J. Biochem. 2012;151(3):263–272 doi:10.1093/jb/mvr135

JB THE JOURNAL OF BIOCHEMISTRY

Role of the transcription factor Sp1 in regulating the expression of the murine cathepsin E gene

Received October 14, 2011; accepted October 31, 2011; published online December 30, 2011

Kuniaki Okamoto[1,*], Yoshiko Okamoto[2], Tomoyo Kawakubo[3], Jun-ichi Iwata[4], Yoshiyuki Yasuda[5], Takayuki Tsukuba[1] and Kenji Yamamoto[3]

[1]Department of Oral Pathopharmacology, Graduate School of Biomedical Sciences, Nagasaki University, Nagasaki 852-8588; [2]Department of Molecular Biology, Pharmaceutical Sciences, Daiichi College, Fukuoka 815-8511; [3]Proteolysis Research Laboratory, Graduate School of Pharmaceutical Sciences, Kyushu University, Fukuoka 812-8582, Japan; [4]Center for Craniofacial Molecular Biology, School of Dentistry, University of Southern California, Los Angeles, CA, USA; and [5]Division of Clinical Cariology and Endodontology, Department of Oral Rehabilitation, School of Dentistry, Health Sciences University of Hokkaido, Hokkaido 061-0293, Japan

*Kuniaki Okamoto, Department of Oral Pathopharmacology, Graduate School of Biomedical Sciences, Nagasaki University, Nagasaki 852-8588, Japan. Tel: +81-95-819-7653, Fax: +81-95-819-7655, email: k-oka@nagasaki-u.ac.jp

Cathepsin E (CE) is an intracellular aspartic proteinase that is exclusively expressed in cells of the gastrointestinal tracts, lymphoid tissues, urinary organs and red blood cells. However, the molecular mechanism by which CE is predominantly expressed in these cells remains unknown. Here, we report the identification of several transcription start sites of the CE gene and their regulatory factors in gastric adenosarcoma cells. We first identified several unique transcription start sites in mouse CE genes by an oligo cap method. Their analysis also revealed the existence of a non-coding region ~24-kb upstream of exon 1 in the CE gene and also the existence of two transcripts for CE. Luciferase analyses in upstream of exon 1 revealed that this site contained putative binding regions for the transcription factors Sp1, AP-1 and cEts-1 essential for the expression of CE gene. Moreover, electrophoretic mobility shift assays revealed that the protein–oligonucleotides complex of the Sp1 site were super-shifted by an anti-Sp1 antibody. The chromatin immunoprecipitation assay showed that Sp1 bound to the CE promoter region. In addition, overexpression of the Sp1 protein increased the expression of the CE protein. Altogether, these results suggest that Sp1 binding plays a particularly important role in the regulation of CE gene expression.

Keywords: aspartic proteinase/cathepsin E/promoter/Sp1.

Abbreviations: AGS, gastric adenosarcoma; CD, Cathepsin D; CE, Cathepsin E; FBS, fetal bovine serum; PCR, polymerase chain reaction; TNF, Tumour necrosis factor; TRAIL, TNF-related apoptosis-inducing ligand.

Cathepsin E (CE) is an intracellular aspartic proteinase that has a limited distribution in certain tissues and cell types such as the gastrointestinal tract tissue, lymphoid tissue, urinary tract tissue and blood cells (*1*, *2*). The intracellular localization of CE appears to vary with cell type. In immune-related cells such as macrophages, dendritic cells and microglia, CE is mainly localized to endosomal/lysosomal structures (*3*, *4*). In erythrocytes, gastric cells, renal proximal tubule cells and osteoclasts, CE is exclusively present in the plasma membrane (*2*, *5*, *6*). CE is also detected in the endoplasmic reticulum and Golgi complex in various cell types, including gastric cells (*2*), M cells (*7*), Langerhans cells and interdigitating reticulum cells (*8*). Thus, the variability in CE cellular localization implies its functional diversity in the cells. Previously, CE has been shown to participate in exogenous antigen processing in primary cultured murine microglia (*4*). Recently, studies in CE-deficient mice have yielded important information about the critical functions of CE. For example, CE-deficient mice spontaneously develop atopic dermatitis when kept under conventional conditions (*9*) and display markedly increased susceptibility to both Gram-positive and Gram-negative bacterial infections (*10*). In addition, CE-deficient mice showed the increased tumour growth, metastasis and mortality after subcutaneous injection of mouse B16 melanoma cells (*11*). In cellular levels, macrophages derived from CE-deficient mice exhibited a novel form of lysosomal storage disorder manifesting the accumulation of lysosomal membrane sialoglycoproteins and elevation of lysosomal pH (*12*).

Compared with the accumulating knowledge of the physiological or pathological functions of CE, little is known about the mechanisms underlying the tissue- or cell-specific gene expression of CE. We have found that the promoter regions of the CE gene have a unique characteristic which contrasts with that of the promoter regions of other analogous aspartic proteinases, i.e. the promoter regions flanking genes for pepsinogen (*13*, *14*), renin (*15*, *16*) and cathepsin D (*17*) possess a consensus TATA box sequence to direct the recruitment of the nuclear protein TFIID. In contrast, the promoter regions of the CE gene from humans (*18*) and mice (*19*) do not contain a TATA box upstream of the translational start site (ATG). Promoters for TATA box-less genes are often found to have a high G + C content and potential Sp1 sites (*17*, *20*). It is thus proposed that the expression of the CE gene is possibly dependent upon the balance between the effects produced by positive-acting, tissue-specific transcription factors such as GATA1 and PU1, and the negative influence of the ubiquitous factor YY1 (*21*). In accordance with the above view, we assumed

that the promoter region for CE might contain an alternative site, including a potential Sp1 site. However, such an alternative site has not yet been identified in previous studies (*18*, *19*, *21*). Given that the primer extension methods used in previous studies may cause a preparation of immature RNA, we have designed a different method to exclude the immature RNA of the CE gene. By an oligo cap method and a dual-luciferase reporter gene assay, we have identified for the first time the multiple transcription start sites, two transcripts for CE and the transcription factors Sp1, AP-1 and cEts-1 for the mouse CE gene. In addition, we showed that, of these transcription factors, Sp1 is particularly important for the regulation of CE gene expression, as revealed by point mutation studies, electrophoretic mobility shift assays (EMSA), chromatin immunoprecipitation (ChIP) assays and the overexpression of Sp1.

Materials and Methods

Identification of transcription start sites

To identify the sequence with the transcription start site, we purified the commercial cDNA library, the Cap site cDNA (Nippon Gene, Co., Inc., Ltd., Toyama, Japan), derived from the mouse spleens, and Gene Racer (invitrogen), derived from the mouse spleen and stomach, according to the manufacturer's instructions (Table 1). The primer sets for the first and second polymerase chain reaction (PCR) were shown in Tables 1 and 2. All the PCR primers were purchased from the FASMAC Co. Ltd. The final PCR products were extracted from the agarose gel after gel electrophoresis and subcloned into the pGEM-T easy vector (Promega) for sequencing.

Table I. Relationship of cDNA library, tissues and primers used to deicide transcription start sites.

TSS[a]	cDNA library	Tissue	1st PCR	2nd PCR
[b]	Cap Site cDNA	spleen	1RC/1RCA1	3RC/3RCA1
[b]	Cap Site cDNA	spleen	1RC/1RCA1	3RC/3RCA1
−20	GeneRacer	spleen	1RC/1RCA2	3RC/2RCA2
1	GeneRacer	spleen	1RC/1RCA2	2RC/2RCA1
2	GeneRacer	spleen	1RC/1RCA2	3RC/2RCA2
−43	GeneRacer	stomach	GR5′(2)/ GRA(964)	GR5′(3)/ GRA(964)
1	GeneRacer	stomach	GR5′(2)/ GRA(964)	
4	GeneRacer	stomach	GR5′(2)/ GRA(964)	

[a]Transcription start site.
[b]Two transcription start site in Fig. 1(A).

Table II. PCR primers used to decide transcription start sites.

Primer	Sequence
1RC	5′-CAAGGTACGCCACAGCGTAT G-3′
2RC	5′-GTACGCCACAGCGTATGATG C-3′
3RC	5′-GTACGCCACAGCGTATGATG CGTAA-3′
1RCA1	5′-GCAGCAGCAGCAGCAGCACA A-3′
1RCA2	5′-TCACTCCTCCAGCAGCCAAT G-3′
2RCA1	5′-CTTCCACCCTTCTTTCCTCT T-3′
2RCA2	5′-ACCCTTCCACAGAGACTTGA T-3′
GR5′(2)	5′-GACTGGAGCACGAGGACACT GAC-3′
GR5′(3)	5′-ACTGGAGCACGAGGACACTG ACA-3′
GRA(964)	5′-GGACTGCGATGGATGGATGG AATACTG-3′

Screening and localization of the non-coding exon

A total of 5×10^4 plaque units of a mouse spleen 139 Svj λFIXII genomic cDNA library (Clontech) were plated with XL1-Blue MRA(P2) and incubated at 37°C for 9 h. Hybond-C extra filters were pre-hybridized in hybridization buffer (48% formamide, 4.8× SCC, 20 mM Tris–HCl (pH 7.6), 1× Denhardt's solution, 10% dextran sulphate, 0.1% SDS) for 1 h at 42°C and then hybridized overnight at 37°C with a radiolabelled probe in hybridization buffer containing 0.1 mg/ml sonicated calf thymus DNA. The 650-bp *Sal*I–*Ase*I fragment/λCE27 was used as a probe. The filters were washed three times for 15 min in 2× SSC and 0.1% SDS at room temperature and then twice for 20 min in 1× SSC and 0.1% SDS at 60°C. Ten positive clones were obtained by the third screening. The DNA inserts of the isolated positive clones were subcloned into pBluescript SK (−) plasmid vectors and sequenced. This sequenced arrangement was searched by using the website http://www.ncbi.nlm.nih.gov/.

Construction of reporter plasmids

The reporter gene plasmids for exon 1 were generated by cloning restriction fragments isolated from the 5′-flanking region of the mouse CE gene. A series of small mouse CE promoter-luciferase reporter constructs was generated using PCR involving a common 5′-primer with an *Mlu*I site at its 5′-end and a 3′-primer with a *Bgl*II site at its 3′-end. The amplified DNA fragments were purified by electrophoresis, digested with *Mlu*I and *Bgl*II, and subcloned into the *Mlu*I–*Bgl*II sites of pGL3-Basic (Promega Co.). The resulting constructs were sequenced for verification. Mutations in the AP-1, Sp1 and cEts-1 sequences were generated by PCR site-directed mutagenesis by using high fidelity Pyrobest DNA polymerase (Takara Bio Inc.) according to the manufacturer's protocol. The oligonucleotides that were used are as follows and the mutation sites are indicated by underlining: AP1-1, 5′-$^{-127}$CGTCGCCTA GTGTACATAGTGTTTACTCTGGGC^{-95}-3′; AP1-2, 5′-$^{-84}$GTG TCAGACCTTATCATTGG^{-65}-3′; Sp1-1, 5′-$^{-52}$TTCAGGGCAG GTCTGGGCAC^{-32}-3′; Sp1-2, 5′-$^{-57}$GACTGGGCAGTTCAG GTCTG^{-37}-3′; cEts-1, 5′-$^{-36}$GCACCATTAAGAGCCAGTGCC CC^{-14}-3′.

RNA extraction and RT–PCR

Total RNA was isolated from the spleen and stomach of C57BL/6 mice using the TRIsol reagent (Invitrogen). One microgram of sample was reversed transcribed in 25 μl containing final concentrations of 20 μg/ml of Oligo d(T)15 (Promega Co.), 1 unit/μl of Ribonuclease Inhibitor (TOYOBO) and 2 units/μl of RevaTra Ace (TOYOBO) to synthesize the first strand cDNA. The PCR amplification was then performed in 25 μl reactions using 2.5 unit/μl of Gene *Taq* NT (Wako). The primers that were used were as follows: exon0/1-s, 5′-AGCCCAACCCATCCACTTAC-3′; exon0/1-as, 5′-C TCACTCTGCTCCGATCTCC-3′; exon6/7-s, 5′-GCTGACTTTCG GAGGCTATG-3′; exon6/7-as, 5′-GAGAGAGGTCCCTGTGTC CA-3′. The reaction conditions were: 95°C for 5 min; followed by 35 cycles of 95°C for 20 s, 62°C for 20 s and 72°C for 30 s; followed by one cycle at 72°C for 5 min. The amplified DNA fragments were visualized on 2% agarose gel.

Cell culture, transient transfection and luciferase assay

Human gastric adenosarcoma (AGS) cells were cultured in Dulbecco's modified Eagle's medium (Nissui) supplemented with 10% foetal bovine serum (FBS). All the cells were maintained at 37°C in a humidified incubator containing 5% CO_2 and were transiently transfected using TransFastTM reagent (Promega Co.) according to the manufacturer's protocol. The cells were then incubated for 48 h and the luciferase activities were measured using the Dual-LuciferaseTM reporter assay system (Promega Co.).

EMSA

Nuclear extracts from AGS cells and mouse stomach cells were prepared using an Active Motif Nuclear Extract Kit (Promega Co.). All the primers were purchased from FASMAC Co. The primers that were used as labelled probes were as follows, and the original and mutation sites are underlined: Sp1, 5′-$^{-63}$TCCTCAGA CTGGGCAGGGCAGGTCTGGGCA^{-34}-3′; mutant Sp1, 5′-$^{-63}$TC CTCAGACTTTGCAGTTCAGGTCTGGGCA^{-34}-3′. The binding reaction was performed for 20 min at 4°C by using a Pierce Chemiluminescent EMSA Kit (comprising 10× binding buffer

(100 mM Tris, 500 mM KCl, 10 mM DTT [pH 7.5], 50 ng/ml poly (dI-dC), 1 M KCl, 100 mM $MgCl_2$, 20 mM EDTA, 50% glycerol, 1% NP–40) containing 20,000–30,000 cpm of labelled probe, and 10 μg of nuclear extracts from AGS cells and murine stomach cells. For the competition assays, unlabelled Sp1 was used. For the supershift assay, anti-Sp1, anti-Sps 2-4 and anti-cEts–1 (Santa Cruz Biotech. Inc.) antibodies were purchased. For the interference assays, 200-fold molar excess of unlabelled competition oligonucleotides or 2 μg of antibody was added to the reaction mixture for 1 h at 4°C before the addition of the radiolabelled probe. The DNA–protein complexes were separated on 4% non-denaturing polyacrylamide gels in 1× Tris borate electrophoresis buffer at 200 V. After completion of the run, the gel was transferred onto 3MM paper (Funakoshi Co.), dried in vacuum and visualized by autoradiography.

ChIP assay

The ChIP assay was performed using a SimpleChIPTM Enzymatic Chromatin IP Kit (Cell Signaling Tech.). Briefly, the cells (4×10^7 cells/dish) were cross-linked in 1% formaldehyde for 10 min at room temperature and the reaction was quenched with glycine. After washing with ice-cold PBS, ice-cold PBS containing PMSF was added. The cells were scraped, resuspended in attachment buffer and centrifuged. After being resuspended in the attachment buffer, Micrococcal Nuclease was added to the cells and they were incubated at 37°C for 20 min to digest DNA to the optimal length. The reaction was stopped by adding 0.5 M EDTA. Cells were centrifuged at 13,000 rpm at 4°C for 1 min and resuspended in ChIP buffer. The mixture was sonicated to break the nuclear membranes and was centrifuged at 10,000 rpm at 4°C for 10 min. One hundred microlitres of the collected supernatant was added to 400 μl of ChIP buffer. Two percent of the mixture was saved as input DNA and processed for further use as a positive control. Immunoprecipitation was performed using 5 μl of antibodies against Sp1, Ap-1, cEts-1, Histone H3 (as a positive control) or normal rabbit IgG (as a negative control) at 4°C with overnight rotation. After adding protein G magnetic beads and being incubated at 4°C for 2 h, the beads were pelleted and washed sequentially several times with low and high salt buffers. The bound proteins were eluted from the beads by incubation with ChIP elution buffer at 65°C for 30 min. Five molar NaCl and proteinase K were added to all samples, including the 2% input sample. All samples were then applied to DNA spin columns for DNA purification, washed and eluted with DNA Elution Buffer. The Sp1, Ap-1 and cEts-1-binding sites were PCR-amplified using purified DNA as a template. The primers used were as follows: the forward primer, 5′-GTTCTCCCGCAGA CTCATA-3′; the reverse primer, 5′-GGGAACGGACTTTCCCTA AC-3′. The reaction conditions were: 95°C for 5 min; followed by 35 or 38 cycles at 95°C for 30 s, 62°C for 30 s and 72°C for 30 s; followed by one cycle at 72°C for 5 min. The amplified DNA fragments were visualized on 2% agarose gels.

Transfection

AGS cells were cultured to 70–80% confluence in 10% FBS-supplemented DMEM and transfected with the Sp1 expression vector using Gene PORTER 3000 (Gene Therapy System, Inc) in FBS-free DMEM. Four hours later, 10% FBS-supplemented fresh DMEM was added to the dishes. After 48 h, the medium was exchanged for 10% FBS-supplemented DMEM containing G418 for selection, and was changed every 3 days. Cloned cells were picked up, cultured and used in the western blot analyses.

DNA sequencing

The nucleotide sequencing reactions were performed by FASMAC Co.

Results

Multiple transcription start sites and the existence of non-coding exons

Tatnell *et al.* (*19*) had reported that the CE transcription start site was 120-bp upstream from methionine, a translational start site for the CE protein; however, the primer extension method in their study may have resulted in the use of immature RNA. Therefore, we reassessed the transcription start site of CE. In order to exclude immature RNA, we used the oligo cap method and found three and five transcription start sites in the mouse stomach and spleen, respectively. Since one of the transcription start sites was common for both, a total of seven unique transcription start sites was identified (Fig. 1A and B). We numbered the transcription start site that was common between the stomach and the spleen as +1. This site was located 65-bp upstream from the ATG translational start site. Interestingly, two transcription start sites were located in a non-coding exon (Fig. 1A). We next used a plaque hybridization method to determine the location of the non-coding exon. Ten positive clones were obtained by the third screening, with the *Sal*I-*Ase*I fragment of the non-coding exon used as a probe. Internet analysis (http://www.ncbi.nlm.nih.gov/) indicated that the non-coding exon was located 24-kb upstream of exon 1, which contains the ATG translational start site (Fig. 1C). The intron between the non-coding exon and exon 1 matched the GT-AG rule of splicing sites.

CE has two transcripts

To examine whether the mRNA of the CE gene containing the non-coding exon really existed, we performed RT–PCR analyses using a forward primer for the non-coding exon and a reverse primer in exon 1 (Fig. 2A). Although the RNA from the murine lung, Kidney, spleen, thymus and intestine reacted with primers containing the non-coding exon, the RNA from the stomach did not (Fig. 2B). However, both tissues reacted with downstream primers containing exons 6–7 (Fig. 2B). These results suggested that the mRNA from the stomach does not contain the non-coding exon, while that from the others does. Furthermore, reactions from downstream primers were increased in the mRNA from the lung, kidney, spleen and intestine compared with that of primers containing the non-coding exon (Fig. 2B). This suggests the possibility that both transcripts with and without the non-coding exon exist in those tissues. To clearly confirm the existence of transcripts with and without the non-coding exon, we used CE knockout mice, in which exons 1–4 were deleted. In the spleen from the CE knockout mice, although a reaction against the primer of exon 0/1 was not observed due to the deletion of exon 1 containing the reverse primer, a reaction against the primer of exon 6/7 was observed (Fig. 2C). On the other hand, in the stomach from CE knockout mice, reactions against both sets of primers were not observed (Fig. 2C). These results show that two transcripts with and without the non-coding exon exist in CE.

Analyses of promoter regions

For the analyses of the CE promoter, we tried using a several kind of cell lines. However, as shown in Fig. 3A, only the AGS cells expressed the CE protein. Therefore, when we performed the luciferase assay in AGS cells to identify the regions that play a pivotal role in CE mRNA expression, while other cell lines,

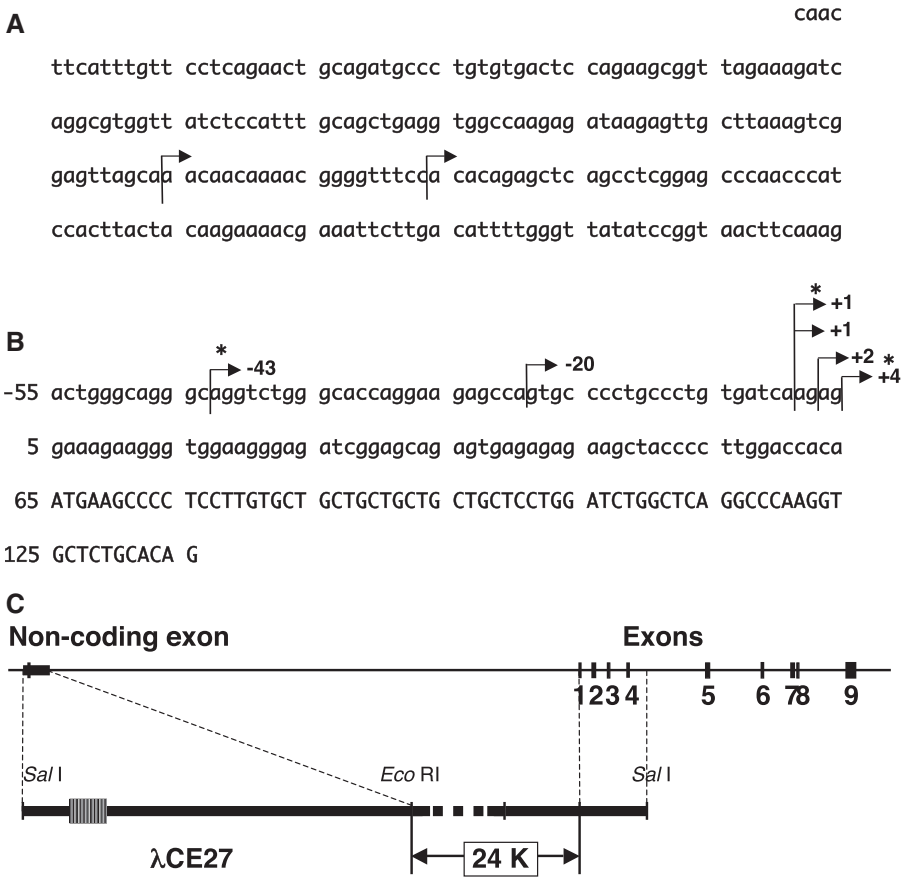

Fig. 1 The mouse CE gene promoter sequence and its putative transcription start sites. The transcription start sites were determined using Cap Site cDNA, rOligo of GeneRacer and GROligo of GeneRacer. Transcription start sites from the Cap site cDNA are indicated by arrows. The asterisk indicates the CE transcription start sites in the stomach, while the other is from the spleen. (A) Non-coding exon and (B) exon 1. Numbering was decided based on the transcription start site that was common between the stomach and the spleen (+1). (C) The relationship between the non-coding exon and exon 1.

such as HeLa, HEK293 and NRK showed little or no luciferase activity (data not shown). We first examined the promoter region of upstream of exon 1 because there were many potential transcription start sites upstream of exon 1 (Fig. 1B) and because there were no transcripts containing a non-coding exon in AGS cells that originated from gastric tissue (Fig. 2C). Using a series of deletion constructs from the 5′-direction of the putative promoter region upstream of exon 1, a remarkable decrease in promoter activity was observed when the region extending from −121 to −78 was deleted (Fig. 3B). This result suggests that the region from −121 to −78 is important for the promoter activities of the CE gene. We next performed a series of deletions from the 3′-direction of the putative promoter region upstream of exon 1. When the region from −46 to −18 was deleted, the promoter activity was strongly decreased (Fig. 3C). To confirm whether this region was crucial for promoter activity, we examined the effect of deleting it and found that the deletion mutant had considerably lower promoter activity than did 186N/Luc, the shortest constructs having promoter activity among the ones used in Fig. 3B were examined (Fig. 3C, 156N/Luc). No promoter activity was observed for the deletion site alone (Fig. 3C, 30N/Luc). These results suggest that the latter site, the region from −46 to −18, is also important for the promoter activities of the CE gene.

Transcription of the CE gene strongly depends on Sp1

When we searched for possible transcription factor sites in the two regions extending from −121 to −78 and from −46 to −18, we found putative binding sites for three transcription factors Sp1, AP-1 and cEts−1. Therefore, we introduced one point mutation into each of these transcription factors by using site-directed mutagenesis (Fig. 4A). The each mutation in two AP-1 sites decreased its promoter activity to 50% less than the promoter activity of the 186N/Luc reporter (Fig. 4B), and the mutations in two Sp1 and one cEts-1 sites led to an even greater decrease in promoter activity relative to that of 186N/Luc (Fig. 4B). To determine whether specific nuclear proteins bind to the Sp1-binding site of the CE gene, we performed an EMSA assay with radiolabelled probes. The ^{32}P-labelled Sp1 oligonucleotides bound to nuclear extracts prepared from AGS cells or murine stomach cells (Fig. 5A and B; lane 2). The specificity of these interactions was confirmed on incubation of the nuclear extracts with mutated Sp1 or a 200-fold molar excess of unlabelled Sp1 oligonucleotides (Fig. 5A and B; lanes 3 and 4). When Sp1 antibody

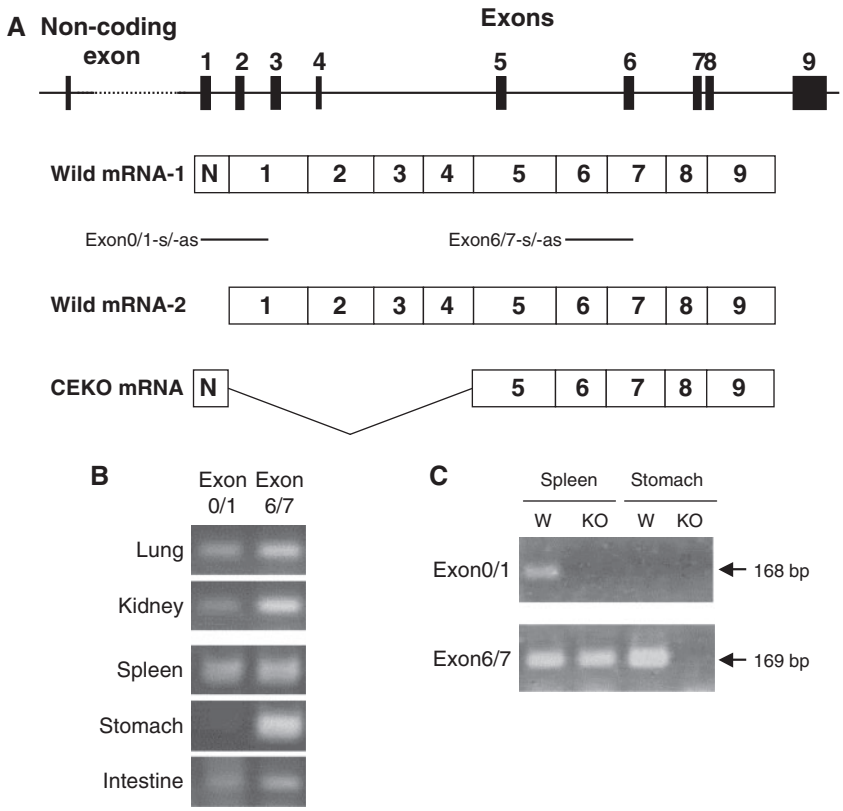

Fig. 2 The two transcripts of CE. (A) A schematic diagram of the two transcripts of CE. (B) Total RNAs were isolated from the murine spleen and stomach and RT–PCR was performed with 1 μg of total RNA. The amplified DNA fragments were visualized on a 2% agarose gel. (C) Total RNAs were isolated from the CE knockout mice spleen and stomach and RT–PCR was performed with 1 μg of total RNA. The amplified DNA fragments were visualized on 2% agarose gel. W and KO indicate wild type mouse (C57BL/6) and CE knockout mouse, respectively.

was added to the binding reaction, these complexes were supershifted (Fig. 5A and B; lane 5). These shifts were not seen when Sp2, Sp3, Sp4 and cEts-1 antibodies were used (Fig. 5A and B; lane 6–9). These results suggest that transcription of the CE gene strongly depends on the binding of the Sp1 transcription factor.

In vivo binding to the CE promoter region of Sp1

To investigate whether Sp1, AP-1 and cEts-1 binds the CE promoter region in AGS cells *in vivo*, we performed a ChIP assay. Immunoprecipitated chromosomal DNA was subjected to PCR using primers designed to amplify the CE promoter region harbouring the Sp1-, AP-1- and cEts-1-binding sites. Sp1 bound to the CE promoter region containing Sp1 sites, whereas AP-1 and cEts-1 exhibited minimal binding (Fig. 6A). To further investigate the relationship between the CE promoter region and Sp1, we overexpressed the Sp1 gene in AGS cells. As shown in Fig. 6B, expression of the CE protein was remarkably increased when Sp1 gene was overexpressed in AGS cells.

Analysis of the non-coding region of the CE gene

To identify the non-coding region crucial to the transcriptional activity of the CE gene, the promoter activities of various lengths of the 5′-flanking region were measured. We designed a series of deletion constructs having a different 3′-end from the upstream region of the non-coding region (Fig. 7). Although the promoter activity of 130NLuc, the shortest construct among them used in this time, was slightly higher than those of the other constructs, this upstream promoter activity of the non-coding exon was lower than that of exon 1 (Fig. 7). We attempted to identify specific transcription factors in the promoter region of non-coding exon. However, we could not identify any TATA box-less related transcription factors such as Sp1 or AP-1.

Discussion

In this study, we defined the transcription start sites by using the Cap site cDNA method to exclude immature RNA. Consequently, we showed that the CE gene had multiple transcription start sites. Furthermore, we identified two regions containing binding sites for transcription factors such as Sp1, AP-1 and cEts-1 by using a dual-luciferase reporter gene assay. Point mutations in these transcription factors decreased luciferase activity. EMSA analysis revealed that the protein–oligonucleotides complex of the Sp1 site were supershifted by an anti-Sp1 antibody. In addition, we showed that Sp1 is involved in the expression of CE *in vivo* by using a ChIP assay and by overexpression Sp1. Finally, we analysed the non-coding region of the CE gene.

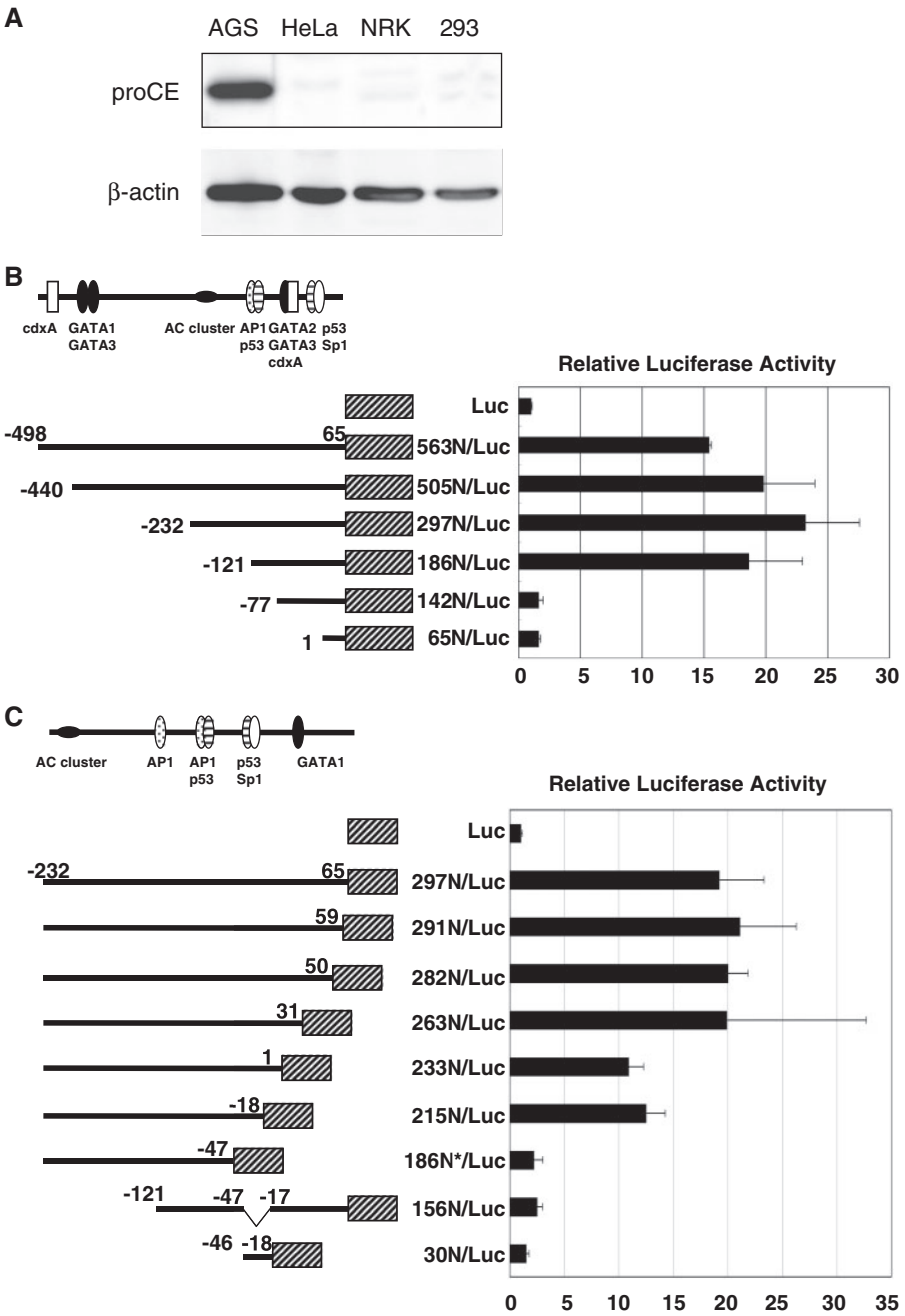

Fig. 3 The transcriptional activity of the deletion constructs from exon 1 transfected in AGS cells. AGS cells were co-transfected with a series of deletion mutant constructs from the 5′- and 3′-flanking region of exon 1 and a TK promoter-driven Renilla luciferase plasmid as an internal control. (A) The western blot analysis for AGS, HeLa, NRK and HEK293T cells against anti-human proCE. (B) The deletion mutant constructs from the 5′-flanking region of exon 1 and their relative luciferase activities. (C) The deletion mutant constructs from the 3′-flanking region of exon 1 and their relative luciferase activities. The deletion construct from −47 to −17 and the construct fused only in this region and their relative luciferase activities. The transcriptional activity is indicated as the relative level compared to pGL3/Basic.

In a previous study, a genomic clone encompassing the murine CE gene was isolated and sequenced (*19*). The gene was found to consist of nine exons, and the 5′-flanking region of the CE gene appeared to be a TATA-less promoter; the absence of features considered typical of TATA-box regulated or housekeeping-type genes was consistent with the low levels of CE gene expression. Furthermore, the transcription start site of the CE gene from mice and humans was determined by primer extension methods (*19*). We tried to define the transcription start site by using the Cap site cDNA method because immature RNA can be excluded using this method. However, we did not find the same transcription start site reported by Tatnell *et al.* (*19*). Instead, we found multiple transcription start sites in the CE gene in the spleen and stomach cells from mice. As some researchers have shown, it is possible for several transcription start sites to exist in a single gene (*17*, *22*), and the transcription start site of the CE gene may represent such

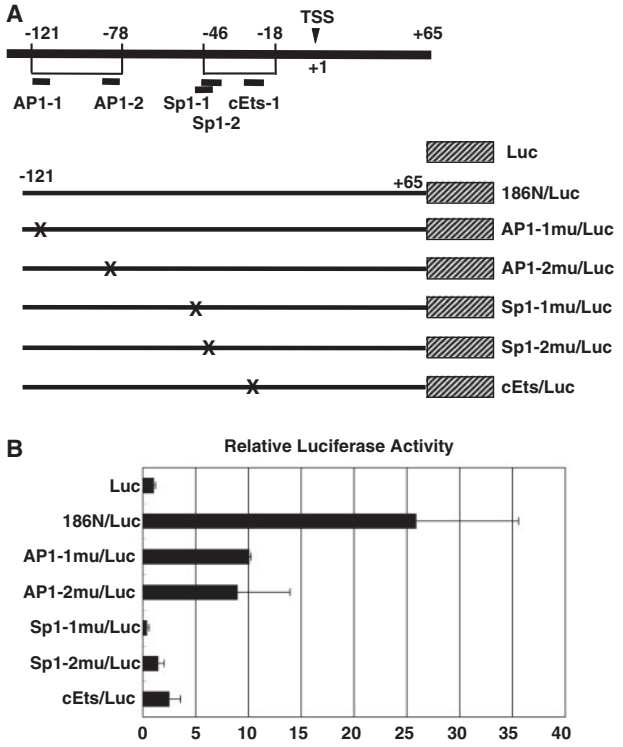

Fig. 4 The transcriptional activity of the point mutant constructs from exon 1 transfected in AGS cells. AGS cells were co-transfected with constructs containing mutations in the regions for the binding of transcription factors such as Sp1, AP-1 and cEts-1 and a TK promoter-driven Renilla luciferase plasmid was used as an internal control. (A) The point mutant constructs. (B) The ralative luciferase activities of each point mutated construct. The transcriptional activity is indicated as the relative level to that of pGL3/Basic.

a case. Therefore, the transcription start site identified by Tatnell *et al.* (*19*) might be only one of several start sites.

To analyse the CE promoter regions, we wanted to examine a several different cell lines because there is thought to be cell and tissue specificity of CE. However, as shown in Fig. 3A, very few cell lines express the CE protein. This may be one of the reasons why it is very difficult to elucidate the cell and tissue specificity of CE. In this study, we used the AGS cells, which abundantly express CE, to identify transcription factors participating in CE mRNA transcription. On performing a refined deletion analysis of the CE promoter in AGS cells, we identified the transcriptionally important sites between −121 and −78 as AP-1-binding sites, and between −46 and −18 as Sp1 and cEts-1-binding sites. The importance of these transcription factors was confirmed by analysis of the luciferase constructs with deletions or mutations in these sites, although the function of AP-1 was less defined than that of Sp1 or cEts-1. Furthermore, from the results of the gel shift assay studies performed using nuclear extracts prepared from AGS cells, we confirmed the binding of the Sp1 protein to these elements. Sp1 was one of the first cellular transcription factors to be identified and cloned from HeLa cells (*23*). It was originally identified as a factor that could bind and

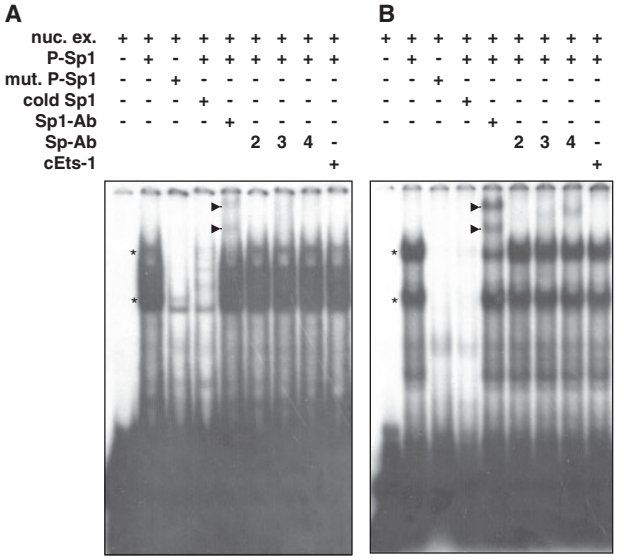

Fig. 5 *In vitro* binding of the Sp group and cEts-1-binding proteins to the CE promoter region in AGS cells. Gel shift assays were performed using AGS nuclear extracts. (A) Nuclear extracts from AGS cells and (B) nuclear extracts from mouse stomach cells. The asterisks indicate ^{32}P-labelled Sp1 oligonucleotides. The supersifted bands of Sp1 are indicated by arrow heads.

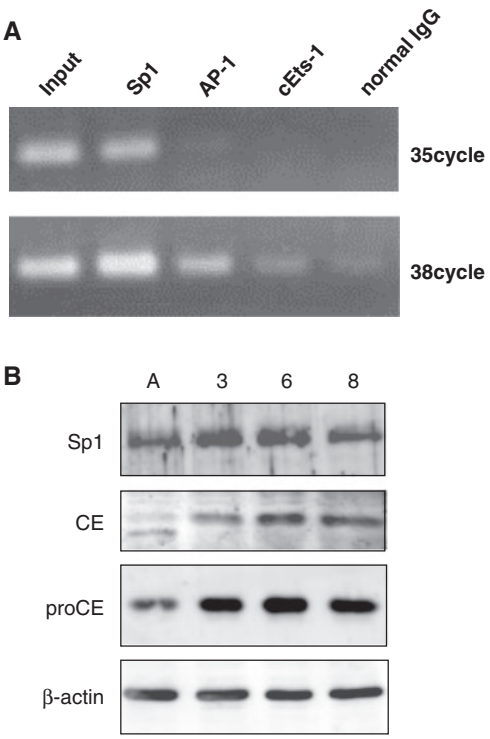

Fig. 6 *In vivo* binding to the CE promoter region in AGS cells. (A) A ChIP assay of the *in vivo* recruitment of Sp1, AP1 and cEts-1 to the CE promoter in AGS cells. DNA purified from the immunoprecipitated chromatin was amplified by PCR. The upper panel shows 35 cycles of PCR and the lower shows the amplification after 38 cycles. The 'Input' shows positive expression, and 'normal IgG' shows the results of negative control. (B) The results of the western blot analysis of the CE protein in AGS cells and AGS cells overexpressing Sp1. The total protein from AGS cells and AGS cells stably transfected with Sp1 was collected and analysed by a Western blotting analysis. β-actin was used as a loading control. The indicated numbers show the clone number of AGS cells overexpressing Sp1.

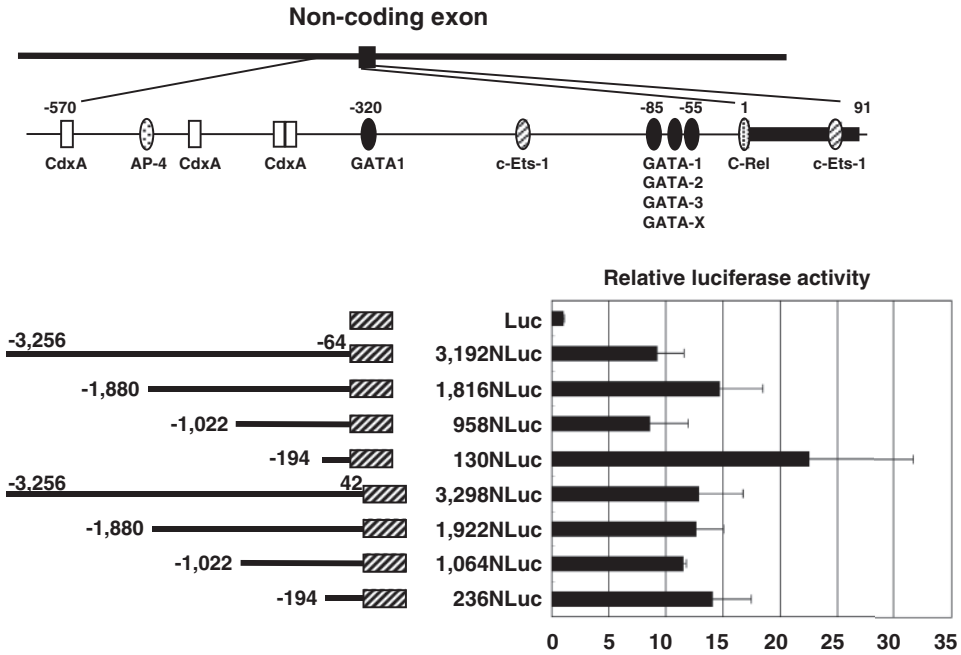

Fig. 7 The transcriptional activity of the CE genomic constructs from the non-coding region transfected in AGS cells. AGS cells were co-transfected with deletion mutant constructs from the 5′-flanking region of exon 1 by using a TK promoter-driven Renilla luciferase plasmid as an internal control. The transcriptional activity is indicated as the relative fold to pGL3/Basic.

activate the SV40 early promoter and thymidine kinase (TK) promoter (*24*, *25*). It binds to GC- or GT-rich DNA sequences in human gene promoters and regulates a large number of housekeeping and tissue-specific genes (*26*–*28*). Although we tried other Sp group proteins such as Sp2, Sp3 and Sp4, we did not observe any supershifted bands on performing the EMSA assay.

We could not identify any reactive bands on performing an EMSA assay of the AP-1 region. However, the ChIP assay revealed the possibility that AP-1 bound to the CE promoter region *in vivo*. A recent study showed that the combination of the ubiquitous transcription factors AP-1 and Sp1 directs keratinocyte-specific and differentiation-specific gene expression *in vitro* (*29*). Furthermore, Gao *et al.* showed that Sp1 and AP-1 together regulated VIL2 promoter activity (*30*). These studies showed that the physical interaction between Sp1 and AP-1 is important for their activity and may lead to tissue-specific expression patterns. Previously, Cook *et al.* had reported that the regulation of human and mouse CE gene expression depended on transcription factors such as YY1 (*21*). Although, unfortunately, we could not find these transcription factors in the −121 to −78 bp and −46 to −18 bp regions upstream of the transcription start site, several researchers have reported an interaction between Sp1 and YY1 (*31*, *32*). Therefore, it is possible that Sp1 also interacts with YY1 for the regulation of CE.

This study also revealed the existence of a non-coding region in the CE gene. The non-coding region had another start site that was located ~24-kb upstream of exon 1. Therefore, it appears that the CE gene consists of a total of 10 exons, including the non-coding exon. From the RT–PCR experiments, we found that CE had two transcripts. Interestingly, it seems that the non-coding exon exists in the spleen and other tissues, but not the stomach (Fig. 2B). This is also confirmed in the results from the CE knockout mice (Fig. 2C). However, it is unclear why the non-coding exon exists or how it functions. It is possible that the non-coding region is responsible for the limited and specific distribution of CE in tissues and cell types such as the gastrointestinal tract tissue, lymphoid tissue, urinary tract tissue and blood cells. In this study, we were unable to identify the specific transcription factors that led to the expression of the non-coding CE exon. A greater number of detailed experiments will be required to clarify the function of the non-coding exon.

Identifying of the role of the transcription factor Sp1 in regulating the expression of the CE gene may therefore be useful for understanding the mechanisms underlying the up-regulation of CE expression in tumour cells. In fact, as shown in Fig. 6B, the over-expression of Sp1 increased the expression of the CE protein. We recently reported that CE prevents tumour growth and metastasis by catalysing the proteolytic release of soluble TRAIL from the tumour cell surface in CE-deficient mice (*11*). This finding suggests that increase in CE expression levels prevents tumour growth and metastasis. Guerardel *et al.* reported that p53 could up-regulate the expression of HIC-1 (hypermethylated in cancer 1) transcripts (*33*). Interestingly, we found that the binding sites for p53 are located in two regions (−121 to −78 bp and −46 to −18 bp regions located upstream of the transcription start site) important for CE gene expression. These results suggest that Sp1 and/or p53 regulate the growth and

metastasis of tumour cells and may be useful for future development of chemotherapeutic agents.

In conclusion, we determined that the transcription factor Sp1 plays a role in regulating the expression of the murine CE gene. This information will be useful for understanding the tissue-specific or cell-specific expression systems of CE.

Acknowledgements

The authors thank Dr Kazuhisa Nishishita, Dr. Eiko Sakai (University of Nagasaki, Japan) for the useful discussions and Dr Kei Suzuki (Research Institute for Microbial Disease, Osaka University, Japan) for the kind gifts of the Sp1 expression vector.

Funding

This work was supported in part by a Grant-in-Aid for Science Research from the Ministry of Education, Culture, Sports, Sciences and Technology of Japan.

Conflict of interest

None declared.

References

1. Sakai, H., Saku, T., Kato, Y., and Yamamoto, K. (1989) Quantitation and immunohistochemical localization of cathepsins E and D in rat tissues and blood cells. *Biochim. Biophys. Acta* **991**, 367–375
2. Saku, T., Sakai, H., Shibata, Y., Kato, Y., and Yamamoto, K. (1991) An Immunocytochemical study on distinct intracellular localization of cathepsin E and cathepsin D in human gastric cells and various rat cells. *J. Biochem.* **110**, 956–964
3. Sastradipura, D.F., Nakanishi, H., Tsukuba, T., Nishishita, K., Sakai, H., Kato, Y., Gotow, T., Uchiyama, Y., and Yamamoto, K. (1998) Identification of cellular compartments involved in processing of cathepsin E in primary cultures of rat microglia. *J. Neurochem.* **70**, 2045–2056
4. Nishioku, T., Hashimoto, K., Yamashita, K., Liou, S.-Y., Kagamiishi, Y., Maegawa, H., Katsube, N., Peters, C., von Figura, K., Saftig, P., Katunuma, N., Yamamoto, K., and Nakanishi, H. (2002) Involvement of cathepsin E in exogeneous antigen processing in primary cultured murine microglia. *J. Biol. Chem.* **277**, 4816–4822
5. Takeda-Ezaki, M. and Yamamoto, K. (1993) Isolation and biochemical characterization of procathepsin E from human erythrocyte membranes. *Arch. Biochem. Biophys.* **304**, 352–358
6. Yoshimine, Y., Tsukuba, T., Isobe, R., Sumi, M., Akamine, A., Maeda, K., and Yamamoto, K. (1995) Specific immunocytochemical localization of cathepsin E at the ruffled border membrane of active osteoclasts. *Cell Tissue Res.* **281**, 85–91
7. Finzi, G., Cornaggia, M., Capella, C., Fiocca, R., Bosi, F., Solcia, E., and Samloff, I.M. (1993) Cathepsin E in follicle associated epithelium of intestine and tonsils: locatization to M cells and possible role in antigen processing. *Histochemistry* **99**, 201–211
8. Solcia, E., Paulli, M., Silini, E., Fiocca, R., Finzi, G., Kindl, S., Boveri, E., Bosi, F., Cornaggia, M., and Capella, C. (1993) Cathepsin E in antigen-presenting Langerhans and interdigitating reticulum cells. Its possible role in antigen processing. *Eur. J. Histochem.* **37**, 19–26
9. Tsukuba, T., Okamoto, K., Okamoto, Y., Yanagawa, M., Kohmura, K., Yasuda, Y., Uchi, H., Nakahara, T., Furue, M., Nakayama, K., Kadowaki, T., Yamamoto, K., and Nakayama, K-I. (2003) Association of Cathepsin E Deficiency with Development of Atopic Dermatitis. *J. Biochem.* **134**, 893–902
10. Tsukuba, T., Yamamoto, S., Yanagawa, M., Okamoto, K., Okamoto, Y., Nakayama, K-I., Kadowaki, T., and Yamamoto, K. (2006) Cathepsin E-deficiency mice show increased susceptibility to bacterial infection associated with the decreased expression of multiple cell surface Toll-like receptors. *J. Biochem.* **140**, 57–66
11. Kawakubo, T., Okamoto, K., Iwata, J., Shin, M., Okamoto, Y., Yasukochi, A., Nakayama, K-I., Kadowaki, T., Tsukuba, T., and Yamamoto, K. (2007) Cathepsin E prevents tumor growth and metastasis by catalyzing the proteolytic release of soluble TRAIL from tumor cell Surface. *Cancer Res.* **67**, 10869–10878
12. Yanagawa, M., Tsukuba, T., Nishioku, T., Okamoto, Y., Okamoto, K., Takii, R., Terada, Y., Nakayama, K-I., Kadowaki, T., and Yamamoto, K. (2007) Cathepsin E deficiency induces a novel form of lysosomal storage disorder showing the accumulation of lysosomal membrane sialoglycoproteins and the elevation of lysosomal pH in macrophages. *J. Biol. Chem.* **282**, 1851–1862
13. Sogawa, K., Fujii-Kuriyama, Y., Mizukami, Y., Ichihara, Y., and Takahashi, K. (1983) Primary structure of human pepsinogen gene. *J. Biol. Chem.* **258**, 5306–5311
14. Hayano, T., Sogawa, K., Ichihara, Y., Fujii-Kuriyama, Y., and Takahashi, K. (1988) Primary structure of human pepsinogen C gene. *J. Biol. Chem.* **263**, 1382–1385
15. Hobert, P.M., Fogliano, M., O'Connor, B.A., Schaefer, I.M., and Chirgwin, J.M. (1984) Human rennin gene: structure and sequence analysis. *Proc. Natl Acad. Sci. USA* **81**, 5026–5030
16. Miyazaki, H., Fukamizu, A., Hirose, S., Hayashi, T., Hori, H., Ohkubo, H., Nakanishi, S., and Murakami, K. (1984) Structure of the human rennin gene. *Proc. Natl Acad. Sci. USA* **81**, 5999–6003
17. Cavailles, V., Augereau, P., and Rochefort, H. (1993) Cathepsin D gene is controlled by a mixed promoter, and estrogens stimulate only TATA-dependent transcription in breast cancer cells. *Proc. Natl Acad. Sci. USA* **90**, 203–207
18. Azuma, T., Liu, W., Vander Laan, D.J., Bowecock, A.M., and Taggarts, R.T. (1992) Human gastric cathepsin E gene. *J. Biol. Chem.* **267**, 1609–1614
19. Tatnell, P.J., Roth, W., Deussing, J., Peters, C., and Kay, J. (1998) Mouse procathepsin E gene: molecular organization and chromosomal localization. *Biochim. Biophys. Acta* **1398**, 57–66
20. Blake, M.C., Jambou, R.C., Swick, A.G., Kahn, J.W., and Azizkhan, J.C. (1990) Transcriptional initiation is controlled by upstream GC-box interactions in a TAT AA-less promoter. *Mol. Cell. Biol.* **10**, 6632–6641
21. Cook, M., Caswell, R.C., Richards, R.J., Kay, J., and Tatnell, P.J. (2001) Regulation of human and mouse procathepsin E gene expression. *Eur. J. Biochem.* **268**, 2658–2668
22. Gum, J.R. Jr., Hicks, J.W., Crawley, S.C., Dahl, C.M., Yang, S.C., Roberton, A.M., and Kim, Y.S. (2003)

Initiation of transcription of the MUC3A human intestinal mucin from a TATA-less promoter and comparison with the MUC3B amino terminus. *J. Biol. Chem.* **278**, 49600–49609

23. Kadonaga, J.T., Carner, K.R., Masiarz, F.R., and Tjian, R. (1987) Isolation of cDNA encoding transcription factor Sp1 and functional analysis of the DNA binding domain. *Cell* **51**, 1079–1090
24. Dynan, W.S. and Tjian, R. (1983) The promoter-specific transcription factor Sp1 binds to upstream sequences in the SV40 early promoter. *Cell* **35**, 79–87
25. Jones, K.A., Yamamoto, K.R., and Tjian, R. (1985) Two distinct transcription factors bind to the HSV thymidine kinase promoter in vitro. *Cell* **42**, 559–572
26. Suske, G. (1999) The Sp-family of transcription factors. *Gene* **238**, 291–300
27. Black, A.R., Black, J., and Azizkhan-Clifford, J. (2001) Sp1 and kruppel-like factor family of transcription factors in cell growth regulation and cancer. *J. Cell. Phys.* **188**, 143–160
28. Resendes, K.K. and Rosmarin, A.G. (2004) Sp1 control of gene expression in myeloid cells. *Crit. Rev. Eukaryot. Gene Expr.* **14**, 171–181
29. Nakamura, Y., Kawachi, Y., Xu, X., Sakurai, H., Ishii, Y., Takahashi, T., and Otsuka, F. (2007) The combination of ubiquitous transcription factors AP-1 and Sp1 directs keratinocyte-specifc and differentiation-specific gene expression in vitro. *Exp. Dermatol.* **16**, 143–150
30. Gao, S.-Y., Li, E.-M., Cui, L., Lu, X.-F., Meng, L.-Y., Yuan, H.-Y., Xie, J.-J., Du, Z.-P., Pang, J.-X., and Xu, L.-Y. (2009) Sp1 and AP-1 regulate expression of the human gene VIL2 in esophageal carcinoma cells. *J. Biol. Chem.* **284**, 7995–8004
31. Seto, E., Lewis, B., and Shenk, T. (1993) Interaction between transcription factors Sp1 and YY1. *Nature* **365**, 462–464
32. Lee, J.-S., Galvin, K.M., and Shi, Y. (1993) Evidence for physical interaction between the zinc-finger transcription factor YY1 and Sp1. *Proc. Natl Acad. Sci. USA* **90**, 6145–6149
33. Guerardel, C., deltour, S., Pinte, S., Monte, D., Begue, A., Godwin, AK., and Leprince, D. (2001) Identification in the human candidate tumor suppressor gene HIC-1 of a new major alternative TATA-less promoter positive regulated by p53. *J. Biol. Chem.* **276**, 3078–3089