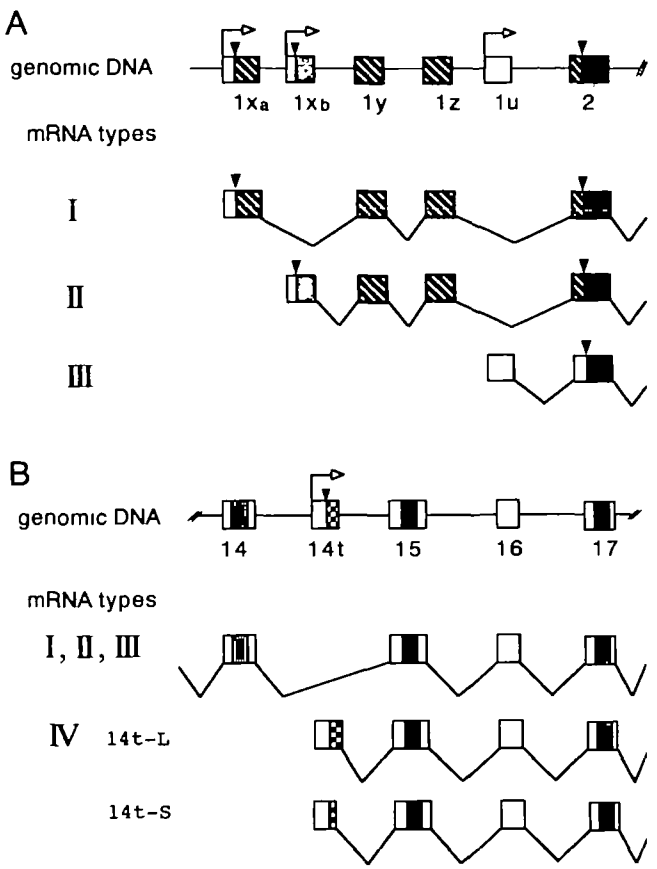


Fig. 7. **Mechanism of the mouse calpastatin isoform generation.** Closed arrowheads and open arrows indicate translation initiation sites and potential transcription initiation sites, respectively Molecular diversities are generated by alternative transcription initiation and exon skipping. A: Generation of the isoforms upstream of exon 2 (Types I, II, and III) B Generation of the testis-specific isoform (Type IV) Alternative splicing causes generation of the long form (14t-L) and the short form (14t-S) of the Type IV isoform in testis.

by protein kinase C (11-SPR, Type I calpastatin) and protein kinase A (41- or 26-KKGS, Type I or II calpastatin), respectively, which are conserved among mouse, human, and bovine calpastatins (Fig. 3 and Ref. *24*). Differences in the N-terminal sequences might affect intracellular distribution of the inhibitors and efficiency of calpain–calpastatin interactions in stimulated cells.

Previously we determined exon-intron junctions in domain 1 of the human calpastatin gene, but those in other inhibitory domains remained unknown (*25*). As we show for the first time in this report, positions of introns in all inhibitory domains are well conserved in the mouse calpastatin gene (Fig. 2 and Table I). Each inhibitory domain possesses 5 exons, except that domain 4 has 1 additional exon at the C-terminal region. The repetitive domain structure accounts for the large number of exons per gene (31 exons, Type I isoform). The average length of the protein coding sequence is 76.3 nt residues/exon or 25.4 amino acid residues/exon. This number is much smaller than the mean exon size (266 bp; median 135 bp) reported for the recently identified 679 genes across the long arm of human chromosome 22, and even smaller than that for catalytic α-subunit of phosphatidyl-inositol-4 kinase (37.9 amino acid residues/exons), which has the largest number of exons per gene (54 exons) in the chromosome (*48*). Thus, the protein coding sequence of the calpastatin gene appears to be split by introns far more frequently than other genes.

Alternative promoter usage has been observed for many genes, some of which contain 5′-untranslated alternative exons. The presence of multiple promoters has been implicated in regulation of gene expression (*49*). (i) The level of transcription initiation can vary between alternative promoters. (ii) Switch to mRNA isoforms with different 5′-exons leads to alteration of turnover rate or translation efficiency. (iii) Alternative promoters can have different tissue specificity or developmental stage specificity, and respond differently to signals. (iv) If the alternative 5′-exons contain translatable sequences in-frame to the downstream coding sequences, polypeptidic isoforms can be generated from a single gene as shown in this work. The number of alternative promoters varies among the genes reported. The rat BDNF (brain-derived neurotrophic factor) gene and the rat γ-glutamyl transpeptidase gene have four and five differentially regulated promoters, respectively (*50*, *51*). In these cases, however, each promoter produces an mRNA containing a unique 5′-untranslated exon, and there is no difference in the product translated from the mRNA isoforms. The calpastatin gene appears to be unique in that it possesses four alternative promoters producing polypeptidic isoforms.

The locus of the human calpastatin gene is assigned to 5q14-21 (*52*). No hereditary diseases directly associated with mutations of this gene have been so far reported. Since all four repetitive domains possess calpain inhibitory activities (*53*, *54*), partial deletion of the gene may not critically affect the function of the inhibitor. In addition to a mechanism to generate multiple isoforms of a protein from the single gene, the presence of multiple promoters may also assure the expression of the gene even if one of them is rendered non-functional by mutation. In other words, redundancy of the inhibitory domains and promoters may be a strategy adopted during evolution to counteract critical mutations during embryonic development of animals. A gross deletion of the calpastatin genes in both alleles might have ultimately lethal effects on embryos. Future studies of the gene targeting of mice by homologous recombination techniques should settle this argument.

We thank T. Kawamura for her technical assistance in the early stage of this work.

REFERENCES

1. Murachi, T. (1983) Calpain and calpastatin. *Trends Biochem. Sci.* **8**, 167–169
2. Balcerzak, D., Cottin, P., Poussard, S., Cucuron, A., Brustis, J.J., and Ducastaing, A. (1998) Calpastatin-modulation of m-calpain activity is required for myoblast fusion. *Eur. J Cell Biol* **75**, 247–253
3. Patel, Y.M. and Lane, M D. (1999) Role of calpain in adipocyte differentiation. *Proc. Natl. Acad. Sci. USA* **96**, 1279–1284
4. Potter, D.A., Tirnauer, J.S., Janssen, R, Croall, D E., Hughes, C.N, Fiacco, K.A., Mier, J.W., Maki, M., and Herman, I.M. (1998) Calpain regulates actin remodeling during cell spreading. *J. Cell Biol.* **141**, 647–662
5. Jang, J.S, Lee, S J, Choi, Y.H, Nguyen, P.M., Lee, J, Hwang, S.G, Wu, M.L., Takano, E, Maki, M., Henkart, P.A., and Trepel, J.B. (1999) Post translational regulation of the retinoblas-

toma gene family member p107 by calpain protease. *Oncogene* **14**, 255–263
6 Chen, F., Lu, Y, Kuhn, D.C, Maki, M, Shi, X., Sun, S C., and Demers, L.M. (1997) Calpain contributes to silica-induced IκB-α degradation and nuclear factor-κB activation *Arch. Biochem Biophys.* **342**, 383–388
7 Lee, K.S, Frank, S., Vanderklish, P, Arai, A., and Lynch, G (1991) Inhibition of proteolysis protects hippocampal neurons from ischemia. *Proc. Natl. Acad. Sci. USA* **88**, 7233–7237
8 Chi, X.J, Hiwasa, T., Maki, M, Sugaya, S., Nomura, J., Kita, K., and Suzuki, N. (1999) Suppression of okadaic acid-induced apoptosis by overexpression of calpastatin in human UVr-1 cells. *FEBS Lett.* **459**, 391–394
9 Shearer, T R. and David, L L (1990) Calpains in lens and cataract in *Intracellular Calcium-Dependent Proteolysis* (Mellgren, R.L and Murachi, T, eds) pp 265–274, CRC Press, Boca Raton, FL
10. Ménard, H. and El-amine, M (1996) The calpain-calpastatin system in rheumatoid arthritis. *Trends Immunol Today* **17**, 545–547
11 Croall, D E. and DeMartino, G N (1991) Calcium-activated neutral protease (calpain) system: Structure, function, and regulation *Physiol. Rev* **71**, 813–847
12. Sorimachi, H, Ishiura, S, and Suzuki, K. (1997) Structure and physiological function of calpains. *Biochem. J.* **328**, 721–732
13 Takano, E. and Maki, M (1999) Structure of calpastatin and its inhibitory control of calpain in *Calpain· Pharmacology and Toxicology of Calcium-Dependent Protease* (Wang, K.K.W and Yuen, P.-W, eds) pp. 25–49, Taylor & Francis, Philadelphia, PA
14 Ma, H, Yang, H Q, Takano, E., Hatanaka, M, and Maki, M. (1994) Amino-terminal conserved region in proteinase inhibitor domain of calpastatin potentiates its calpain inhibitory activity by interacting with calmodulin-like domain of the proteinase. *J. Biol. Chem.* **269**, 24430–24436
15. Yang, H Q, Ma, H., Takano, E, Hatanaka, M, and Maki, M. (1994) Analysis of calcium-dependent interaction between amino-terminal conserved region of calpastatin functional domain and calmodulin-like domain of μ-calpain large subunit *J Biol. Chem.* **269**, 18977–18984
16. Takano, E, Ma, H, Yang, H.Q., Maki, M., and Hatanaka, M (1995) Preference of calcium-dependent interactions between calmodulin-like domains of calpain and calpastatin subdomains *FEBS Lett* **362**, 93–97
17. Emori, Y., Kawasaki, H, Imajoh, S., Minami, Y, and Suzuki, K. (1987) Endogenous inhibitor for calcium-dependent cysteine protease contains four internal repeats that could be responsible for its multiple reactive sites. *Proc. Natl Acad Sci USA* **84**, 3590–3594
18 Takano, E., Maki, M., Mori, H, Hatanaka, M, Marti, T., Titani, K., Kannagi, R., Ooi, T., and Murachi, T. (1988) Pig heart calpastatin identification of repetitive domain structures and anomalous behavior in polyacrylamide gel electrophoresis. *Biochemistry* **27**, 1964–1972
19. Asada, K., Ishino, Y, Shimada, M, Shimojo, T., Endo, M, Kimizuka, F, Kato, I, Maki, M., Hatanaka, M, and Murachi, T (1989) cDNA cloning of human calpastatin: sequence homology among human, pig, and rabbit calpastatins. *J Enzyme Inhib.* **3**, 49–56
20. Ishida, S, Emori, Y., and Suzuki, K. (1991) Rat calpastatin has diverged primary sequence from other mammalian calpastatins but retains functionally important sequences. *Biochim. Biophys. Acta* **1088**, 436–438
21. Killefer, J. and Koohmaraie, M (1994) Bovine skeletal muscle calpastatin· cloning, sequence analysis, and steady-state mRNA expression. *J Anim. Sci.* **72**, 606–614
22 Tullio, R.D., Sparatore, B, Salamino, F., Melloni, E., and Pontremoli, S (1998) Rat brain contains multiple mRNAs for calpastatin. *FEBS Lett.* **422**, 113–117
23. Cong, M., Thompson, V.F., Goll, D E, and Antin, P.B. (1998) The bovine calpastatin gene promoter and a new N-terminal region of the protein are targets for cAMP-dependent protein kinase activity. *J. Biol Chem* **273**, 660–666
24. Takano, J, Kawamura, T, Murase, M., Hitomi, K., and Maki, M. (1999) Structure of mouse calpastatin isoforms. implications of species-common and species-specific alternative splicing. *Biochem Biophys. Res. Commun.* **260**, 339–345
25 Lee, W.J., Ma, H., Takano, E, Yang, H.Q., Hatanaka, M, and Maki, M (1992) Molecular diversity in amino-terminal domains of human calpastatin by exon skipping. *J Biol. Chem.* **267**, 8437–8442
26. Shizuya, H., Birren, B., Kim, U.J., Mancino, V, Slepak, T, Tachiiri, Y, and Simon, M (1992) Cloning and stable maintenance of 300-kilobase-pair fragments of human DNA in *Escherichia coli* using an F-factor-based vector *Proc. Natl. Acad. Sci. USA* **89**, 8794–8797
27 Birnboim, H C and Doly, J (1979) A rapid alkaline extraction procedure for screening recombinant plasmid DNA. *Nucleic Acids Res.* **7**, 1513–1523
28. Chomczynski, P. and Sacchi, N (1986) Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction *Anal Biochem.* **162**, 156–159
29 Lee, W.J., Hatanaka, M., and Maki, M, (1992) Multiple forms of rat calpastatin cDNA in the coding region of functionally unknown amino-terminal domain. *Biochim. Biophys. Acta* **1129**, 251–253
30. Shapiro, M B and Senapathy, P (1987) RNA splice junctions of different classes of eukaryotes sequence statistics and functional implications in gene expression. *Nucleic Acids Res.* **15**, 7155–7174
31. Cong, M., Goll, D.E, and Antin, P.B. (1998) cAMP responsiveness of the bovine calpastatin gene promoter. *Biochim. Biophys. Acta* **1443**, 186–192
32. Ito, E, Toki, T, Ishihara, H, Ohtani, H., Gu, L, Yokoyama, M, Engel, J D, and Yamamoto, M. (1993) Erythroid transcription factor GATA-1 is abundantly transcribed in mouse testis. *Nature* **362**, 466–468
33 McCarrey, J R., Berg, W M, Paragioudakis, S J, Zhang, P L., Dilworth, D D, Arnold, B L., and Rossi, J.J. (1992) Differential transcription of Pgk genes during spermatogenesis in the mouse *Dev Biol* **154**, 160–168
34 Goto, M., Tamura, T., Mikoshiba, K., Masamune, Y., and Nakanishi, Y (1991) Transcription inhibition of the somatic-type phosphoglycerate kinase 1 gene in vitro by a testis-specific factor that recognizes a sequence similar to the binding site for Ets oncoproteins. *Nucleic Acids Res.* **19**, 3959–3963
35. Nishiyama, S, Masamune, Y, and Nakanishi, Y (1994) The silencer of mouse Pgk-2 consists of two discrete DNA elements that individually have no effect *Gene* **141**, 261–266
36 Takano, E., Kitahara, A., Sasaki, T., Kannagi, R, and Murachi, T. (1986) Two different molecular species of pig calpastatin Structural and functional relationship between 107 kDa and 68 kDa molecules. *Biochem. J* **235**, 97–102
37. Yokota, H., Katayama, M., Hino, F., Kato, I., Takano, E., Maki, M., Hatanaka, M, and Murachi, T (1991) Direct measurement of calpastatin subtypes by sandwich enzyme immunoassay using monoclonal antibodies. *Mol Cell. Probes* **5**, 261–269
38. Inomata, M, Nakamura, M., Imajoh-Ohmi, S., and Kawashima, S. (1993) A variety of calpain/calpastatin systems in mammalian erythrocytes. *Biochim. Biophys. Acta* **1178**, 207–214
39. Imajoh, S., Kawasaki, H, Emori, Y., and Suzuki, K. (1987) Calcium-activated neutral protease inhibitor from rabbit erythrocytes lacks the N-terminal region of the liver inhibitor but retains three inhibitor units. *Biochem. Biophys. Res. Commun.* **146**, 630–637
40 Scharz-Benmeir, N., Glaser, T., Barnoy, S., and Kosower, N.S. (1994) Calpastatin in erythrocytes of young and old individuals. *Biochem. J* **304**, 365–370
41. Rojas, F.J., Brush, M., and Moretti-Rojas, I. (1999) Calpain-calpastatin: a novel, complete calcium-dependent protease system in human spermatozoa. *Mol. Human Reproduct* **5**, 520–526
42. Towler, D.A., Gordon, J.I., Adams, S.P., and Glaser, L. (1988) The biology and enzymology of eukaryotic protein acylation.

Annu. Rev. Biochem. **57**, 69–99
43. Mellgren, R.L. (1987) Calcium-dependent proteases: an enzyme system active at cellular membranes?. *FASEB J.* **1**, 110–115
44. Barnoy, S., Glaser, T., Kosower, N.S. (1998) The calpain-calpastatin system and protein degradation in fusing myoblasts. *Biochim. Biophys. Acta* **1402**, 52–60
45. Temm-Grove, C.J., Wert, D., Thompson, V.F., Allen, R.E., and Goll, D.E. (1999) Microinjection of calpastatin inhibits fusion in myoblasts. *Exp. Cell Res.* **247**, 293–303
46. Glaser, T. and Kosower, N.S. (1986) Calpain-calpastatin and fusion: fusibility of erythrocytes is determined by a protease-protease inhibitor (calpain-calpastatin) balance. *FEBS Lett.* **206**, 115–120
47. De Tullio, R., Passalacqua, M., Averna, M., Salamino, F., Melloni, E., and Pontremoli, S. (1999) Changes in intracellular localization of calpastatin during calpain activation. *Biochem. J.* **343**, 467–472
48. Dunham, I., Shimizu, N., Roe, B.A., Chissoe, S., *et al.* (1999) The DNA sequence of human chromosome 22. *Nature* **402**, 489–495
49. Ayoubi, T.A.Y. and Van de Ven, W.J.M. (1996) Regulation of gene expression by alternative promoters. *FASEB J.* **10**, 453–460
50. Timmusk, T., Palm, K., Metsis, M., Reintam, T., Paalme, V., Saarma, M., and Persson, H. (1993) Multiple promoters direct tissue-specific expression of the rat BDNF gene. *Neuron* **10**, 475–489
51. Chikhi, N., Holic, N., Guellaen, G., and Laperche, Y. (1999) Gamma-glutamyl transpeptidase gene organization and expression. a comparative analysis in rat, mouse, pig and human species. *Comparative Biochem. Physiol. Part B*, **122**, 367–380
52. Inazawa, J., Nakagawa, H., Misawa, S., Abe, T., Minoshima, S., Fukuyama, R., Maki, M., Murachi, T., Hatanaka, M., and Shimizu, N. (1990) Assignment of the human calpastatin gene (CAST) to chromosome 5 at region q14-q22. *Cytogenet. Cell Genet.* **54**, 156–158
53. Maki, M., Takano, E., Mori, H., Sato, A., Murachi, T., and Hatanaka, M. (1987) All four internally repetitive domains of pig calpastatin possess inhibitory activities against calpains I and II. *FEBS Lett.* **223**, 174–180
54. Emori, Y., Kawasaki, H., Imajoh, S., Minami, Y., and Suzuki, K. (1988) All four repeating domains of the endogenous inhibitor for calcium-dependent protease independently retain inhibitory activity. Expression of the cDNA fragments in *Escherichia coli*. *J. Biol. Chem.* **263**, 2364–2370