Characterization of human cathepsin L promoter and identification of binding sites for NF-Y, Sp1 and Sp3 that are essential for its activity

Didier JEAN, Nathalie GUILLAUME and Raymond FRADE[1]

Immunochimie des Régulations Cellulaires et des Interactions Virales, INSERM U.354, Centre INSERM, Hôpital Saint-Antoine, 184 rue du Faubourg Saint-Antoine, 75012 Paris, France

Cathepsin L is a cysteine protease whose overexpression in human melanoma cells increases their tumorigenicity and switches their phenotype from non-metastatic to highly metastatic. Regulation of the transcription of the gene encoding human cathepsin L has not been yet studied and only preliminary data exist on the promoter regulation of the gene encoding rodent cathepsin L. In the present study we identified molecular elements involved in the transcriptional regulation of human cathepsin L in melanoma cells. The sequence of the 5′-flanking region of the gene encoding human cathepsin L was determined up to 3263 bp upstream of the translation start site. The major transcription intiation site was located. Three mRNA splice variants, differing in their 5′ untranslated ends, were identified. Regulatory regions crucial for cathepsin L promoter activity were characterized between −1489 and −1646 bp. In this region, two GC boxes (−1590/−1595 and −1545/−1550) and a CCAAT motif (−1571/−1575) were involved in specific DNA–protein interactions. An electrophoretic mobility-shift assay demonstrated that Sp1 and Sp3 transcription factors bound to these GC boxes, and only the transcription factor nuclear factor Y (NF-Y) bound to the CCAAT motif. Mutagenesis studies demonstrated that these binding sites contributed at least 85% of cathepsin L promoter activity. Thus structural and functional analysis demonstrated that binding sites for NF-Y, Sp1 and Sp3 are essential for transcription of the gene encoding human cathepsin L in melanoma cells.

Key words: complement, cysteine protease, melanoma, transcription factor.

INTRODUCTION

Cathepsin L is a papain-type cysteine proteinase [1–3] whose gene is located on human chromosome 9q21–22 [4]. Cathepsin L, initially translated as preprocathepsin L, is then transferred through the Golgi as procathepsin L and stored in lysosomes as mature cathepsin L [5]. As a lysosomal endopeptidase, mature cathepsin L degrades a wide range of intracellular cytoplasmic and nuclear proteins [6,7]. In addition, procathepsin L can be secreted by normal or transformed cells [5] and can therefore also be involved in the cleavage of extracellular proteins, thus controlling their functions. We demonstrated that procathepsin L secreted by human melanoma cells cleaved human C3, the third component of complement, thus inhibiting the cytolytic activity of complement, i.e. controlling the humoral immune system [8,9]. Cathepsin L secreted by an endothelial tumour cell line generated endostatin from collagen XVIII, suggesting that cathepsin L might have a role in angiogenesis [10] and in tumour progression [11,12]. Anti-sense inhibition of cathepsin L mRNA decreased tumour growth in murine myeloma [13]. Pretreatment of tumorigenic and highly metastatic human melanoma cells with anti-(cathepsin L) antibody strongly inhibited their tumorigenicity and significantly decreased their metastatic potential in nude mice [14]. Cathepsin L is highly expressed in transformed cells in culture [15] and human tumours [16]. More recently, we demonstrated that overexpression of procathepsin L in human melanoma cells increases their tumorigenicity and switches their phenotype from non-metastatic to highly metastatic [17]. Therefore the enforced expression and secretion of procathepsin L by human melanoma cells arms them with the ability to inactivate complement-mediated cell lysis and contributes to tumour growth and metastasis [18].

All these results emphasized the need to elucidate the exact mechanisms of regulation of expression of the gene encoding human cathepsin L and to analyse its transcriptional regulation. Indeed, incomplete and very preliminary data existed in this field only for rodents [19–22]. In human, only partial sequence data have been described: cathepsin L cDNA was cloned [23,24] and the intron–exon structure of its gene was determined [4]. However, although the ATG translation start site was located in exon 2 and the DNA sequences of intron 1 and exon 1 were determined, the DNA sequence upstream of the transcript region has remained unknown. In addition, the transcription initiation site and transcription factors involved in its promoter regulation have not yet been identified.

In the present study we therefore analysed molecular elements involved in the transcriptional regulation of human cathepsin L. We first focused on the cloning of the 5′-flanking region of the human cathepsin L gene and on the identification of transcription initiation sites. Then we identified the region of the promoter and

Abbreviations used: AP-3, activator protein 3; C/EBP, CCAAT-enhancer-binding protein; CMV, cytomegalovirus; CTF/NF-1, CCAAT-binding transcription factor/nuclear factor 1; EMSA, electrophoretic mobility-shift assay; GAPDH, glyceraldehyde-3-phosphate dehydrogenase; mAb, monoclonal antibody; NF-Y, nuclear factor Y; RLM–RACE, RNA-ligase-mediated rapid amplification of cDNA ends.

[1] To whom correspondence should be addressed (e-mail frade354@easynet.fr).

The nucleotide sequence data reported will appear in DDBJ, EMBL and GenBank® Nucleotide Sequence Databases under the accession number AF394055.

transcription factors that regulate cathepsin L promoter activity. We demonstrated that the binding sites specifically recognized by nuclear factor Y (NF-Y), Sp1 and Sp3 transcription factors were essential for human cathepsin L gene promoter activity.

MATERIALS AND METHODS

Cell lines and culture conditions

The DM-4 human melanoma cell line was established from surgical specimens of lymph node metastasis [25]. The A375SM cell line was established from nude-mice lung metastasis produced by the A375-P human melanoma cell line isolated from a lymph node metastasis [26,27]. Both cell lines are highly tumorigenic and metastatic in nude mice and were kindly provided by Dr M. Bar-Eli (M. D. Anderson Cancer Center, Houston, TX, U.S.A.). The melanoma cell lines were maintained in culture as adherent monolayers in DMEM (Dulbecco's modified Eagle's medium) supplemented with 10% (v/v) fetal calf serum, Glutamax-1, glucose, sodium pyruvate, non-essential amino acids and penicillin/streptomycin (Life Technologies) at 37 °C in an incubator containing air/CO_2 (19:1).

Cloning of the 5′-flanking region of the human cathepsin L gene

To identify the unknown sequence of the 5′-flanking region of the human cathepsin L gene, we used a PCR-based method for DNA walking in uncloned genomic DNA (GenomeWalker kit; Clontech). Nested PCRs were realized on adaptator-ligated human genomic DNA provided with the GenomeWalker kit by using two primer pairs. A first step was performed with primers specific for the ligated adaptor (AP1 and AP2) and two primers specific for the 5′ region of human cathepsin L cDNA: the primary primer was 5′-GCTTTCCTGCTGCGGTCGTAGCT-3′ and the nested primer was 5′-GAAGATCTCCGAGACTCC-ACGGATGCCGCTCAAGGTTG-3′. Then, in a second step, the sequence of a larger DNA fragment was determined with primers AP1 and AP2 and two primers specific for the first fragment of the 5′-flanking region of the human cathepsin L gene that we sequenced: the primary primer was 5′-CTCTA-GGCTATGGATTATATCTTCC-3′ and the nested primer was 5′-GAAGATCTGGTAACCAAGTCGTTCTCTTGAGAT-3′. The underlined *Bgl*II site was added to the 5′-end of nested primers to facilitate subcloning. PCR products were analysed by electrophoresis on a 1% (w/v) agarose gel, purified from the gel with Sephaglas BandPrep kit (Amersham Biosciences), subcloned in pGL3 basic vector (Promega), digested with *Bgl*II and *Mlu*I and sequenced. Sequence analyses were performed by Genome Express.

Characterization of the transcription initiation site

Total RNA was isolated with Trizol (Life Technologies) and mRNA was purified with an Oligotex mRNA kit (Qiagen). To obtain full-length 5′ ends of the cathepsin L cDNA, full-length RNA-ligase-mediated rapid amplification of cDNA ends (RLM–RACE) was performed with the GeneRacer kit (Invitrogen). This kit ensures the amplification of only full-length transcripts by eliminating truncated message from the amplification process. The procedure was performed as described in the manufacturer's protocol by using mRNA of the human melanoma cell line DM-4 and a specific primer (PCL-PE3) of cathepsin L cDNA located downstream of the translation start site 5′-AAAGGCAGCAAGGATGAGTGTAGGATTCAT-3′. DNA fragments resulting from the RLM–RACE PCR were analysed by electrophoresis on a 3% (w/v) agarose gel, cloned with a TOPO TA Cloning kit (Invitrogen) and sequenced (Genome Express).

Construction of the reporter gene

Genomic DNA was prepared from human melanoma DM-4 cells with a Qiagen Genomic-tip system. The nucleotide upstream of the translation start site was numbered −1. The high-fidelity DNA polymerase Dynazyme EXT (Finnzymes) was used for PCR. The underlined restriction-enzyme site in each primer was added to the 5′ end to facilitate subcloning. The 3263 bp genomic fragment upstream of the translation start site was generated by PCR from the genomic DNA of DM-4 cells. A first PCR was performed with a primer upstream of the translation start site 5′-GAAGATCTGTTTTAAAACCTAGGAAGGGAAC-3′ (primer −1/−23) and a primer 5′-CCGACGCGTTTCCTCTA-TCTACAACATTTCTCCA-3′ complementary to −2366/−2390 fragment of the 5′ region of the cathepsin L gene. The PCR product was subcloned in the promoterless pGL3 basic vector (Promega) digested with *Bgl*II and *Mlu*I, generating the construct pGL3-PCL-(−1/−2390). A second PCR was performed with primers 5′-GAAGATCTTACTCTAGTCCAGTGGCTC-CTGTTT-3′ (−1891/−1915) and 5′-CCGACGCGTTTCT-ACACATTGTGACTGGTTCT-3′ (−3241/−3263). By using the *Xho*I restriction site (−1224/−1230) in the 5′ region of the cathepsin L gene, the second PCR product was digested by *Xho*I and *Bgl*II and subcloned in pGL3-PCL-(−1/−2390) to generate pGL3-PCL-(−1/−3263). Other constructs were generated from cathepsin L gene reporter vector. pGL3-PCL-(−1/−1388), pGL3-PCL-(−1/−1646) and pGL3-PCL-(−1/−2085) were obtained by PCR with primer −1/−23 and respectively primers 5′-CCGACGCGTGACAGGGACTGGAAGAGAG-GAC-3′ (−1367/−1388), 5′-CCGACGCGTAGTCAGTAAA-CAAGCCACGAAC-3′ (−1625/−1646), and 5′-CCGACGC-GTAGTCTCCTCCCTCCTAGGGTCT-3′ (−2064/−2085). pGL3-PCL-(−1465/−1646) was also obtained by PCR with primer 5′-GAAGATCTTCAAGGTTGCGGAGGTCGCGGT-TCT-3′ (−1465/−1489) and primer −1625/−1646. pGL3-PCL-(−1/−460), pGL3-PCL-(−1/−644) and pGL3-PCL-(−1/−1205) were constructed by digestion with restriction enzymes for *Stu*I, *Sma*I and *Kpn*I sites respectively. PGL3-GAPDH (in which GAPDH stands for glyceraldehyde-3-phosphate dehydrogenase) was constructed by subcloning the 1419 bp genomic fragment upstream of the translation start site of the gene encoding GAPDH in pGL3 basic vector digested by *Bgl*II and *Hin*dIII. Primers used for the PCR amplification of genomic DNA from DM-4 cells were 5′-GAAGATCTCGTGTG CCCA-AGACCTCTTT-3′ and 5′-CCCAAGCTTATGGTGTCTGAG-CGATGTGG-3′. pRL-CMV (in which CMV stands for cytomegalovirus) was purchased from Promega. Constructs were confirmed by sequence analysis (Genome Express).

Site-directed mutagenesis

Mutagenesis was performed with the QuickChange site-directed mutagenesis kit (Stratagene). The following oligonucleotides (single-stranded sense and anti-sense oligonucleotides) were used as primers: 5′-GCACCCAGAGTCCCCGaaCAGCTGCCGG-CACAGC-3′ for mutation in the GC.1 binding site, 5′-GCAGC-CAGGCGGCGGaaCGGTGCCGGCCGAACC-3′ for mutation in the GC.2 binding site and 5′-GCTGCCGGCACAGtCA-gTCGCAGCGCAGCC-3′ for mutation in the CCAAT motif. Mutated nucleotides are indicated by lower-case letters. In accordance with the manufacturer's instructions, PCRs were performed with wild-type reporter vector as template; the PCR products, digested by *Dpn*I, were used for transformation. The

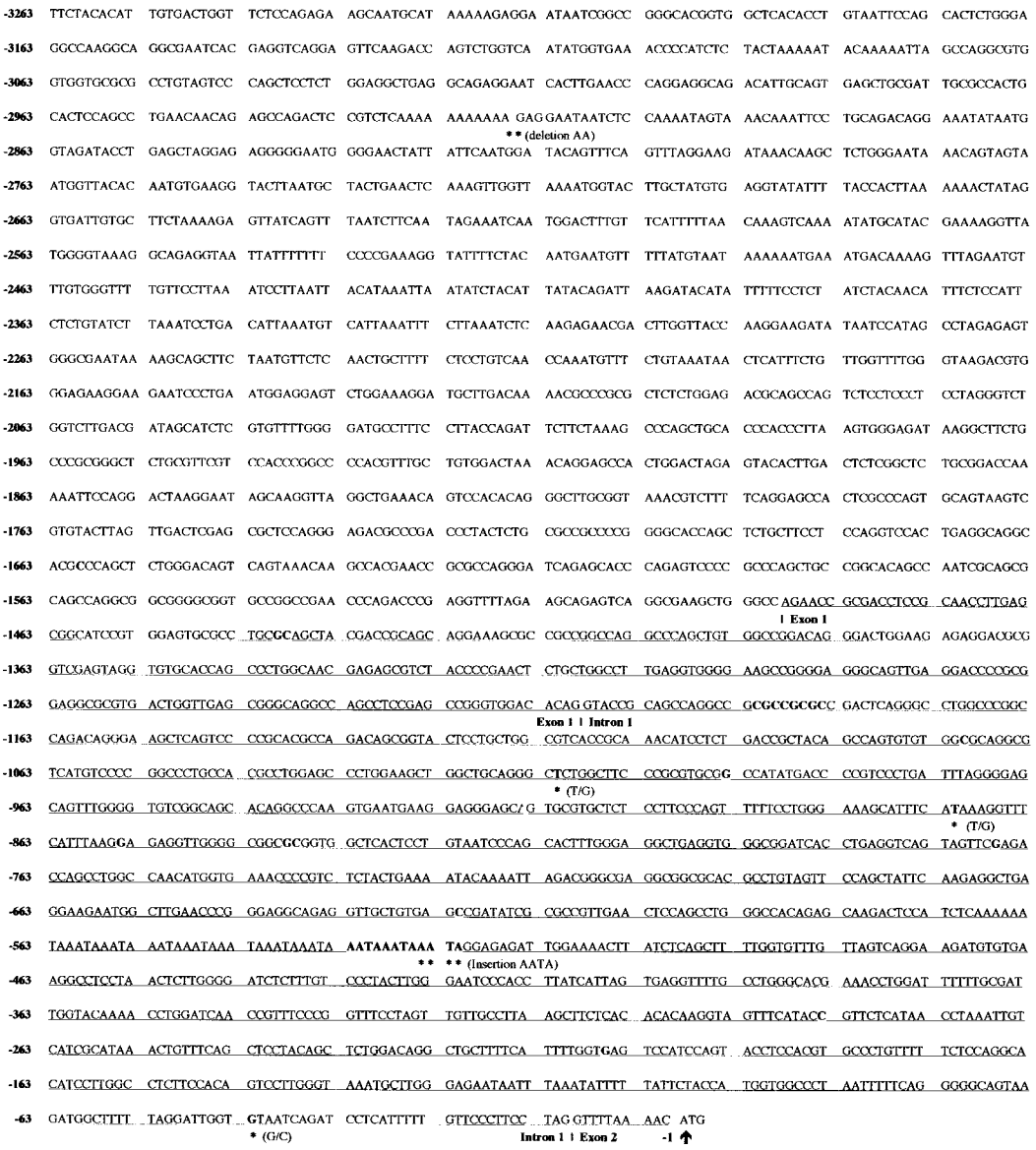

Figure 1 5′ region of the human cathepsin L gene

DNA sequence of the 3263 bp genomic fragment upstream of the translation start site of the human cathepsin L gene cloned from DM-4 human melanoma cells. The translation start codon is indicated with an arrow. The region of our sequence that corresponded to the sequence of exon 1, intron 1 and exon 2 published by Chauhan et al. [4] is underlined; within this region, sequences not underlined correspond to differences between both sequences. Differences between the genomic DNA of DM-4 cells and the working draft sequence segment of human chromosome 9 (NT_024008) are indicated by asterisks.

presence of mutations was verified by sequencing (Genome Express). Reporter vectors containing mutations were chosen for large-scale DNA preparation and used in transfection experiments. Oligonucleotides containing mutations were also annealed and used in electrophoretic mobility-shift assays (EMSAs) as probes or as competitors to verify the effect of mutations.

Transient transfections and luciferase assays

DM-4 and A375SM melanoma cells were seeded in 12-well tissue culture plates, cultured for 24 h and transfected by using 2 μl and 0.8 μl of the Lipofectin reagent (Life Technologies) respectively, in accordance with the manufacturer's protocol. Transfections were performed with 800 ng of pGL3 vector reporter constructs and 25 ng of pRL-CMV (Promega) as a control for transfection efficiency. After incubation overnight, the cells were washed and transfection medium was replaced for 24 h by serum-containing culture medium. The cells were then washed with PBS and harvested in Passive Lysis Buffer (Promega). After incubation for 30 min, cell extracts were centrifuged at 12000 ***g*** for 30 min to pellet the cell debris and were subjected to luciferase assays. Quantification of firefly and *Renilla* luciferase activities was performed with the Dual-Luciferase Reporter Assay System (Promega). The relative firefly luciferase activity was calculated by normalizing transfection efficiency with the *Renilla* luciferase activity. The experiments were performed in triplicate and similar results were obtained from at least three independent experiments.

Figure 2 5′ RLM–RACE analysis of transcription initiation sites

(**A**) 5′ RLM–RACE PCR products were analysed on an ethidium bromide-stained 3% (w/v) agarose gel (lane 2). The DNA molecular-size standard is indicated (lane 1; scale at the left). (**B**) DNA sequence of the cathepsin L gene (−1500 to +40). The underlined sequence corresponds to the PCL-PE3 primer used. Transcription initiation sites are identified by boxes. Arrows indicate the 5′ splice site and the asterisk indicates the 3′ splice site.

Preparation of nuclear extracts

DM-4 melanoma cells at 70–80% confluence were harvested by scraping in cold PBS and washed, then lysed with a Dounce tissue homogenizer in 10 mM Hepes buffer, pH 7.9, containing 1.5 mM $MgCl_2$, 10 mM KCl, 0.5 mM dithiothreitol and complete protease-inhibitor cocktail (Roche). After centrifugation, the nuclear pellet was resuspended in 20 mM Hepes buffer, pH 7.9, containing 25% (v/v) glycerol, 1.5 mM $MgCl_2$, 600 mM KCl, 0.2 mM EDTA, 0.5 mM dithiothreitol and protease inhibitors, then incubated on ice for 30 min. Soluble proteins were separated by centrifugation at 15000 ***g*** for 30 min. Nuclear extracts were dialysed in 20 mM Hepes buffer, pH 7.9, containing 20% (v/v) glycerol, 100 mM KCl, 0.2 mM EDTA, 0.5 mM dithiothreitol and protease inhibitors, then stored at −80 °C before use. Protein concentration was determined by the Bio-Rad protein assay.

EMSA

DNA fragments −1326/−1646 and −1326/−1489 of the 5′ region of the cathepsin L gene were amplified by PCR by using Dynazyme II (Finnzymes) with primers 5′-GAAGATCTGTTT-TAAAACCTAGGAAGGGAAC-3′ and 5′-CCGACGCGTA-GTCAGTAAACAAGCCACGAAC-3′ or 5′-CCGACGCGTA-GAACCGCGACCTCCGCAAC-3′ respectively. PCR products were then separated on an agarose gel and purified with a DNA Sephaglas Bandprep kit. Single-stranded sense and anti-sense oligonucleotides spanning the −1476/−1646 region of the cathepsin L gene promoter, oligonucleotides containing the consensus binding site for NF-Y and CCAAT-enhancer-binding protein (C/EBP) [28] and an oligonucleotide containing a mutated Sp1 binding site (5′-ATTCGATCGGTTGGGGGCG-AGC-3′) were synthesized by Sigma–Genosys; double-stranded oligonucleotides were prepared by annealing. Double-stranded oligonucleotides containing consensus binding sites for Sp1, CCAAT-binding transcription factor/nuclear factor 1 (CTF/NF-1) and activator protein 3 (AP-3) were purchased from Stratagene. Probes were labelled with T4 polynucleotide kinase (Life Technologies) and [γ-^{32}P]dATP (Amersham Biosciences) and unincorporated nucleotides were eliminated on a Microspin G-25 column (Amersham Biosciences). EMSAs were performed by incubating nuclear extracts (5 μg) for 30 min with ^{32}P-end-labelled double-stranded DNA (20000 c.p.m.) in binding reaction buffer [25 mM Hepes buffer (pH 8) containing 4% (v/v) glycerol, 50 mM NaCl, 0.5 mM EDTA, 0.5 mM dithiothreitol and 1 μg of poly(dI-dC)] on ice. Samples were resolved on 5% or 6% (w/v) non-denaturing polyacrylamide gel in 0.5 × Tris/borate/EDTA buffer. Gels were dried and exposed to X-ray film. In competition assays, nuclear extracts were preincubated with a 100-fold molar excess of unlabelled competitor oligonucleotide for 30 min and a binding reaction to the labelled probe was performed. In supershift assays, antibodies were added after the binding reaction and incubation was continued for 60 min on ice. The following antibodies were used in supershift analyses: anti-Sp1 and anti-Sp3 polyclonal antibodies (Santa Cruz Biotechnology), monoclonal antibody (mAb) anti-(NF-YA) (Pharmingen) and anti-p53 mAb (Oncogene). Polyclonal anti-body anti-(NF-YB) was kindly provided by Dr R. Mantovani (Università di Modena, Modena, Italy).

RESULTS

Cloning and characterization of the 5′-flanking region of the human cathepsin L gene

To sequence the 5′-flanking region of the human cathepsin L gene, we used a PCR-based method as described in the Materials and methods section. In brief, the use of specific primers of human cathepsin L cDNA led us to identify a first 930 bp sequence of the 5′ -flanking region of the human cathepsin L gene from a human genomic library. Then an additional 863 bp sequence of this region was determined by using the same protocol with primers specific for the first sequence identified above. These results allowed us to design specific primers and to clone by PCR the 5′ region of human cathepsin L from the genomic DNA of DM-4, a human melanoma cell line. Thus we

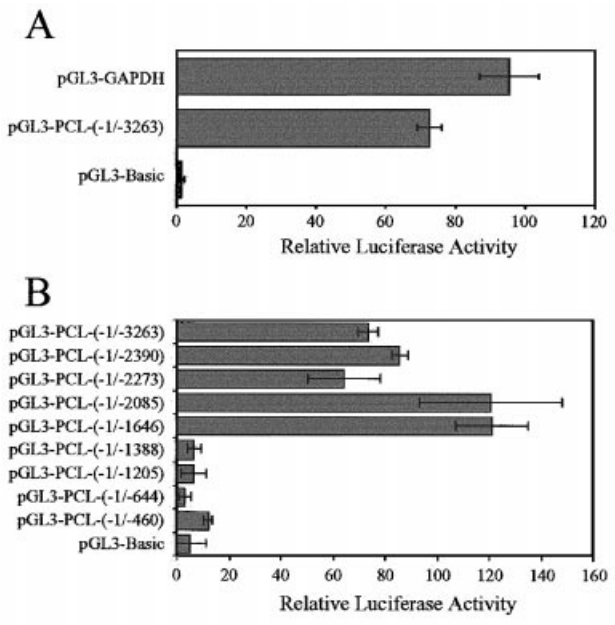

Figure 3 Promoter activity of the 5′ region of the human cathepsin L gene

Reporter vectors were transfected into DM-4 cells and firefly and *Renilla* luciferase activities were quantified with the Dual-Luciferase Reporter Assay System 2 days after transfection. PRL-CMV reporter vector, which contained the reporter gene *Renilla* luciferase driven by the CMV promoter, was introduced in each experiment to correct for variations of transfection efficiency. (**A**) A reporter vector containing the 5′ region of human cathepsin L subcloned in front of the firefly luciferase gene [pGL3-PCL-(−1/−3263)] was transfected into DM-4 cells. A promoterless vector (pGL3 Basic) and a reporter vector containing the GAPDH promoter (pGL3-GAPDH) were used as negative and positive controls respectively. (**B**) A 5′-deletion mutant of human cathepsin L promoter subcloned into the pGL3 vector was used for transient transfection. Results are means ± S.D.

determined the complete nucleotide sequence of the 3263 bp genomic fragment upstream of the translation start site of human cathepsin L (Figure 1). Comparison of the relevant part of our data with the sequences of exon 2, intron 1 and exon 1 described previously by Chauhan et al. [4] demonstrated differences that will be detailed in the Discussion section.

Next we focused on the identification of the transcription initiation site of the cathepsin L gene. For this purpose, full-length RLM–RACE was performed on mRNA isolated from DM-4 cells, with a specific primer of the coding sequence of cathepsin L cDNA (PCL-PE3). Three major DNA fragments resulting from the 5′ RLM–RACE PCR were obtained (Figure 2A). PCR products were cloned and 60 clones were analysed. DNA sequence analysis allowed us to localize the major transcription initiation site between 1489 and 1491 bp from the ATG translation start site (Figure 2B). Furthermore, in 13 of these clones, an additional transcription initiation site was identified between 1371 and 1382 bp from the translation start site (Figure 2B). Three splice variants were also identified in the 5′ untranslated region of cathepsin L mRNA. As shown in Figure 2(B), alternative splicing of cathepsin L mRNA in DM-4 cells occurred with a 5′ splice site in the three positions 1210, 1300 and 1355 bp from the translation start site. We observed that the frequency of splice variants was the same and splicing occurred independently of the transcription initiation site. Taking into account the transcription initiation site at −1489 bp from the translation start site and a splicing involving one of the three 5′ splice sites, the size of the 5′ RLM–RACE PCR products was 220, 275 and 365 bp, corresponding to the size of the three major DNA fragments observed on the agarose gel (Figure 2A).

Taken together, these data constitute the structural basis for the study of the regulation of the human cathepsin L gene promoter.

Regulatory sequences within human cathepsin L gene promoter

To analyse the promoter activity of the cloned 5′ region, two sets of experiments were performed. First, the 3263 bp sequence upstream of the translation start site of the cathepsin L gene was subcloned in front of the luciferase reporter gene into the pGL3-Basic vector [pGL3-PCL-(−1/−3263)]. This construct was co-transfected into DM-4 cells with the pRL-CMV plasmid, used as an internal control to measure the transfection efficiency. Promoter activity was analysed with the Dual-Luciferase Reporter System. As shown in Figure 3(A), a marked increase in luciferase activity was observed in cells transfected with the pGL3-PCL-(−1/−3263) vector, in comparison with the pGL3-Basic vector. In DM-4 cells, promoter activity of this 5′ region of the human cathepsin L gene was similar to the GAPDH promoter activity used as positive control. Secondly, to define the DNA regions involved in cathepsin L gene expression, we constructed 5′-deletion mutants of the cathepsin L promoter in the reporter vector pGL3-Basic. These constructs were transfected into DM-4 cells and luciferase activities were quantified (Figure 3B). No significant promoter activity was detected with reporter vectors that contained the first 1388 bp sequence upstream of the translation start site of the cathepsin L gene, in comparison with luciferase activity obtained with the pGL3-Basic vector. However, a strong luciferase activity was detected when DM-4 cells were transfected with the vector pGL3-PCL-(−1/−1646). These results demonstrate that the 3263 bp sequence upstream of the translation start site of the cathepsin L gene had a strong promoter activity and that the region between −1388 and −1646 bp was crucial for cathepsin L gene transcription. Furthermore, the decrease in luciferase activity observed with the construct pGL3-PCL-(−1/−2273) suggested the presence of negative regulator sequences between −2085 and −2273.

Taking into account the fact that the major transcription initiation site was identified between −1489 and −1491 from the translation start site and the above data, we focused on the identification of transcription factors that bound to the proximal promoter −1489/−1646 of the cathepsin L gene. For this purpose, EMSA was performed with nuclear extracts of DM-4 cells and, as a probe, a labelled DNA fragment corresponding to −1326/−1646 sequence of cathepsin L promoter. As shown in Figures 4(A) and 4(B) (lanes 2), five DNA–protein complexes were detected. These DNA–protein interactions (Figure 4A) were specific, because they were totally inhibited by an excess of unlabelled fragment −1326/−1646 (lane 3) and were not affected by an excess of a non-specific fragment corresponding to the −1/−313 DNA sequence of the cathepsin L gene promoter (lane 4). Complexes 1 and 5 were abolished in presence of oligonucleotide −1326/−1489 (Figure 4A, lane 5). The identification of regions within cathepsin L proximal promoter involved in the formation of complexes 2, 3 and 4 was performed by competition studies with 30 bp oligonucleotides spanning the −1476/−1646 region of the cathepsin L gene promoter (Figure 4B). The formation of complexes 2 and 4 was inhibited by oligonucleotides −1576/−1606 and −1536/−1566 (Figure 4B, lanes 6 and 8). The band intensity of complex 3 decreased with oligonucleotides −1576/−1606, −1556/−1586 and −1536/−1566 (Figure 4B, lanes 6 to 8); other oligonucleotides (lanes 3–5 and 9–11) had no effect. The presence of consensus binding sites

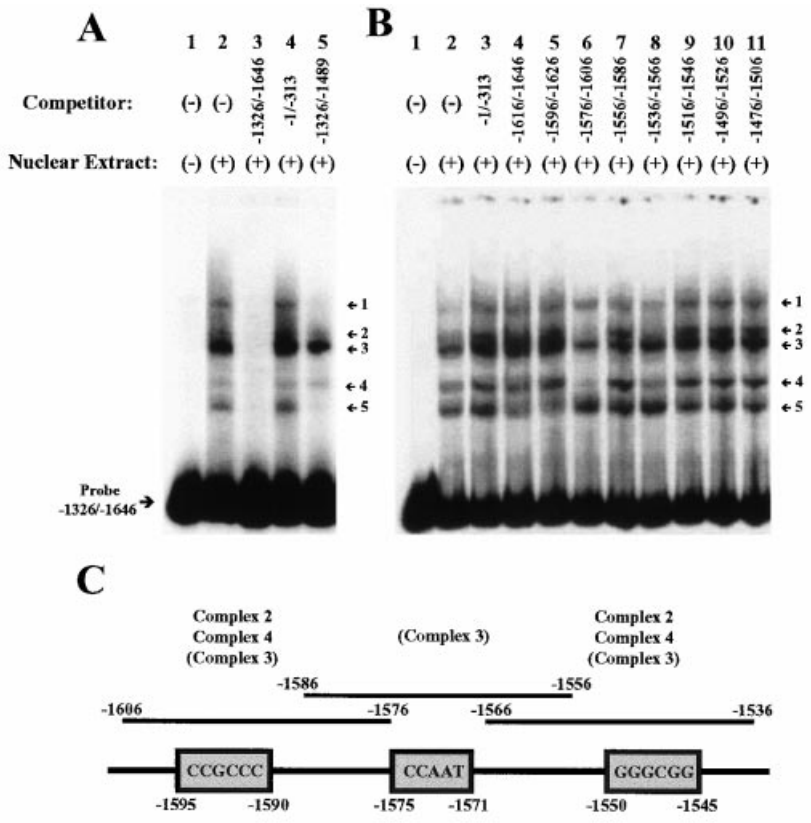

Figure 4 Formation of specific DNA–protein complexes within the region −1489/−1646 of human cathepsin L promoter

EMSAs were performed with nuclear extracts isolated from DM-4 human melanoma cells and, as a probe, the radiolabelled −1326/−1646 fragment of the human cathepsin L gene promoter (**A** and **B**). Competition was performed with unlabelled fragments of human cathepsin L gene promoter in the presence of 100-fold molar excess of either large oligonucleotides (**A**) or 30 bp oligonucleotides spanning the −1476/−1646 region (**B**). Specific DNA–protein complexes numbered 1–5 are indicated with arrows. (**C**) Schematic representation and localization of GC boxes and the CCAAT motif.

of transcription factors on the inhibiting oligonucleotides was analysed. As shown in Figure 4(C), oligonucleotides −1576/−1606 and −1536/−1566 had a GC box at positions −1590/−1595 and −1545/−1550 respectively, and oligonucleotide −1556/−1586 contained a CCAAT motif at position −1571/−1575.

Sp1 and Sp3 transcription factors regulate the human cathepsin L gene promoter

The GC box was described as a binding site for Sp1 and Sp3, which belong to the Sp family of transcription factors [29]. Thus we performed complementary experiments to analyse the possible contribution of members of this transcription-factor family in melanoma cells. First, we analysed whether GC boxes present in oligonucleotides −1576/−1606 and −1536/−1566 were involved in the formation of complexes 2, 3 and 4. For this purpose, competition studies were performed with the labelled fragment −1326/−1646 as a probe and nuclear extracts from DM-4 cells. As shown in Figure 5(A), preincubation with an oligonucleotide containing the Sp1 consensus binding site completely inhibited the formation of complexes 2 and 4 and, in part, complex 3 (lane 3), whereas preincubation with oligonucleotides containing the CTF/NF-1 consensus binding site or a double-mutated Sp1-binding site did not modify the formation of these complexes (lanes 4 and 5). Secondly, the involvement of Sp1 and Sp3 transcription factors in the formation of these complexes was analysed by using specific antibodies in supershift studies. As shown in Figure 5(B), the formation of complexes 2 and 4 was inhibited in the presence of a specific anti-Sp3 antibody (lane 4), whereas the band intensity of complex 3 decreased slightly in the presence of a specific anti-Sp1 antibody (lane 3). Thirdly, the presence of an active transcription factor in DM-4 cells that bound to the Sp1 consensus binding site was analysed. EMSA was performed by using radiolabelled consensus Sp1 oligonucleotide and nuclear extracts prepared from these cells. As shown in Figure 5(C), three DNA–protein complexes were identified, namely two slower-migrating bands (complexes 1 and 2) and a faster-migrating band (complex 3) (lane 2). These interactions were specific because they were totally inhibited by unlabelled consensus Sp1 oligonucleotide (Figure 5C, lane 3) and not by unlabelled consensus CTF/NF-1 oligonucleotide (lane 4). In addition, preincubation of these samples with specific anti-Sp1 antibody (Figure 5C, lane 5) supershifted complex 1, and preincubation with specific anti-Sp3 antibody (lane 6) inhibited the formation of complexes 2 and 3. Fourthly, the binding of Sp1 and Sp3 present in DM-4 nuclear extracts to each GC box present in the cathepsin L promoter was analysed. EMSAs were performed with radiolabelled oligonucleotides −1576/−1606 (Figure 5D) or −1536/−1566 (Figure 5E) as probes. The DNA–protein complexes, which presented a similar specific pattern to that observed with the Sp1 consensus oligonucleotide

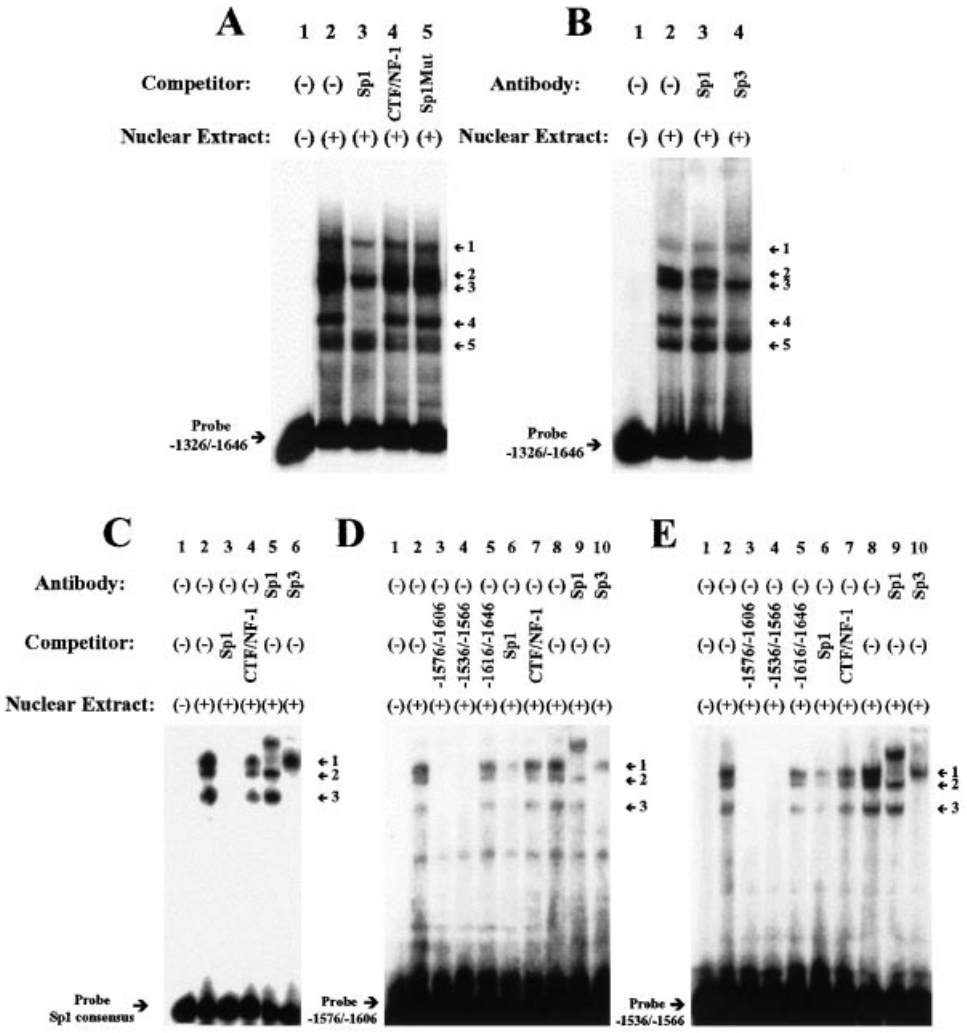

Figure 5 Interactions of Sp1 and Sp3 with GC boxes of human cathepsin L promoter

(**A** and **B**) EMSAs were performed with DM-4 nuclear extracts and the radiolabelled −1326/−1646 fragment of the human cathepsin L promoter as a probe; (**A**) competition studies and (**B**) supershift studies. Other probes used for EMSAs were the oligonucleotide containing the Sp1 consensus binding site (**C**) and the two oligonucleotides −1576/−1606 (**D**) and −1536/−1566 (**E**) corresponding to fragments of cathepsin L human promoter. To determine the specificity of the DNA–protein interaction, competition studies were performed in the presence of a 100-fold molar excess of the indicated oligonucleotides. Specific polyclonal antibody against anti-Sp1 or anti-Sp3 was used to identify proteins present in DNA–protein complexes. Specific numbered DNA–protein complexes are indicated at the right.

(Figures 5D and 5E, lane 2), were totally inhibited by unlabelled oligonucleotide −1576/−1606 (lanes 3) or −1536/−1566 (lanes 4) and by the Sp1 consensus oligonucleotide (lanes 6), but not modified by the non-specific oligonucleotide −1616/−1646 (lanes 5) or the CTF/NF-1 consensus oligonucleotide (lanes 7). Incubation with specific anti-Sp1 or anti-Sp3 antibody induced similar modifications of DNA–protein complexes to that observed with the Sp1 consensus sequence (Figures 5D and 5E, lanes 9 and 10). Fifthly, the functional contribution of these two GC boxes on cathepsin L promoter activity was studied. Mutational analysis was performed in DM4 and A375SM, two human melanoma cell lines. As shown in Figure 6(A), mutations were introduced in each GC box (GC.1 and GC.2). EMSA allowed us to verify that oligonucleotides containing these mutations did not form any complexes with nuclear proteins of DM-4 cells and did not inhibit the binding of Sp1 and Sp3 to oligonucleotides −1576/−1606 and −1536/−1566 (results not shown). The effects of GC-box mutations on cathepsin L promoter activity were analysed by using pGL3-PCL-(−1/−1646), i.e. the luciferase reporter construct containing the 1646 bp sequence upstream of the translation start site of the cathepsin L gene. Results demonstrated that mutations in both GC boxes inhibited up to 40% of the promoter activity in A375SM and DM-4 cells, in comparison with the wild-type promoter. We also analysed the effects of mutations in GC boxes on the promoter activity of a fragment of the cathepsin L promoter that initiated close to the major transcription initiation site, identified above. For this purpose we used the construct pGL3-PCL-(−1465/−1646), wild-type or containing mutated GC boxes, in transfection assays. Luciferase assays demonstrated that mutations of both GC boxes in this construct decreased promoter activity by 65% and 60% in DM-4 and A375SM cells respectively. Mutations in either the GC.1 or the GC.2 box decreased promoter activity by approx. 30%, suggesting that each GC box contributed equally to cathepsin L promoter activity.

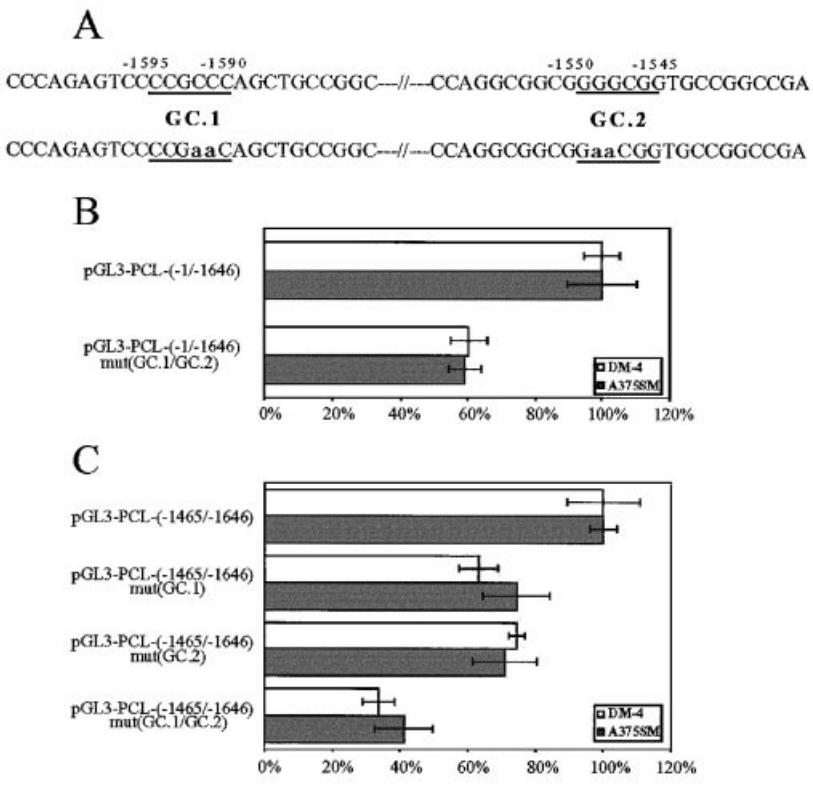

Figure 6 Effect of GC-box mutations on human cathepsin L promoter activity

(**A**) Schematic representation of nucleotide modifications between the wild-type and mutant GC boxes. (**B** and **C**) GC boxes individually (GC.1 and GC.2) or in combination (GC.1/GC.2) were mutated in luciferase reporter vectors containing a fragment of cathepsin L promoter: pGL3-PCL-(−1/−1646) (**B**) or pGL3-PCL-(−1465/−1646) (**C**). Transient transfections were performed with these constructs in DM-4 and A375SM melanoma cells, and promoter activities were determined with the Dual-Luciferase Reporter Assay System. The PRL-CMV reporter vector was introduced in each experiment to correct for variations in transfection efficiency. All activities are presented as percentages of wild-type promoter activity. Results are means ± S.D.

Taken together, these results show clearly that, in human melanoma cells, cathepsin L promoter activity was under the control of the two GC boxes in positions −1590/−1595 and −1545/−1550, to which the transcription factors Sp1 and Sp3 bound specifically.

NF-Y transcription factor regulates human cathepsin L gene promoter

As mentioned above, the oligonucleotide −1556/−1586 contained a CCAAT motif, which could be recognized by several transcription factors. In addition, the above demonstration that mutations in GC boxes did not totally inhibit cathepsin L promoter activity also supported the notion that other transcription factors were involved in its regulation. Indeed, complex 3 obtained with the fragment −1326/−1646 could contain at least two different DNA–protein complexes, one formed with Sp1 and one with protein that bound to oligonucleotide −1556/−1586. First, we verified this point by EMSA with labelled oligonucleotide −1326/−1646 in the presence of an excess of unlabelled consensus Sp1 oligonucleotide. As shown in Figure 7(A), in the presence of the consensus Sp1 oligonucleotide, nuclear extracts induced a pattern of two DNA–protein complexes (lane 2); in agreement with the results described above, complexes 2 and 4 were not detected because they resulted from the interaction of Sp3 with the fragment −1326/−1646. Complex 3 was totally inhibited by oligonucleotide −1556/−1586 (Figure 7A, lane 6) and not by other 30 bp oligonucleotides spanning the −1476/−1646 region of the cathepsin L gene promoter (lanes 3–5 and 7–10).

Secondly, this result and the property of the CCAAT motif to be recognized by such transcription factors as C/EBP, NF-Y or CTF/NF-1 [30] led us to identify the transcription factor that bound to fragment −1326/−1646. For this purpose, competition and supershift analyses were performed in the presence of the consensus Sp1 oligonucleotide. As shown in Figure 7(B), only the NF-Y consensus binding site competed with the complex 3 (lane 1), whereas the consensus binding site for CTF/NF-1, C/EBP or AP-3 had no effect (lanes 2–4). Because NF-Y results from the interaction of three subunits (NF-YA, NF-YB and NF-YC), we incubated the above samples with an anti-(NF-YA) mAb, which induced a supershift of complex 3 (Figure 7B, lane 7). The control, an anti-p53 mAb, had no effect (Figure 7B, lane 8). In addition, EMSA was performed with labelled oligonucleotide −1556/−1586 and DM-4 nuclear extracts. As shown in Figure 7(C), only one DNA–protein complex was detected (lane 2). This complex was specifically inhibited by the unlabelled oligonucleotide −1556/−1586 (Figure 7C, lane 3) and not by the non-specific oligonucleotide −1616/−1646 (lane 4). The complex was totally inhibited by unlabelled NF-Y consensus binding site (Figure 7C, lane 5) but was not modified by consensus binding sites for CTF/NF-1, C/EBP or AP-3 (lanes 6–8).

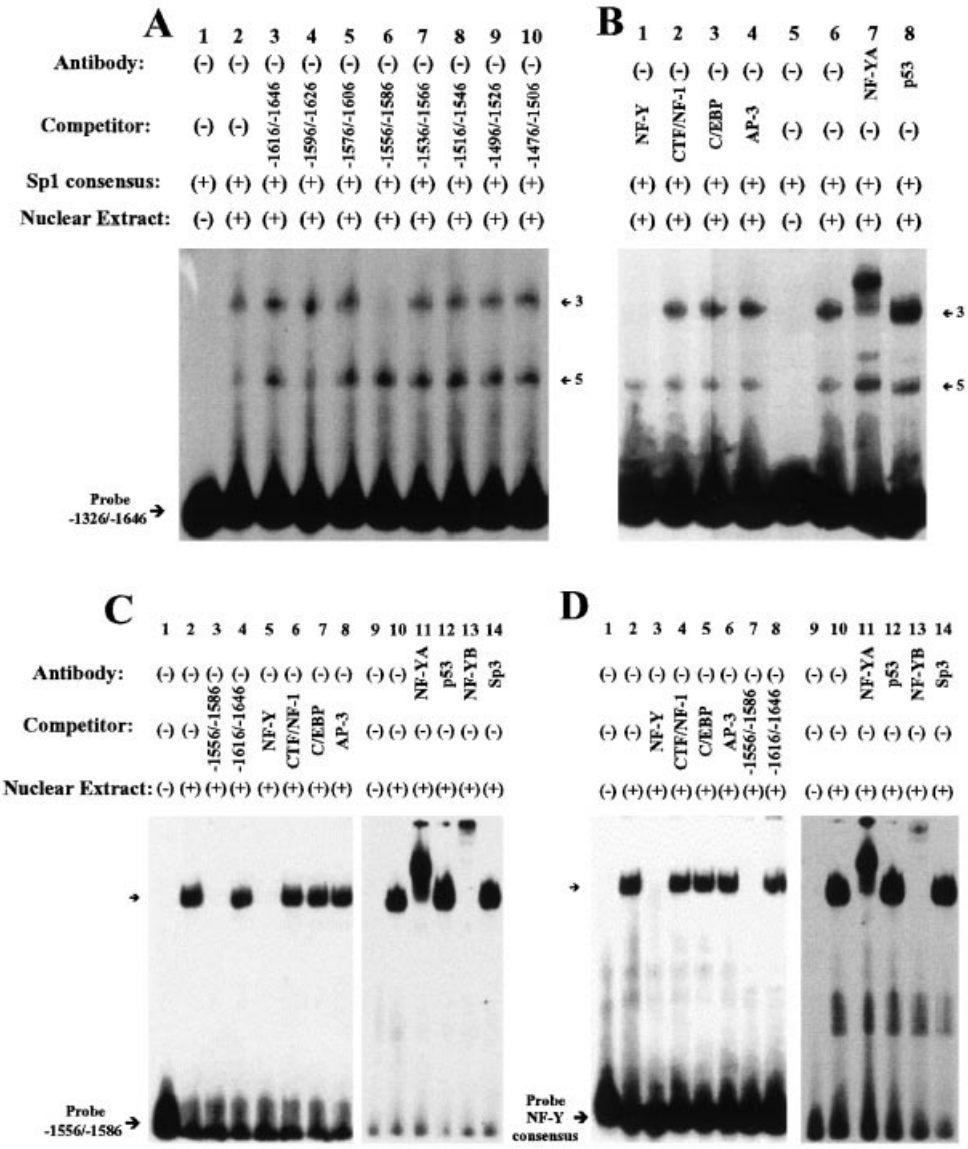

Figure 7 Interaction of NF-Y with the CCAAT motif of human cathepsin L promoter

(**A** and **B**) EMSAs were performed by incubating DM-4 nuclear extracts with the radiolabelled −1326/−1646 fragment of the human cathepsin L promoter in the presence of an excess of an unlabelled oligonucleotide containing the Sp1 consensus binding site. The following competitors were also added: fragments of human cathepsin L promoter (**A**) or oligonucleotides containing consensus binding sites for the identified transcription factors (**B**). (**C** and **D**) EMSAs were also performed by incubating DM-4 nuclear extracts with radiolabelled probes: the −1556/−1586 fragment of the human cathepsin L promoter (**C**) or an oligonucleotide containing the NF-Y consensus binding site (**D**). For competition studies the indicated competitors were used at 100-fold molar excess. In supershift studies, the indicated specific antibodies were added to the samples. Numbered specific DNA–protein complexes are indicated with arrows.

Furthermore, the addition of anti-(NF-YA) mAb supershifted this complex (Figure 7C, lane 11), whereas the control anti-p53 mAb had no effect (lane 12). This complex formation was inhibited by a polyclonal anti-(NF-YB) antibody (Figure 7C, lane 13) and not by polyclonal anti-Sp3 antibody (lane 14). Identical results were obtained in EMSAs with an oligonucleotide containing the NF-Y consensus binding site as a probe (Figure 7D). Taken together, these results demonstrate clearly that NF-Y transcription factor specifically recognized the CCAAT site present in the human cathepsin L gene promoter.

Thirdly, the function of NF-Y binding to the CCAAT motif was analysed by site-directed mutagenesis and reporter assay, with the two human melanoma cell lines A375SM and DM-4. Two mutations were introduced into the CCAAT motif (Figure 8A). EMSA allowed us to verify that the oligonucleotide containing these two mutations did not bind NF-Y and failed to inhibit NF-Y binding to oligonucleotide −1556/−1586 (results not shown). Mutations in the CCAAT motif were tested either alone or in combination with mutations in both GC boxes with the luciferase reporter construct pGL3-PCL-(−1/−1646) (Figure 8B). Mutation of the CCAAT motif decreased promoter activity by 30% and 35% in DM-4 and A375SM cells respectively; mutations of the CCAAT motif and GC boxes together decreased the activity of the cathepsin L promoter by 85% and 70% in DM-4 and A375SM cells respectively. These results demonstrate that, in addition to binding sites for Sp1 and

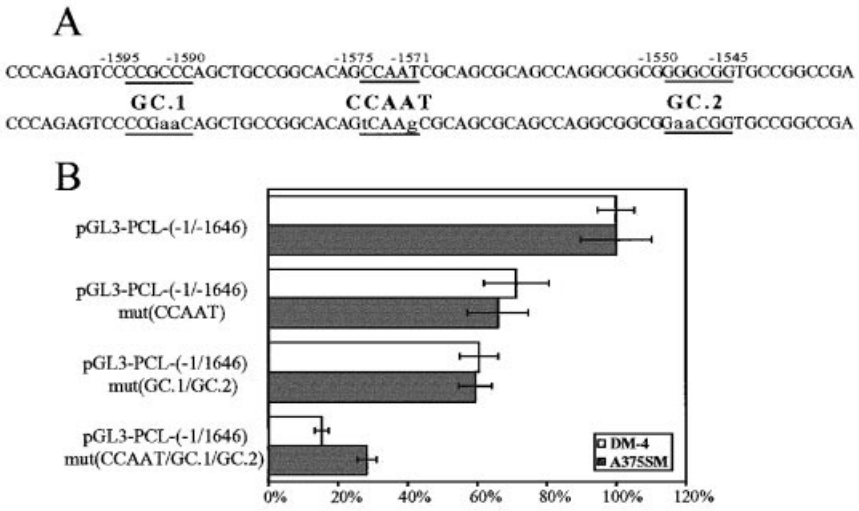

Figure 8 Effect of CCAAT motif mutations on human cathepsin L promoter activity

(**A**) Schematic representation of nucleotide modifications between the wild-type and mutant GC boxes and the CCAAT motif. (**B**) Mutations were introduced in luciferase reporter vector containing a fragment of the cathepsin L promoter pGL3-PCL-(−1/−1646). Constructs were transiently transfected into DM-4 and A375SM melanoma cells, and promoter activities were determined with the Dual-Luciferase Reporter Assay System. PRL-CMV reporter vector was introduced in each experiment to correct for variations of transfection efficiency. All activities are presented as percentages of wild-type promoter activity. Results are means ± S.D.

Sp3, the NF-Y-binding site is also important for cathepsin L promoter activity in human melanoma cells.

DISCUSSION

The demonstration that the enforced expression and secretion of procathepsin L by human melanoma cells contribute to tumour growth and metastasis [18] and the few results, mainly obtained in rodents, on molecular mechanisms responsible for regulation of cathepsin L transcription, led us to identify new and basic elements that were involved in the regulation of human cathepsin L expression.

For this purpose, we first identified the 5′-flanking region of the human cathepsin L gene, up to 3263 bp upstream of the translation start site. Comparison of the corresponding part of our data with the sequence of exon 1, intron 1 and exon 2 (1489 bp upstream of the translation start site), previously described by Chauhan et al. [4], demonstrated differences at 15 distinct positions. However, 12 of these 15 differences were not found between our sequence and that of a working draft sequence segment of human chromosome 9 (NT_024008.2), on which the cathepsin L gene locus was previously mapped, i.e. chromosome 9q21–22 [4]. Indeed, these two sequences share up to 99% identity; differences, found at only five positions, were more probably due to polymorphism or analytical errors.

In addition, in the human melanoma DM-4 cell line we identified the major and a minor transcription initiation sites of the human cathepsin L gene, localized between 1489–1491 and 1371–1382 bp from the translation start site respectively. Our results demonstrate that alternative splicing occurred in DM-4 melanoma cells and allowed us to identify three cathepsin L splice variants: one new had a 5′ splice site at position −1355 generating a small transcript, and two were identical with those described by others [4,31], with a 5′ splice site at positions −1300 and −1210. Indeed, the two first species of human cathepsin L mRNA were described as hCATL-A [23] and hCATL-B [24]: hCATL-A resulted from the splicing of intron 1 between −11 and −1210 from the translation start site; no splicing occurred in hCATL-B because the 5′ end of hCATL-B was identical with the 3′ end of intron 1 [4]. Rescheleit et al. [31] also identified two other splice variants in the 5′ untranslated region of cathepsin L mRNA, one form lacking 27 nt (hCATL-A I) and another lacking 90 nt (hCATL-A II) in exon 1 in comparison with hCATL-A [4,23], suggesting a 5′ splice site at positions −1300 and −1237. Analysis of the 50 bp DNA sequence downstream of the translation start site of the cathepsin L gene with the human expressed sequence tag (EST) database (Megablast; National Centre for Biotechnology Information) confirmed that we had identified the three most common splice variants in the 5′ non-coding sequence of cathepsin L mRNA; 50% of Blast hits corresponded to the new splice variant that we identified at position −1355. In contrast, splicing similar to hCATL-B [24] or hCATL-A II [31] was not found in the EST database. Thus our results show that in DM-4 melanoma cells, several cathepsin L mRNA species differed at their 5′ untranslated ends, owing to the selection of alternative transcription initiation sites and to differences in mRNA splicing.

We characterized the proximal promoter of the cathepsin L gene and identified sites essential for its regulation. The absence of typical mammalian TATA box close to the major transcription initiation site suggested that human cathepsin L gene has a TATA-less promoter. One of the major findings was to demonstrate that a 50 bp region, containing one CCAAT motif and two GC boxes and located 60 bp from the major transcription initiation site, contained transcription-factor-binding sites required for cathepsin L promoter activity. The CCAAT motif was found in cathepsin L promoter in a direct orientation at 82 bp from the major transcription initiation site. Interestingly, in the higher eukaryotic proximal promoter of many genes, the CCAAT motif has typically been described between 80 and 100 bp upstream of the transcription initiation site [30], in both direct and inverted orientations. Here we have clearly identified NF-Y as the protein that bound specifically to the CCAAT motif present in the cathepsin L promoter. NF-Y, also named CBF or

CP1, is a heterotrimeric transcription factor composed of three subunits (NF-YA, NF-YB and NF-YC) all necessary for DNA binding [32,33]. CTF/NF-1 and C/EBP, two other CCAAT-binding proteins [30], did not interact with the CCAAT motif present in cathepsin L promoter in DM-4 cells. This is the first demonstration that the NF-Y binding site is involved in the transcriptional regulation of a member of the papain-type cysteine protease family.

Furthermore, close to this CCAAT motif, we identified two GC boxes that were recognized by members of the Sp family of transcription factors (Sp1, Sp2, Sp3 and Sp4). We demonstrated that Sp1 and Sp3 transcription factors are involved in the regulation of the cathepsin L promoter in DM-4 melanoma cells. Further studies are needed to determine whether other members of the Sp family might be involved in regulation of the cathepsin L promoter. In addition, the respective functions of Sp1 and Sp3 in cathepsin L promoter activity remain to be explained. Indeed, Sp1 is described as a general activator of transcription, whereas Sp3 can act as an activator or as a repressor of Sp1-mediated activation, depending on the sequence context, the number of Sp1-binding sites and the availability of specific co-activators, co-repressors or other transcription factors [29]. The presence of an Sp1-binding site on the rat cathepsin L promoter has been suggested, in a very preliminary study [22], without identification of the transcription factor that might interact specifically with this site and without functional studies to demonstrate the relevance of this site on rat cathepsin L promoter activity.

The proximity of the two GC boxes near the CCAAT motif in the cathepsin L promoter suggested that Sp1 and/or Sp3 might co-operate with NF-Y in the regulation of cathepsin L promoter. Indeed, NF-Y might regulate the transcription of various promoters by co-operative interactions with other transcription factors [29]. Functional synergism between NF-Y and Sp1 was also described in other genes such as the major histocompatibility complex class II-associated invariant chain [34], rat fatty acid synthase [35] and human tissue inhibitor of metalloproteinases 2 [36]. The promoters of these genes have in common one or several Sp1-binding sites located close to (20–30 nt) a CCAAT motif. Direct interactions between both Sp1 and Sp3 and subunit NF-YA of NF-Y have recently been demonstrated [37,38]. These interactions depended on the expression level of each transcription factor, because for example the expression of NF-Y subunits varied during the growth and differentiation of individual cell lineages [33,39].

In conclusion, in the present study we characterized the human cathepsin L promoter and a specific region, upstream of the initiation site, containing a CCAAT motif and two GC boxes involved in the regulation of this promoter. In addition, we clearly identified three transcription factors, NF-Y, Sp1 and Sp3, that bound to these regulator elements of cathepsin L promoter. Further studies will allow us to determine the contribution of these transcription factors on cathepsin L overexpression in human melanoma cells.

We thank Gérard Drevet for technical assistance, and Dr Monique Barel for reading the manuscript. This work was supported by INSERM, Ministère de l'Education Nationale et de la Recherche, Association de Recherche Contre le Cancer (ARC, Villejuif), Ligue National Contre le Cancer (Comité de Paris) and Fondation de France.

REFERENCES

1 Ishidoh, K., Towatari, T., Imajoh, S., Kawasaki, H., Kominami, E., Katunuma, N. and Suzuki, K. (1987) Molecular cloning and sequencing of cDNA for rat cathepsin L. FEBS Lett. **223**, 69–73

2 Troen, B. R., Gal, S. and Gottesmann, M. M. (1987) Cloning and expression of the gene for the major excreted protein of transformed mouse fibroblasts. A secreted lysosomal protease regulated by transformation. J. Biol. Chem. **263**, 254–261

3 Turk, B., Turk, D. and Turk, V. (2000) Lysosomal cysteine proteases: more than scavengers. Biochem. Biophys. Acta **1477**, 98–111

4 Chauhan, S. S., Popescu, N. C., Ray, D., Fleischmann, R., Gottesman, M. M. and Troen, B. R. (1993) Cloning, genomic organization, and chromosomal localization of human cathepsin L. J. Biol. Chem. **268**, 1039–1045

5 Ishidoh, K. and Kominami, E. (1998) Gene regulation and extracellular functions of procathepsin L. Biol. Chem. **379**, 131–135

6 Gal, S. and Gottesmann, M. M. (1986) The major excreted protein (MEP) of transformed mouse cells and cathepsin L have similar protease specificity. Biochem. Biophys. Res. Commun. **139**, 156–162

7 Kirschke, H. and Barrett, A. J. (1987) Lysosomal cysteine proteases. In Lysosomes. Their Role in Protein Breakdown (Glaumann, H. and Ballard, F. J., eds.), pp. 193–238, Academic Press, London

8 Jean, D., Hermann, J., Rodrigues-Lima, F., Barel, M., Balbo, M. and Frade, R. (1995) Identification on melanoma cells of p39, a cysteine proteinase that cleaves C3, the third component of complement: amino-acid-sequence identities with procathepsin L. Biochem. J. **312**, 961–969

9 Jean, D., Rodrigues-Lima, F., Cassinat, B., Hermann, J., Cabane, J. and Frade, R. (1997) Co-expression and secretion of C3, the third component of complement and a C3-cleaving cysteine proteinase in a highly metastatic human melanoma cell line. Immunol. Lett. **58**, 107–112

10 Felbor, U., Dreier, L., Bryant, R. A., Ploegh, H. L., Olsen, B. R. and Mothes, W. (2000) Secreted cathepsin L generates endostatin from collagen XVIII. EMBO J. **19**, 1187–1194

11 Rozhin, J., Wade, R. L., Honn, K. V. and Sloane, B. F. (1989) Membrane-associated cathepsin L: a role in metastasis of melanomas. Biochem. Biophys. Res. Commun. **164**, 556–561

12 Maciewicz, R. A., Wardale, R. J., Etherington, D. J. and Paraskeva, C. (1989) Immunodetection of cathepsins B and L present in and secreted from human pre-malignant and malignant colorectal tumour cell lines. Int. J. Cancer **43**, 478–486

13 Kirschke, H., Eerola, R., Hopsu-Havu, V. K., Bromme, D. and Vuorio, E. (2000) Antisense RNA inhibition of cathepsin L expression reduces tumorigenicity of malignant cells. Eur. J. Cancer **36**, 787–795

14 Jean, D., Bar-Eli, M., Huang, S., Xie, K., Rodrigues-Lima, F., Hermann, J. and Frade, R. (1996) A cysteine proteinase, which cleaves human C3, the third component of complement, is involved in tumorigenicity and metastasis of human melanoma. Cancer Res. **56**, 254–258

15 Gottesmann, M. M. and Sobel, M. E. (1980) Tumor promoters and Kirsten sarcoma virus increase synthesis of a secreted glycoprotein by regulating levels of translatable mRNA. Cell **19**, 449–455

16 Chauhan, S. S., Goldstein, L. J. and Gottesman, M. M. (1991) Expression of cathepsin L in human tumors. Cancer Res. **51**, 1478–1481

17 Frade, R., Rodrigues-Lima, F., Huang, S., Xie, K., Guillaume, N. and Bar-Eli, M. (1998) Procathepsin L, a proteinase that cleaves human C3 (the third component of complement), confers high tumorigenic and metastatic properties to human melanoma cells. Cancer Res. **58**, 2733–2736

18 Frade, R. (1999) Structure and functions of proteases which cleave human C3 and are expressed on normal or tumor human cells: some are involved in tumorigenic and metastatic properties of human melanoma cells. Immunopharmacology **42**, 39–45

19 Troen, B. R., Ascherman, D., Atlas, D. and Gottesman, M. M. (1988) Cloning and expression of the gene for the major excreted protein of transformed mouse fibroblasts. A secreted lysosomal protease regulated by transformation. J. Biol. Chem. **263**, 254–261

20 Ishidoh, K., Kominami, E., Suzuki, K. and Katunuma, N. (1989) Gene structure and 5′-upstream sequence of rat cathepsin L. FEBS Lett. **259**, 71–74

21 Troen, B. R., Chauhan, S. S., Ray, D. and Gottesman, M. M. (1991) Downstream sequences mediate induction of the mouse cathepsin L promoter by phorbol esters. Cell Growth Differ. **2**, 23–31

22 Ishidoh, K., Taniguchi, S. and Kominami, E. (1997) Egr family member proteins are involved in the activation of the cathepsin L gene in v-src-transformed cells. Biochem. Biophys. Res. Commun. **238**, 665–669

23 Gal, S. and Gottesman, M. M. (1988) Isolation and sequence of a cDNA for human pro-(cathepsin L). Biochem. J. **253**, 303–306

24 Joseph, L. J., Chang, L. C., Stamenkovich, D. and Sukhatme, V. P. (1988) Complete nucleotide and deduced amino acid sequences of human and murine preprocathepsin L. An abundant transcript induced by transformation of fibroblasts. J. Clin. Invest. **81**, 1621–1629

25 Zhang, R. D., Price, J. E., Schackert, G., Itoh, K. and Fidler, I. J. (1991) Malignant potential of cells isolated from lymph node or brain metastases of melanoma patients and implications for prognosis. Cancer Res. **51**, 2029–2035

26 Kozlowski, J. M., Hart, I. R., Fidler, I. J. and Hanna, N. (1984) A human melanoma line heterogeneous with respect to metastatic capacity in athymic nude mice. J. Natl. Cancer Inst. **72**, 913–917

27 Li, L., Price, J. E., Fan, D., Zhang, R. D., Bucana, C. D. and Fidler, I. J. (1989) Correlation of growth capacity of human tumor cells in hard agarose with their in vivo proliferative capacity at specific metastatic sites. J. Natl. Cancer Inst. **81**, 1406–1412
28 Park, G. T. and Morasso, M. I. (1999) Regulation of the Dlx3 homeobox gene upon differentiation of mouse keratinocytes. J. Biol. Chem. **274**, 26599–26608
29 Suske, G. (1999) The Sp-family of transcription factors. Gene **238**, 291–300
30 Mantovani, R. (1998) A survey of 178 NF-Y binding CCAAT boxes. Nucleic Acids Res. **26**, 1135–1143
31 Rescheleit, D. K., Rommerskirch, W. J. and Wiederanders, B. (1996) Sequence analysis and distribution of two new human cathepsin L splice variants. FEBS Lett. **394**, 345–348
32 Sinha, S., Maity, S. N., Lu, J. and de Crombrugghe, B. (1995) Recombinant rat CBF-C, the third subunit of CBF/NFY, allows formation of a protein–DNA complex with CBF-A and CBF-B and with yeast HAP2 and HAP3. Proc. Natl. Acad. Sci. U.S.A. **92**, 1624–1628
33 Mantovani, R. (1999) The molecular biology of the CCAAT-binding factor NF-Y. Gene **239**, 15–27
34 Wright, K. L., Moore, T. L., Vilen, B. J., Brown, A. M. and Ting, J. P. Y. (1995) Major histocompatibility complex class II-associated invariant chain gene expression is up-regulated by cooperative interactions of Sp1 and NF-Y. J. Biol. Chem. **270**, 20978–20986
35 Roder, K., Wolf, S. S., Beck, K. F. and Schweizer, M. (1997) Cooperative binding of NF-Y and Sp1 at the DNase I-hypersensitive site, fatty acid synthase insulin-responsive element 1, located at −500 in the rat fatty acid synthase promoter. J. Biol. Chem. **272**, 21616–21624
36 Zhong, Z. D., Hammani, K., Bae, W. S. and DeClerck, Y. A. (2000) NF-Y and Sp1 cooperate for the transcriptional activation and cAMP response of human tissue inhibitor of metalloproteinases-2. J. Biol. Chem. **275**, 18602–18610
37 Roder, K., Wolf, S. S., Larkin, K. J. and Schweizer, M. (1999) Interaction between the two ubiquitously expressed transcription factors NF-Y and Sp1. Gene **234**, 61–69
38 Yamada, K., Tanaka, T., Miyamoto, K. and Noguchi, T. (2000) Sp family members and nuclear factor-Y cooperatively stimulate transcription from the rat pyruvate kinase M gene distal promoter region via their direct interactions. J. Biol. Chem. **275**, 18129–18137
39 Maity, S. N. and de Crombrugghe, B. (1998) Role of the CCAAT-binding protein CBF/NF-Y in transcription. Trends Biochem. Sci. **23**, 174–178

Received 14 August 2001/28 September 2001; accepted 22 October 2001