Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.
Structural organization, expression and chromosomal mapping of the mouse cystatin-C-encoding gene (Cst3)

(Recombinant genomic DNA; gene mapping; exon; intron; promoter sequence; proteinase inhibitor; multigene family)

Changgoo Huh a, James W. Nagle b, Christine A. Kozak c, Magnus Abrahamson d and Stefan Karlsson a

a Molecular and Medical Genetics Section, Developmental and Metabolic Neurology Branch, National Institute of Neurological Disorders and Stroke, National Institutes of Health, Bethesda, MD 20892, USA; b Neurogenetic Section, National Institute of Neurological Disorders and Stroke, National Institutes of Health, Bethesda, MD 20892, USA. Tel. (1-301) 496-8801; c Laboratory of Molecular Microbiology, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, MD 20892, USA. Tel. (1-301) 496-0972; and d Department of Clinical Chemistry, University Hospital, S-221 85 Lund, Sweden. Tel. (46-46) 173-445

Received by J. Marmur: 15 March 1994; Revised/Accepted: 20 May/23 May 1994; Received at publishers: 17 October 1994

SUMMARY

Cystatin C (CstC) is a potent cysteine-proteinase inhibitor. The structure of the mouse CstC-encoding gene (Cst3) was examined by sequencing a 6.1-kb genomic DNA containing the entire gene, as well as 0.9 kb of 5' flanking and 1.7 kb of its 3' flanking region. The sequence revealed that the overall organization of the gene is very similar to those of the genes encoding human CstC and other type-2 Cst, with two introns at positions identical to those in the human gene. The promoter area does not contain typical TATA or CAAT boxes. Two copies of a Sp1-binding motif, GGGCGG, are present in the 5' flanking region within 300 bp upstream from the initiation codon. A hexa-nucleotide, TGTTCT, which is a core sequence of the androgen-responsive element (ARE), is found in the promoter region. This region also contains a 21-nucleotide sequence, 5'-AGACTAGCAGCTGACTGAAGC, which contains two potential binding sites for the transcription factor, AP-1. The mouse Cst3 mRNA was detected in all of thirteen tissues examined by Northern blot analysis. Cst3 was mapped in the mouse to a position on distal chromosome 2.

INTRODUCTION

The cystatins (Cst) are a group of potent cysteine-proteinase inhibitors. There are at least five distinct types in the Cst superfamily, each type consisting of several proteins (Rawlings and Barrett, 1990; Devos et al., 1993). CstC belongs to the family of type-2 Cst and consists of 120 aa, with two intrachain disulfide bonds (Barrett et al., 1986). Although the proteinase-inhibiting function of the CstC has been thoroughly investigated, less is known about its broader biological role. Recent reports indicate that CstC may play a role in cancer progression (Sloane, 1990), bone resorption (Lerner and Grubb, 1992), modulation of neutrophil chemotactic activity and inflammation (Leung-Tack et al., 1990a,b), and resistance to viral infection (Collins and Grubb, 1991). Furthermore, a point mutation in the CST3 gene, resulting in a Leu→Gln substitution, is the primary cause of autosomal dominant hereditary disorder, hereditary CstC amyloid angiopathy (HCCAA) (Grubb et al., 1984). As young adults, carriers

Correspondence to: Dr. C. Huh, National Institutes of Health, Building 10, Room 3D04, 9000 Rockville Pike, Bethesda, MD 20892, USA. Tel. (1-301) 496-8236; Fax (1-301) 496-9480; e-mail: huh@pop.nih.gov

Abbreviations: aa, amino acid(s); ARE, androgen-responsive element(s); bp, base pair(s); Chr, chromosome; Cst, cystatin(s); Cst3, gene (DNA) encoding CstC; Cstl, gene (DNA) encoding CstSN; Cst2, gene (DNA) encoding CstA; Cst4, gene (DNA) encoding CstD; HCCAA, hereditary CstC amyloid angiopathy; kb, kilobase(s) or 1000 bp; nt, nucleotide(s); RFLP, restriction-fragment length polymorphism; W, A or T.

SSDI 0378-1119(94)00728-4
of this mutation suffer from repeated and massive brain hemorrhages due to deposition of the mutant protein in the walls of the cerebral arteries.

The gene structures for several type-2 Cst have been determined. The human CST1 and CST2 genes (Saitoh et al., 1987), and the human CST3 (Abrahamson et al., 1990) and CST4 (Freije et al., 1991) genes show very similar structural organization with respect to the number and position of the introns. Structural analysis of the genes for these proteins will be necessary to understand the function and evolution of the members in the cystatin multigene superfamily. In this paper, we report the structural organization and expression of the mouse Cst3 gene, compare its regulatory elements with that of other Cst genes and map the Cst3 gene in the mouse.

EXPERIMENTAL AND DISCUSSION

(a) Isolation of the mouse Cst3 gene

Using two primers myc3: (5'-ATG GCC AGC CCG CTG CGC TCC TTG-3') and myc4: (5'-GGC ATT TTT GCA GCT GAA TTT TGT CAG-3'), a 420-bp DNA fragment within the coding region was generated from mouse Cst3 cDNA, labeled with 32P by random primer extension, and used as a probe to screen a λFixII genomic DNA library from 129/Sv mice (purchased from Stratagene, La Jolla, CA, USA). Among several hybridizing clones, one clone, λcyg13, was chosen for characterization. Southern blot and PCR analysis indicated the presence of the mouse Cst3 gene in a 15-kb DNA insert. Southern blot analysis of the genomic DNA did not show any evidence for the presence of a Cst3 pseudogene. A 6.1-kb genomic DNA fragment containing the entire gene, as well as 0.9 kb of 5' flanking and 1.7 kb of 3' flanking region, was subjected to nt sequencing.

(b) Structural analysis of the mouse Cst3 gene

The 6125-nt sequence covering the entire mouse Cst3 gene is shown in Fig. 1. Comparison of the mouse Cst3 gene sequence with that of the corresponding cDNA (Solem et al., 1990) revealed that the gene contains two intron sequences located between the nt triplets encoding aa 55–56 and 93–94 of the proposed mature polypeptide chain, exactly as in the human CST3 gene. The presence of two introns, at homologous positions, has also been reported in the other type-2 Cst genes fully characterized to date: the human CST1, CST2 and CST4 genes. The intron-exon junctions in the mouse Cst3 gene are all close matches to the consensus sequences for the donor and acceptor splice sites of introns (Mount, 1982).

Some differences were observed between the exons of the genomic DNA sequence and the published cDNA. The five differences are summarized in Table I. Two of these positions, 994 and 3421, are in the coding region. The nt994 in exon 1 results in a GCC codon (coding for Ala) towards the C-terminal part of the leader sequence. However, a GCC codon (coding for Gly) was reported at this position in the published cDNA. This GCC codon found in the genomic DNA exists in the corresponding site of the rat and human Cst3 cDNAs. The nt3421 in exon 2 forms a TTG codon for Leu (TTT coding for Phe in cDNA). The differences between genomic and cDNA sequence may be due to an error during cDNA synthesis or due to polymorphism between the mouse strains 129/Sv and BALB/c.

(c) 5'-flanking sequence of the mouse Cst3 gene

The sequence of the 0.9-kb segment flanking exon 1 of the mouse Cst3 gene at the 5' end, did not reveal a typical TATA or CAAT box in the suggested promoter area. However, a TATA-box-like TAAAA sequence is present at 78–82 nt upstream from the start codon. A similar slightly atypical TATA-box is found at the homologous position in the human CST3 gene (ATAAAA), the human CST4 gene (ATAAAT), the human CST1 and CST2 genes (ATAAAA). The TATA-box is preceded by a Sp1-binding GC-box sequence (Pugh and Tjian, 1990) with the core consensus sequence, GGGCGG, ending 23 nt upstream from the AT-sequence. A corresponding sequence in the human CST3 gene is located slightly closer to the TATA-box (distance 16 nt). In the human CST4 gene, a GC-box is also found in the immediate 5'-flanking region (upstream distance from the AT-sequence 41 nt). By contrast, in the human CST1 and CST2 genes, a segment similar to the CAAT consensus is found instead of the GC box.

A core sequence of the ARE, TGTTTC, is found 22 nt downstream from the TAAAA sequence. This hexanucleotide is located in a partial palindromic setting. Some naturally occurring sequences and synthetic constructs containing this core sequence in a partial palindromic structure were shown to be inducible with androgens (Ham et al., 1988). This ARE is not found in the promoters of any other type-2 Cst genes published. However, it is present in Cst-related protein-encoding genes whose expression is regulated by androgen in the ventral prostate and lachrymal gland (Chamberlain et al., 1983).

An exact match of nine nt with the pituitary transcription factor (Pit-1) recognition element is centered around nt −795 from the start codon, but is probably of low significance for the expression of the gene because multiple recognition elements have been shown to be needed for markedly increased expression of the rat prolactin gene by Pit-1 (Ingraham et al., 1988). The motif
recognized by the leader binding protein (LBP-1), 5'-WCTGGG-3' or its inverse, that is present in several copies in the HIV-1 promoter and contribute to its basal function (Jones et al., 1988), is strikingly abundant in the 5'-flanking region of the mouse Cs3 gene. Another five LBP-1 motifs are found within a 300-bp segment in the first part of the first intron.

Transcription factor AP-1-binding sites that bind the Jun-Fos protooncogene complexes contain the consensus sequence 5'-TGACTCAGC. The mouse Cs3 promoter contains two AP-1-like binding sequences within the sequence 5'-AGACTAGCAGCTGACTGAAGC, immediately upstream from the first Sp1-binding site. This 21-mer sequence contains direct repeats of two adjacent potential AP-1-binding sites, each slightly deviating from the consensus sequence. It has been shown that two adjacent AP-1-like binding sites act synergistically to confer inducibility beyond that observed for a single AP-1 consensus sequence (Fricing et al., 1992). The presence of the two AP-1-like binding sites in the promoter indicates that
displays several Spl-binding sites and contains a high degree of methylation due to constant transcription of several type-2 the gene (Bird, 1986). Proposed promoter regions of promoters and the human extremely high G + C content. Rather it is similar to these is not a typical housekeeping gene promoter having number of CpG dinucleotides. This may indicate a low ratio is 1/1.3 in the 300 bp region upstream from the start codon. The GC content of approx. 60%. However, the CpG/GpC ratio is determined on both strands (number of independent sequencing runs in parentheses).

A strong positive regulation of the Cst3 gene by TGF-β in serum free mouse embryo cells has also been reported (Solem et al., 1990). The presence of AP-1-like binding sites in the mouse Cst3 promoter suggests that cystatin C induction by TGF-β may be mediated by the AP-1 complex.

The 5′-flanking region of the human CST3 gene has a notably high G + C content, with >70% G + C in the 400 bp sequence upstream from the start codon. The G+C-rich region also includes the coding part of exon 1 and the 5′ part of the first intron, which together represents a 900-bp segment with a G+C content of 73%, and contains CpG/GpC dinucleotides in a ratio close to unity (Abrahamson et al., 1990). The immediate 5′-flanking region of the mouse Cst3 gene does not have such a strikingly high G + C content, but is more similar to the human CST1, CST2 and CST4 genes in having a GC content of approx. 60%. However, the CpG/GpC ratio is 1/1.3 in the 300 bp region upstream from the start codon (as compared to 1/4.1 over the entire 6.1-kb sequenced region), differing markedly from ratios of 1/6, 1/9 and 1/16 for the human CST4, CST1 and CST2 genes, respectively. Thus, the mouse Cst3 gene 5′-flanking region is not a typical housekeeping gene promoter having extremely high G + C content. Rather it is similar to these promoters and the human CST3 promoter because it displays several Spl-binding sites and contains a high number of CpG dinucleotides. This may indicate a low degree of methylation due to constant transcription of the gene (Bird, 1986). Proposed promoter regions of several type-2 Cst are compared in Fig. 2.

### TABLE 1

| Position | Genomic DNA | cDNA | aa (genomic/cDNA) |
|----------|-------------|------|-------------------|
| 994      | C (11)      | G    | Ala/Gly           |
| 3421     | G (5)       | T    | Leu/Phe           |
| 4393     | C (5)       | T    |                   |
| 4564-5   | GC (3)      | AT   |                   |

(a) The nt positions refer to Fig. 1.  
(b) The nt refers to the genomic DNA sequence. Sequence was determined on both strands (number of independent sequencing runs in parentheses).  
(c) The nt refers to the cDNA sequence.  
(d) The aa encoded by nt in columns Genomic DNA and cDNA.

4564-5 GC (3) AT

(d) 3′-Flanking sequence of the mouse Cst3 gene

Sequence determination of the mouse Cst3 gene 1.7-kb 3′-flanking segment revealed no alternative polyadenylation signals in addition to the one present in the corresponding cDNA. 213 bp downstream from the stop codon. Analysis of short tandem repeats within the entire gene sequence revealed the presence of (GT)_{12} and (GA)_{20} in the region immediately 3′-flanking the polyadenylation signal and three stretches of perfect CT repeats, (CT)_{13}, (CT)_{14} and (CT)_{25}, 1300 bp further downstream.

(e) Genetic mapping of the mouse Cst3 gene

Analysis of two multilocus crosses was used to define the Chr location for the mouse Cst3 gene: (NFS/N or C58/J × M. m. musculus)× M. m. musculus (Kozak et al., 1990) and (NFS/N× M. spretus)× M. spretus or C58/J (Adamson et al., 1991). DNAs extracted from parental mice and progeny of the crosses were typed by Southern blotting analysis for RFLPs of Cst3 using the mouse Cst3 DNA as probe. SstI digestion produced fragments of 8.6 and 6.0 kb in M. spretus and 6.0 and 2.8 kb in NFS/N and C58/J mice, and BanHI fragments of 22.0 and 18.8 kb were detected in M. m. musculus and NFS/N, respectively.

Inheritance of the parental fragments was followed in both crosses and compared with inheritance of almost 650 markers previously typed and mapped to all 19 autosomes and the X chromosome. As shown in Fig. 3, Cst3 showed linkage to markers on distal Chr 2 and could be positioned proximal to Psp (encoding parotid secretory protein) and Emv15 (ectropic murine virus 15).
A Snap
Cst3
Psp, Emv15
M. spretus cross 30 26 3 1 0 3
M. m. musculus cross 9 14 2 1 0 0

Recombination:
Snap - 7.9±2.9 - Cst3 - 3.4±1.9 - Psp, Emv15

B

20pter-p12 73 - Prn 5
77 - Snap 8
86 - Psp 3
89 - Emv15
90 - Src
95 - Ada

Fig. 3. Genetic mapping of the mouse Cst3 gene. (A) Segregation of Cst3 with markers on Chr 2. Blackened squares represent heterozygous mice, open squares are homozygous mice. The numbers under each column represent the number of progeny in each of the two backcrosses with the indicated genotype. Recombination frequencies are given below the figure. The numbers given between adjacent loci represent percent recombination±the standard error calculated according to Green (1981). (B) Chr 2 linkage maps. The map on the right was generated from the two crosses described here and indicates the position of Cst3 relative to the other markers typed in this cross. Distances between adjacent markers (in centiMorgans) are indicated to the immediate left of the map. The map on the left is an abbreviated version of the composite genetic map (Siracusa and Abbott, 1993). Numbers to the left of the map are centiMorgan distances from the centromere. Human map locations for homologs of the underlined mouse genes are indicated to the far left of this map.

and distal to Snap (encoding synaptosomal associated protein), markers which were typed in these crosses as previously described (Joseph et al., 1990; Grimaldi et al., 1992). It has recently been shown that the human homolog of this gene maps to 20p11 (Schnittiger et al., 1993). This is consistent with our results, since the distal end of mouse Chr 2 contains a substantial region of linkage conservation with human Chr 20 (Siracusa and Abbot, 1993; Fig. 3b). The human CST3 is part of a cluster which includes up to eight members of the CST gene family (Schnittiger et al., 1993) further suggesting that the mouse homologs of these genes are likely to map to the same site on Chr 2.

(f) Expression of the mouse Cst3 gene

We examined the expression of the mouse Cst3 gene in 13 different tissues by Northern blot analysis using the mouse Cst3 cDNA probe. As expected, Cst3 mRNA was detected in all tissues examined, including stomach, brain, intestines, liver, muscle, spleen, heart, kidney, lung, pancreas, testis, uterus and ovary (data not shown). The pattern of the mouse Cst3 gene expression is similar to that of its human counterpart. Both species show expression of the gene in all tissues examined with high level of Cst3 messenger RNA in brain and testis, and lowest level in pancreas. This overall similarity between the two species indicates that mouse may be suitable for generating an animal model for the human genetic disease HCCAA.

ACKNOWLEDGEMENT

We thank Dr. Jakob Reiser for critical reading of the manuscript.

REFERENCES

Abrahamson, M., Olafsson, I., Palsdottir, A., Ulvsbäck, M., Lundwall, A., Jansson, O. and Grubb, A.: Structure and expression of the human cystatin C gene. Biochem. J. 268 (1990) 287–294.

Adamson, M.C., Silver, J. and Kozak, C.A.: The mouse homolog of the gibbon ape leukemia virus receptor: genetic mapping and a possible receptor function in rodents. Virology 183 (1991) 778–781.

Barrett, A.J., Fritz, H., Grubb, A., Isemura, S., Järvinen, M., Katunuma, N., Machleidt, W., Müller-Esterl, W., Sasaki, M. and Turk, V.: Nomenclature and classification of the proteins homologous with the cysteine- proteinase inhibitor chicken cystatin. Biochem. J. 236 (1986) 312.

Bird, A.P.: CpG-rich islands and the function of DNA methylation. Nature 321 (1986) 209–213.

Chamberlain, L.L., Mpanias, O.D. and Wang, T.Y.: Isolation, properties, and androgen regulation of a 20-kilodalton protein from rat ventral prostate. Biochemistry 22 (1983) 3072–3077.

Collins, A.R. and Grubb, A.: Inhibitory effects of recombinant human cystatin C on human coronavirus. Antimicrob. Agents Chemother. 35 (1991) 2444–2446.
Devos, A., De Clercq, N., Vercaeren, I., Heyns, W., Rombauts, W. and Peeters, B.: Structure of rat genes encoding androgen-regulated cystatin-related proteins (CRPs): a new member of the cystatin superfamily. Gene 125 (1993) 159–167.

Freije, J.P., Abrahamson, M., Ólafsson, Í., Velasco, G., Grubb, A. and López-Otin, C.: Structure and expression of the gene encoding cystatin D, a novel human cysteine proteinase inhibitor. J. Biol. Chem. 261 (1996) 20538–20543.

Fritling, R.S., Bergelson, S. and Daniel, V.: Two adjacent AP-1-like binding sites form the electrophilic-responsive element of the murine glutathione S-transferase Ya subunit gene. Proc. Natl. Acad. Sci. USA 89 (1992) 668–672.

Green, E.L.: Genetics and Probability in Animal Breeding Experiments. MacMillan, New York, NY, 1981, pp. 77–113.

Grimaldi, L.C., Torres, R., Kozak, C.A., Chang, R., Clark, E.A., Howard, M. and Cockayne, D.A.: Genomic structure and chromosomal mapping of the murine CD40 gene. J. Immunol. 149 (1992) 3921–3926.

Grubb, A., Jensson, O., Gudmundsson, G., Arnason, A., Löfberg, H. and Malm, J.: Abnormal metabolism of γ-trace alkaline microprotein. N. Engl. J. Med. 311 (1984) 1547–1549.

Ham, J., Thomson, A., Needham, M., Webb, P. and Parker, M.: Characterization of response elements for androgens, glucocorticoids and progesterin in mouse mammary tumour virus. Nucleic Acids Res. 16 (1988) 5263–5276.

Ingraham, H.A., Chen, R., Mangalam, H.J., Elsholz, H.P., Flynn, S.E., Lin, C.R., Simmons, D.M., Swanson, L. and Rosenfeld, M.G.: A tissue-specific transcription factor containing a homeodomain specifying a pituitary phenotype. Cell 55 (1988) 519–529.

Jones, K.A., Angel, P., Lafayt, R., Hattori, K., Kim, K.Y., Sporn, M.B., Karin, M. and Roberts, A.B.: Autoinduction of transforming growth factor β1 is mediated by the AP-1 complex. Mol. Cell. Biol. 10 (1990) 1492–1497.

Kozak, C.A., Peyser, M., Kral, M., Mariano, T.M., Kumar, C.S., Pestka, S. and Mock, B.A.: Molecular genetic markers spanning mouse chromosome 10. Genomics 8 (1990) 519–524.

Leung-Tack, J., Tavera, C., Martinez, J. and Colle, A.: Neutrophil chemotactic activity is modulated by human cystatin C, an inhibitor of cysteine proteinases. Inflammation 14 (1990a) 247–258.

Leung-Tack, J., Tavera, C., Genc, M.C., Martinez, J. and Colle, A.: Modulation of phagocytosis-associated respiratory burst by human cystatin C role of the N-terminal tetrapeptide Lys-Pro-Pro-Arg. Exp. Cell Res. 188 (1990b) 16–22.

Lerner, U.H. and Grubb, A.: Human cystatin C, a cysteine proteinase inhibitor, inhibits bone resorption in vitro stimulated by parathyroid hormone and parathyroid hormone-related peptide of malignancy. J. Bone Miner. Res. 7 (1992) 433–440.

Mount, S.M.: A catalogue of splice junction sequences. Nucleic Acids Res. 10 (1982) 459–472.

Pugh, B.F. and Tjian, R.: Mechanism of transcriptional activation by Sp1; evidence for coactivators. Cell 61 (1990) 1187–1197.

Rawlings, N.D. and Barrett, A.J.: Evolution of proteins of the cystatin superfamily. J. Mol. Evol. 30 (1990) 60–71.

Saitoh, E., Kim, H.-S., Smithies, O. and Maeda, N.: Human cysteine proteinase inhibitors: nucleotide sequence analysis of three members of the cystatin gene family. Gene 61 (1987) 329–338.

Schnittger, S., Gopal Rao, V.V.N., Abrahamson, M. and Hannum, I.: Cystatin C (CST3), the candidate gene for hereditary cystatin C amyloid angiopathy (HCCA/A), and other members of the cystatin gene family are clustered on chromosome 20p11.2. Genomics 16 (1993) 50–55.

Siracusa, L. D. and Abbot, C.M.: Mouse chromosome 2. Mammal. Genome 4 (1993) S31–S46.

Sloane, B.F.: Cathespin B and cystatins: evidence for a role in cancer progression. Cancer Biol. 1 (1990) 137–152.

Solem, M., Rawson, C., Lindberg, K. and Barnes, D.: Transforming growth factor beta regulates cystatin C in serum-free mouse embryo (SFME) cells. Biochem. Biophys. Res. Commun. 172 (1990) 945–951.