

Primary structure of bovine cathepsin S

Comparison to cathepsins L, H, B and papain

Bernd Wiederanders¹, Dieter Broemme¹, Heidrun Kirschke¹, Nisse Kalkkinen², Ari Rinne³, Thomas Paquette³ and Penelope Toothman³

¹Institute of Biochemistry, Faculty of Medicine, University of Halle, Germany, ²Institute of Biotechnology, University of Helsinki, Finland and ³Institute of Medical Biology, PO Box 977, University of Tromsø, 9001 Tromsø, Norway

Received 26 March 1991; revised version received 23 May 1991

The primary structure of bovine cathepsin S was determined by combining results of protein and peptide sequencing with the sequence deduced from nucleic acid sequencing. Using polymerase chain reaction (PCR) technology, cDNA clones commencing at amino acid 22 of the mature enzyme and continuing through the 3' untranslated region of bovine cathepsin S mRNA were isolated and sequenced. The open reading frame in these overlapping clones correctly predicts the determined amino acid sequence of 13 tryptic peptides derived from purified bovine spleen cathepsin S. The deduced amino acid sequence shows that mature bovine cathepsin S consists of 217 amino acids corresponding to a molecular weight of 23.7 kDa. Cathepsin S belongs to the papain superfamily of lysosomal cysteine proteinases and shares 41% identity with papain. Amino acid sequence identities of bovine cathepsin S to human cathepsins L, H, and B are 56%, 47% and 31% respectively.

Cathepsin S; Cysteine proteinase; Amino acid sequence; cDNA cloning; PCR; Bovine spleen

1. INTRODUCTION

Lysosomal cysteine proteinases are important for cellular protein breakdown. The cathepsins B (EC 3.4.22.1), H (EC 3.4.22.16) and L (EC 3.4.22.15) are the best characterized enzymes of this group [1]. The recently described cathepsin S [2,3] has a high specific activity towards protein substrates, a property it shares with cathepsin L. However, cathepsin S differs from cathepsin L since it is a single chain enzyme like papain. Furthermore, it differs from all cysteine proteinases since it is stable and active *in vitro* at neutral pH.

The N-terminal part (35 amino acid residues) of the bovine cathepsin S sequence was reported by Turk et al. [4] as that of bovine cathepsin L. At that time, cathepsin S was insufficiently characterized, and therefore, the enzyme isolated from bovine spleen was thought to be a species variant of cathepsin L. We report here the primary structure of bovine cathepsin S. This information is necessary for the ultimate determination of the structural basis for the difference in properties of the cathepsins. We have compared the amino acid sequence of the cathepsins in the region of cleavage of cathepsins L and H into heavy and light chains and propose an explanation based on secondary structure for the failure of cathepsin S to be similarly cleaved in this region.

2. MATERIALS AND METHODS

2.1. Purification, alkylation, and trypsin digestion of cathepsin S

Cathepsin S was purified from bovine spleen as described previously [3]. The purified protein (25 µg) was alkylated with 4-vinylpyridine [5] and desalted by reversed phase chromatography on a 0.46 × 3 cm TSK TMS250 (CI) column. Alkylated cathepsin S was dissolved in 60 µl 0.1 M ammonium bicarbonate and treated with 4% (w/w) TPCK-trypsin (Sigma) for 8 h. Tryptic peptides were separated by reversed phase chromatography on a 0.46 × 15 cm Vydac 218TPB5 column using a linear gradient of acetonitrile (0–60% in 90 min) in 0.1% trifluoroacetic acid.

2.2. Amino acid sequencing

Unalkylated cathepsin S as well as the tryptic peptides were degraded (Applied Biosystems program 03RPTH) in a gas/pulsed liquid sequencer [6] connected to an on-line PTH-amino acid analyzer consisting of a Brownlee MicroGradient syringe pump, Rheodyne 7126 pneumatic injector, Jones chromatography oven and a Spectra Physics SP 8450 detector. PTH-amino acids were separated on an Applied Biosystems PTH C-18 column using conditions described for the Applied Biosystems 120A PTH-analyzer. For degradation, samples were applied on Polybrene (2 mg) pretreated glass fiber discs.

2.3. cDNA cloning

RNA was purified from bovine spleen [7] and mRNA was isolated [8]. mRNA (0.5 µg) was reverse transcribed by avian myeloblastosis virus reverse transcriptase (Bethesda Research Laboratories) using oligo-(dT)12–18 or oligo-(dT)-adapter as primer according to the manufacturers recommendations. Amplification of cathepsin S cDNA by polymerase chain reaction (PCR) was essentially according to Frohman et al. [9]. Reactions were amplified for 25 cycles. The PCR cycle used was 1'–3'–1' at 55°C, 65°C, and 94°C, respectively. Oligo nucleotide primers are listed in Fig. 1.

After chloroform extraction and ethanol precipitation PCR products were treated with DNA polymerase I Klenow fragment (Bethesda Research Labs) to ensure the ends were flush. Agarose gel purified cDNAs were directly cloned into the *Sma*I site of

Correspondence address: P. Toothman, Dept. of Microbiology, PO Box 977, University of Tromsø, 9001 Tromsø, Norway. Fax: (47) (83) 80850

Primers Used to Clone Bovine Cathepsin S

Primers	Amino Acid Sequence
1. TGTGGCTCTTGCTGGGNCCT	CGSCWAF
2. GTTTTGGACAAGCAUTAUTC*	DYWLKVN
3. ATATCCCTGATCTCCUAAUTG*	HFGDQGY
4. CATTGTGGAATTGCTAAATAYCC	HCGIANYP
5. AACTGGTGTCTACTATGACC	KTGVYYP
6. AACCATGGTGTACTCGTCGT	NHGVLVV
7. GGAAGCTTCTGAATCGATGC*	GIDSEAS
8. GACTCGAGTCGACATCGA(T) ₁₅	oligo-dT primer/ adapter
9. GACTCGAGTCGACATCG	primer/ adapter

Fig. 1. *Indicates primers which are antisense. Primers 1-4 were chosen from the possible codons for the corresponding peptide sequences. Wobble bases were included in these primers towards the 3' end to ensure efficient replication by Taq polymerase (N = G,A,T,C; U = G,A; Y = C,T). The sequences of primers 5, 6, and 7 were taken from the nucleic acid sequence of the first cathepsin S clones isolated.

pGem-7Zf(+) (Promega Biotec) by blunt-end ligation using T4 ligase (Bethesda Research Labs) according to the manufacturers instructions.

2.4. DNA sequencing

Plasmid DNA was isolated by the alkaline lysis procedure [10] followed by polyethylene glycol precipitation. Double stranded DNA was sequenced by the dideoxy chain termination method using the T7 DNA polymerase enzyme, Sequenase (US Biochemicals) according to the manufacturers specifications.

2.5. Estimation of size of cathepsin S mRNA

Glyoxylated RNA samples [11] were electrophoresed in 25 mM sodium phosphate buffer, pH 6.5, 2% Nusieve agarose, 0.1% Seakem agarose for 320 V.h at 4°C with buffer recirculation. The RNA was transferred to Nytran membrane (Schleicher and Schuell) according to the manufacturers instructions. Cathepsin S cDNA amplified with primer pair 1 and 2 was gel purified and 50 ng were labeled with dCT32P by random prime labelling [12]. Hybridization and washing conditions have been described [13].

3. RESULTS AND DISCUSSION

For direct primary structure analysis cathepsin S was purified from bovine spleen as described [3]. The purified enzyme (2 µg) was subjected to 24 cycles in the sequencer. The determined sequence of the first 24 amino acid residues (Fig. 2) shows three differences from the previously published one [4]; Met instead of Val in position 5, Cys instead of Gly in position 12 and Glu instead of Pro in position 15. These differences are also confirmed by the sequenced tryptic peptides cathepsin S-(1-8)-peptide and cathepsin S-(11-17)-peptide. The sequences of the purified tryptic peptides derived from alkylated cathepsin S are also shown in Fig. 2.

The cDNA clones for cathepsin S were derived from reverse transcribed bovine spleen poly(A⁺) RNA after PCR amplification. From the amino acid sequences of 16 bovine cathepsin S tryptic peptides, four peptides were chosen and the corresponding oligonucleotide

Bovine Cathepsin S cDNA

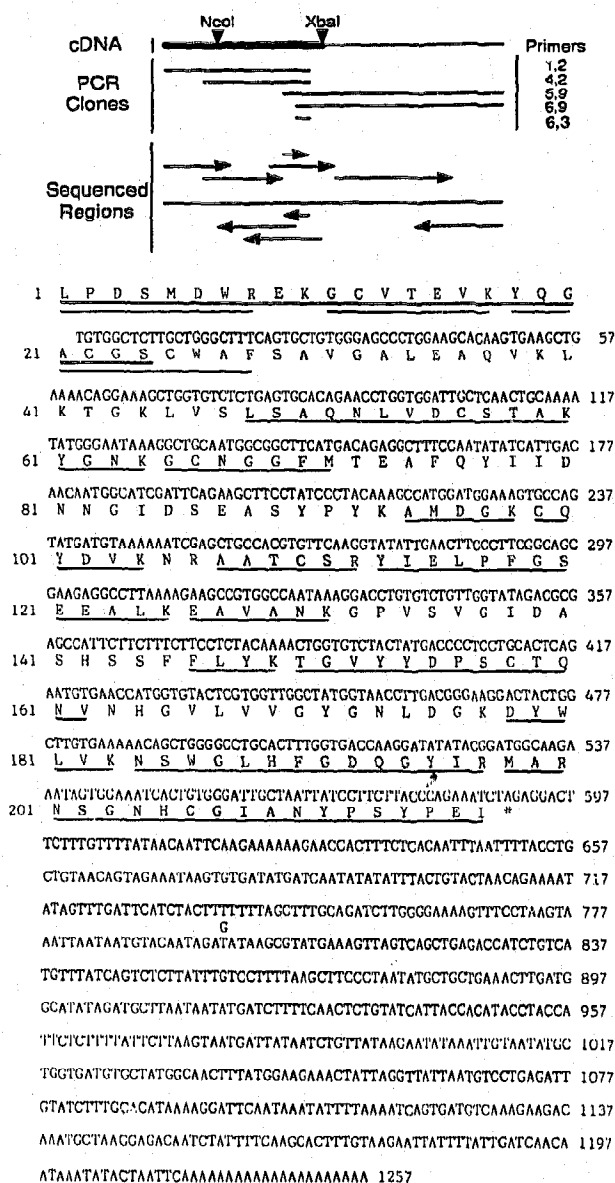


Fig. 2. The heavy line in the map corresponds to the amino acid coding region. The lengths of cDNA clones isolated and the primers used are indicated. The solid arrows show the length and direction sequenced using the SP6 and T7 primer sites in the plasmid. The first 20 nucleotides are the sequence of primer no. 1 and not necessarily the sequence found in the bovine cathepsin S gene. The result from the N-terminal sequence analysis of the purified enzyme as well as results from sequence analyses of purified tryptic peptides are underlined.

primers were synthesized. The primers used to isolate cathepsin S clones are described in Fig. 1. The initial cathepsin S cDNA clone was isolated with primer combination 1 and 2. The sequence of primer 1 encodes a conserved peptide near the amino terminus of the papain enzyme family which contains the active site cysteine and the sequence of primer 2 is antisense code for another conserved peptide near the carboxyl-terminus.

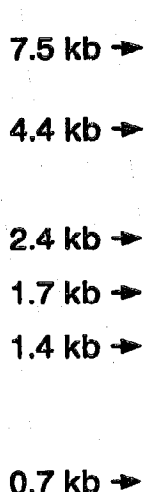
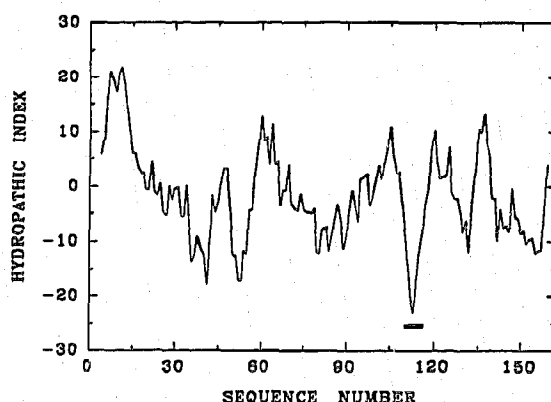


Fig. 3. RNA blot hybridization predicts a single species of bovine cathepsin S in the spleen. The estimated size is 1.7 kb, using ethidium bromide stained RNAs as mobility markers.

Two cDNA clones amplified with primers 1 and 2 were isolated from independent PCR reactions. The nucleic acid sequence of these clones predicts an open reading frame which would code for 10 of the analyzed tryptic peptides generated from the cathepsin S protein. Three additional cathepsin S specific primers and the adapter described in Fig. 1 were used to amplify the additional cathepsin S cDNAs and resulting clones described in Fig. 2. All tryptic peptide sequences determined were present in the single open reading frame identified as cathepsin S. A stop codon immediately follows the nucleotides for the last determined amino acid (Ile) in cathepsin S-(201-217)-peptide. This shows that mature cathepsin S is not processed at its C-terminus and consists of 217 amino acid residues corresponding to a molecular weight of 23.7 kDa. This molecular weight is in excellent agreement with the apparent molecular weight 24 kDa determined by SDS-PAGE.

Only a single band of 1.7 kb can be detected after blot hybridization to bovine spleen poly(A⁺) RNA indicating that the cDNA clones were probably derived from a single molecular species of RNA (Fig. 3). The cDNA amplified with primer combination 4 and 2 results from unpredicted priming of primer 4 at nucleotides 205-228. One clone predicts a proline (CCT) at amino acid 143 whereas the other two independent clones covering the same region predict a serine (TCT). Another single nucleotide discrepancy occurred at nucleotide 798. These differences may reflect the fidelity of replication of the TAQ polymerase.

Bovine cathepsin S primary structure shows 56% identity to cathepsin L [14], 47% to cathepsin H [15], 31% to cathepsin B [16] and 41% to papain [17] (Fig. 4). Overall there are 38 identities (18%) shared by all five enzymes. Cys-25 and His-164 of bovine cathepsin S



	1		45
S	LPDSMDWRER	G.CUTEVKY..	..QGACGSCW
L	APRSVDWRER	G.YUTPVKN..	..QGACGSCW
H	YPPSVDRERK	GNFVSPVKN..	..QGACGSCW
Papain	IPFYSVDWRQK	G.AVTPVKN..	..QGACGSCW
B	LPASFDAREQ	WFQCPTIKEX	RDQSGSCGSCW
			AFSAVGALAE
			QVKLKTG.KL
			QMFRTKG.RL
			ALAIATG.KH
			IIKIRTG.NH
			RICHTNARV
	46		94
S	.VSLAQNLV	DCSTAKYGNK	GCNGGFMTEA
L	.ISLSEQLV	DCSGFQ.GNE	GCNGGLMDYA
H	.ISLAEQQLV	DCAGD.FNNY	GCNGGLPSQA
Papain	N.QYSEQLL	DCDRRSY...	GCNGGYPMSA
B	SVEVSAEDLL	TCCGSMCGD.	GCNGGYPALAE
			WNFWR.KGL
			VSGGL..YES
	95		119
S	MDGKCQ.Y..DVKNRAATC
L	TEESCK.Y..NPKYSVAND
H	KDGYCK.F..QFKAIGFVK
Papain	VQRYCR.....S REKGYAAKT
B	HVG.CRFYSI	PFCEHNVGS	RPFCGEGEDT
			PKCKICEPG
			YSPTYQDKH
	120		158
S	SEELKEAVA	NKGPVSUGID
L	QERALKAVA	TUVPISVAID
H	DEEAMVEVA	LYNPVSFAFE
Papain	NQGALLYIA	NO.FVSVVLQ
B	YGVDYSVSN	SEKDIAMET	KNGFVEGAFS
			VY.GDFLLYK
			SGVYQVNT..
	159		202
S	.TQNVNKG	LVUGVGNLGD	KD....YMLV
L	.SEDMHGV	LVUGVGFEST	ESDNNKYMV
H	KTFDKVNHAV	LAUGYGERNG	IF....YMY
Papain	.GKVDHAY	AAUGYNPG...YLL
B	.GEMMGHAI	RILGWGVNG	TP....YMLV
			ANSNTDWDG
			NGFFKILRGQ
	203		217
S	GNI...CGIA	NYPSPYET...	
L	RNI...CGIA	SAASYPV...	
H	KIM...CGIA	ACASYPPLV	
Papain	GNSYGVGGLY	TSSFYEVKN	
B	.DH...CGIE	SEVVAGIPRT	D

Fig. 4. Alignment of the amino acid sequences of bovine cathepsin S, human cathepsins L [14], H [15], and B [16], and papain [17]. The gaps have been introduced to achieve maximal homology. Shaded amino acids are identical in all five enzymes.

correspond to the papain active site [17]. Amino acids directly flanking Cys-25 are highly conserved whereas those flanking His-164 are not identical but have similar side chains. The greater similarity of cathepsin S to cathepsin L is in keeping with the similar enzymatic properties of the two enzymes [2,3,18].

Cathepsins S and papain are single chain enzymes whereas cathepsins H, L and B are two chain enzymes. Cathepsin B is cleaved close to the N-terminus of the polypeptide and is stable to cleavage between residues 164 and 186 like cathepsin S and papain. Fig. 5 shows the predicted secondary structure of cathepsins and average flexibility parameters in the region where cathepsins H and L are cleaved to produce the heavy and light chains, respectively. This region is not conserved in cysteine proteinases but is flanked by two conserved regions which correspond to cathepsin S

Enzyme	Position*	Secondary Structure	Flexibility	Reference
Consensus		HavllvGvg---(n)---YwlvkNSW		
human L	276	t-BBBBBB---t-tt-TBBBBBttth	1.089	19
mouse L	276	t-BBBBBB-tt-TTTTTBBBBBtttt	1.071	20
rat L	276	t-BBBBBB-tt-TTTTTBBBBBtttt	1.065	21
human H	281	BBBBBttt-TT-BBBBBBtttT	1.058	22
rat H	166	BBBBBttt-TTBBBBBttt-	1.058	23
bovine S	164	BBBBBttttttt-BBBBBtttb	1.048	**
human S	164	BBBBBttttttt-BBBBBtttT	1.040	***
human B	278	hhhhhh---tt---BBBBBttt	1.043	24
rat B	197	BBBBB---tt---BBBBBttB	1.004	25
mouse B	278	BBBBBBB---tt-BBBBBBttt-	1.003	24
bovine B	199	BBBBB---tt---BBBBB???	1.043	26
papain	292	HHHHHHttt-BBBBBtttT		27

Fig. 5. The secondary structure and chain flexibility adjacent to the site of cleavage during formation of cathepsins L and H heavy and light chains. Upper case and lower case B, T and H indicate strong and weak predictions of β sheet, turn and helix, respectively. The cleavage sites are shown by the arrows. The consensus sequence for this region is shown at the top, where *n* represents a variable length, unconserved sequence ranging in length: 10 amino acids in cathepsin L, 6 amino acids in cathepsins S, H and B, and 2 amino acids in papain. The secondary structure symbols corresponding to the conserved amino acids of the consensus are underlined. The average of the flexibility values for the cathepsins in the variable region is shown. The value for papain is 0.989. Secondary structure (Chou-Fasman [28]) and flexibility (Karplus and Schulz [29]) were calculated by using the Peptidestructure program of the Genetics Computer Group, Madison, WI. *Refers to the position of the conserved H in the enzyme or prepro-enzyme sequence. **This manuscript. ***B. Wiederanders (unpublished results).

(164-173) and (178-186). Secondary structure calculations predict that the conserved regions tend to form β sheets. The intervening sequences which are cleaved in cathepsins H and L are strongly predicted to form reverse turn structures and to display more flexibility than in cathepsins S and B which are not cleaved in this region. A comparison of the means of the flexibility values for cathepsins H and L to those of S and B indicates a statistically significant difference (*t*-test, $P=0.0046$). Thus, the absence of a reverse turn and/or insufficient chain flexibility may render cathepsin S resistant to cleavage between amino acids (173-178).

Acknowledgements: This work was supported by an Olav Aakre foundation (Tromsø, Norway) grant to A.R. and a NAVF North Norway Biotechnology Program fellowship to P.T.

REFERENCES

- [1] Kirschke, H. and Barrett, A. (1987) in: *Lysosomes: Their Role in Protein Breakdown* (Glaumann, H. and Ballard, F.J. eds.) pp. 193-238, Academic Press, London.
- [2] Kirschke, H., Schmidt, I. and Wiederanders, B. (1986) *Biochem. J.* 240, 455-459.
- [3] Kirschke, H., Wiederanders, B., Broemme, D. and Rinne, A. (1989) *Biochem. J.* 264, 467-473.
- [4] Turk, V., Brzin, J., Kopitar, M., Kregar, I., Locknikar, P., Longer, M., Popovic, T., Ritonja, A., Vitale, L., Machleidt, W., Giraldi, T. and Sava, G. (1983) in: *Proteinase Inhibitors, Medical and Biological Aspects* (Katunuma, N., Umezawa, H. and Holzer, H. eds.) pp. 125-134, Japan Sci. Soc. Press, Tokyo.
- [5] Thomsen, J. and Bayne, S. (1988) *J. Prot. Chem.* 7, 295-296.
- [6] Kalkkinen, N. and Tilgmann, C. (1988) *J. Prot. Chem.* 7, 242-293.
- [7] Chirgwin, J.M., Przybyla, A.E., McDonald, R.J. and Rutter, W.J. (1979) *Biochemistry* 18, 5294-5299.
- [8] Aviv, H. and Leder, P. (1972) *Proc. Natl. Acad. Sci. USA* 69, 1408-1412.
- [9] Frohmann, M.A., Dush, M.K. and Martin, G.R. (1988) *Proc. Natl. Acad. Sci. USA* 85, 8998-9002.
- [10] Birnboim, H.C. and Dolby, J. (1977) *Nucleic Acids Res.* 7, 1513-1518.
- [11] McMaster, G.K. and Carmichael, G.G. (1977) *Proc. Natl. Acad. Sci. USA* 74, 4835-4838.
- [12] Feinberg, A.P. and Vogelstein, B. (1983) *Anal. Biochem.* 132, 6-13.
- [13] Toothman, P. and Paquette, T. (1987) *Mol. Endocrinol.* 1, 413-419.
- [14] Joseph, L.J., Chang, L.C., Stamenkovich, D. and Sukhatme, V.D. (1988) *J. Clin. Invest.* 81, 1621-1629.
- [15] Fuchs, R., Machleidt, W. and Gassen, H.-G. (1988) *Biol. Chem. Hoppe-Seyler* 369, 469-475.
- [16] Ritonja, A., Popovic, T., Turk, V., Wiedenmann, K. and Machleidt, W. (1985) *FEBS Lett.* 181, 169-172.
- [17] Husain, S.S. and Lowe, G. (1969) *Biochem. J.* 114, 279-288.
- [18] Broemme, D., Steinert, A., Friehe, S., Fittkau, S., Wiederanders, B. and Kirschke, H. (1989) *Biochem. J.* 264, 4475-4481.
- [19] Gal, S. and Gottesman, M.M. (1988) *Biochem. J.* 253, 303-306.
- [20] Portnoy, D.A., Erickson, A.H., Kochan, J., Ravetch, J.V. and Unkeless, J.C. (1986) *J. Biol. Chem.* 261, 14697-14703.
- [21] Ishidoh, K., Towatari, T., Imajoh, S., Kawasaki, H., Kominami, E., Katunuma, N. and Suzuki, K. (1987) *FEBS Lett.* 223, 69-73.
- [22] Fuchs, R. and Gussin, H. G. (1989) *Nucleic Acids Res.* 17, 9471.
- [23] Ishidoh, K., Kominami, E., Katunuma, N. and Suzuki, K. (1989) *FEBS Lett.* 253, 103-107.
- [24] Chan, S.J., San Segundo, B., McCormick, M.B. and Steiner, D.F. (1986) *Proc. Natl. Acad. Sci. USA* 83, 7721-7725.
- [25] Takio, K., Towatari, T., Katunuma, N., Teller, D.C., and Titani, K. (1983) *Proc. Natl. Acad. Sci. USA* 80, 3666-3670.
- [26] Meloun, B., Pohl, J., and Kostka, V. (1986) in: *Cysteine Proteinases and Their Inhibitors* (Turk, V. ed.) pp. 19-29, Walter de Gruyter, Berlin/New York.
- [27] Cohen, L.W., Coghlan, V.M. and Dihel, L.C. (1986) *Gene* 48, 219-227.
- [28] Chou, P. Y. and Fasman, G. D. (1978) *Adv. Enzymol.* 47, 45-148.
- [29] Karplus, P.A. and Schulz, G.E. (1985) *Naturwissenschaften* 72, 212-213.