Volume 283, number 2, 329–331 FEBS 09812 June 1991

ADONIS 001457939100520#

The complete amino acid sequence of bovine cathepsin S and a partial sequence of bovine cathepsin L

A. Ritonja[1], A. Čolić[2], I. Dolenc[1], T. Ogrinc[1], M. Podobnik[1] and V. Turk[1]

[1] *Department of Biochemistry, Jožef Stefan Institute, Jamova 39, 61000 Ljubljana, Yugoslavia* and [2] *Centre for Genetic Engineering and Biotechnology, Filipa Kljajića 12, 71000 Sarajevo, Yugoslavia*

Received 18 April 1991

The complete amino acid sequence of bovine spleen cathepsin S has been determined. The single-chain protein contains 217 residues and has a M_r of 23 682. The primary structure was determined by sequencing of native protein and the peptides obtained by proteolytic cleavage with β-trypsin, papaya proteinase IV and by chemical cleavage with cyanogen bromide. Comparison of the amino terminal sequences of the heavy and the light chain of bovine cathepsin L with that of bovine cathepsin S clearly indicates that the enzymes are structurally different.

Cathepsin S; Cathepsin L; Cysteine proteinase; Amino acid sequence

1. INTRODUCTION

The mammalian lysosomal cysteine proteinases cathepsins B (EC 3.4.22.1), H (EC 3.4.22.16), L (EC 3.4.22.15) and cathepsin S (EC 3.4.22.-) belong to the group of closely related proteins of the papain superfamily [1]. The enzymes are considered to play an important role in the initial or terminal stages of protein degradation [2]. Whereas cathepsins B, H and L have been extensively studied and their primary structures are known [3–6], there is only limited knowledge about cathepsin S. This enzyme has been purified and partially characterized for the first time in our laboratory from bovine lymph nodes [7] and spleen [8–11]. It was also purified and characterized from bovine spleen [12,13] and rabbit spleen [14]. Bovine cathepsin S is a single chain protein with Mw of about 24 kDa [8,11] and a preliminary N-terminal amino acid sequence has been determined [15]. The bond-specificity of cathepsin S hydrolysis is in part similar to that of cathepsin L [15,16] and for some time it was thought that differences between both enzymes might be ascribed to species and/or tissue variations in primary structure [17]. This problem may be solved by sequence studies of both enzymes from the same species.

Recent efforts in our laboratory were therefore undertaken in order to prepare sufficient quantities of pure bovine spleen cathepsin S for determination of its primary structure. To determine whether bovine cathepsin S and L indeed differ in their primary structures, we extended our studies to cathepsin L of the same origin.

Correspondence address: A. Ritonja, Department of Biochemistry, Jožef Stefan Institute, Jamova 39, 61000 Ljubljana, Yugoslavia.

Here we report the complete amino acid sequence of bovine cathepsin S and the partial amino acid sequence of bovine cathepsin L. Both sequences clearly show that bovine cathepsin S and L are different enzymes, although they are homologous with other known cysteine proteinases.

2. MATERIALS AND METHODS

Cathepsin S from bovine spleen was purified as previously described [8]. As a last purification step to remove traces of impurities from cathepsin S, ion-exchange rechromatography on CM-cellulose was performed and 2 mg of enzyme were obtained [9]. Cathepsin L was isolated essentially as described in [18]. β-Trypsin was prepared according to [19]. The papaya proteinase IV (PPIV) was a generous gift of Dr A.J. Barrett and prepared as described [20]. 4-Vinylpyridine monomer was from Fluka. *N*-Methylmorpholine, CNBr, guanidine hydrochloride, maleic anhydride and 2-mercaptoethanol were from Pirce. The Cromspher C8 HPLC column was from Chrompack. The chemicals used for Edman degradation were of Sequenal grade from Applied Biosystems. All other chemicals were of analytical grade.

Native active cathepsin S was reduced with 2-mercaptoethanol in 6 M guanidine hydrochloride, and alkylated with 4-vinylpyridine monomer [21]. Hydrolysis of the alkylated protein with β-trypsin was performed in 0.5 M *N*-methylmorpholine pH 8.2 at 38°C for 30 min at the enzyme/substrate ratio 0.01 (w/w). PPIV digestion occurred at 20°C for 7 min at the enzyme/substrate ratio 0.01 (w/w). All other conditions of the hydrolysis were reported previously [22]. Aspartyl-prolyl bond hydrolysis occurred in 80% formic acid and 2% 2-mercaptoethanol (v/v) at 38°C for 40 h. Reversible blocking of amino groups with maleic anhydride and unblocking were as described [23]. Peptides were purified by gel chromatography on Sephacryl S-200 and HPLC on the C8 column eluted with aqueous acetonitrile containing trifluoroacetic acid. Native active bovine cathepsin L was reduced and alkylated as cathepsin S mentioned above.

Samples were sequenced with an Applied Biosystems liquid phase sequenator model 475A. Phenylthiohydantoin derivatives were identified on-line with the 120A HPLC [24]. The reduced and alkylated protein and peptides were hydrolysed with 6 M HCl at 110°C for 24 and 72 h. Amino acid compositions were determined by using a

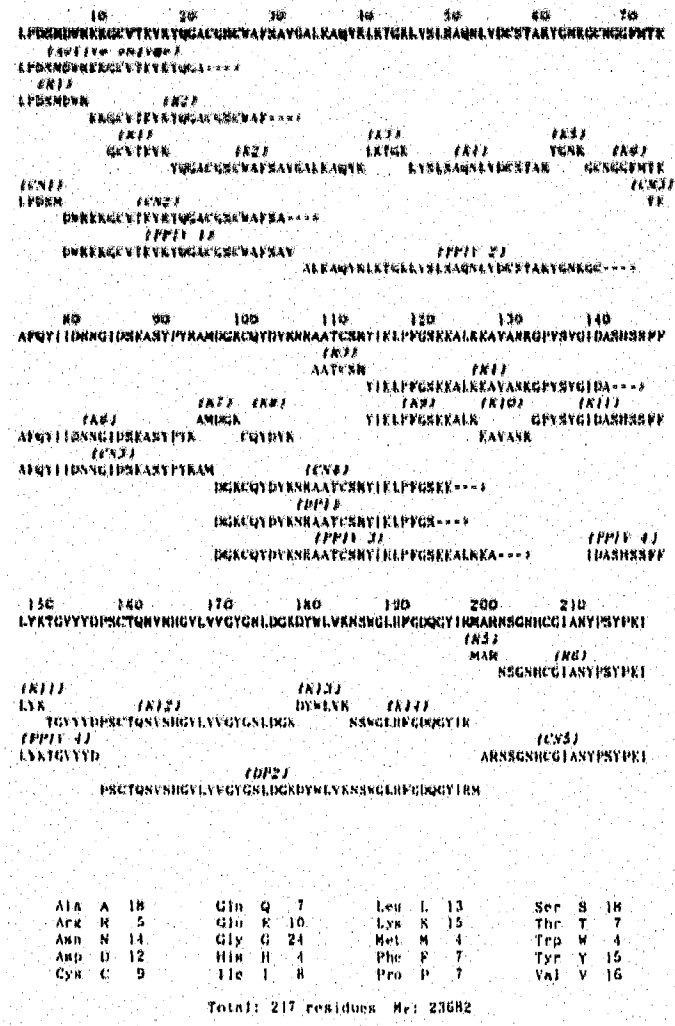

Fig. 1. Amino acid sequence and composition of bovine cathepsin S, and strategy of the sequence determination. The amino acid sequence was determined by automated isothiocyanate degradation of the native enzyme, the 6 tryptic peptides from the maleylated protein (R1–R6) and 14 sub-fragments from R2 and R4 by further digestion with trypsin after de-maleylation (K1–K14). Five CNBr peptides (CN1–CN5) derived from the intact protein, 2 sub-fragments derived from CN4 by acid hydrolysis (DP1, DP2) and 4 PPIV fragments obtained from CN2 and DP1 peptides by further digestion with PPIV (PPIV 1–PPIV 4).

Beckman 118CL analyzer, by post-column fluorescence detection after reaction with o-phthaldialdehyde.

3. RESULTS AND DISCUSSION

Figure 1 describes the complete amino acid sequence as well as the strategy of the sequence determination of a single-chain bovine cathepsin S. These sequence data indicate that cathepsin S is composed of 217 amino acid residues, with a Mw calculated to be 23 682 Da, which is in good agreement with the value obtained by SDS-PAGE [8,11]. Neither a glycosylated residue nor a potential glycosylation site was found. The NH_2-terminal sequence was determined by the sequenator up to residue 21. After reduction, alkylation and reversible blocking of amino groups the first set of peptides was generated by arginine-specific hydrolysis with β-trypsin. Six peptides (R1–R6) were isolated and subjected to amino acid analysis (not shown) and sequence determination. Two of them, R2 (residues 9–106) and R4 (residues 113–199) were not sequenced completely due to their length. After demaleylation both peptides, R2 and R4 were cleaved with trypsin at the Lys-residues. With the exception of di- and tri-peptides, all other peptides were separated and sequenced.

To determine the order of tryptic peptides, a chemical cleavage of the protein with CNBr was performed. In these experiments 5 peptides were obtained (CN1–CN5) although only 3 of them (CN1, CN3 and CN5) were short enough to be sequenced completely. The sequence of the rest of the CN4 peptide was obtained from the DP2 peptide. This peptide generated together with the DP1 peptide from the acid hydrolysis of aspartyl-prolyl bond (residues 155–156) and the sequences of both peptides covered nearly 80% of the CN4 peptide. The peptides CN2 and DP1 were further hydrolyzed with PPIV which specifically cleaves glycyl bonds [22] in the sequence -Val-Gly- to obtain the last missing overlapping sequences. The hydrolysis was complete and the 4 peptides (PPIV 1 - PPIV 4) obtained overlapped tryptic peptides and showed the order as summarized in Fig. 1.

```
                    10        20        30        40        50        60        70
cath.S bov  LPDSMDWREKGCVTEVKYQGACGSCWAFSAVGALEAQV-KLKTGKLVSLSAQNLVDCSTAKYGNKGCNGGFMTE-AFQY
cath.L hu   A*R*V*******Y**P**N**Q*********T****G*MFR-***R*I***E********GPQ-**E*****L*-DY****
cath.L bov  V*W*V**TK***Y**P**N*.....................................................
papain      I*EYV***Q**A***P**N**S**********VTI*GII-*IR**N*NEY*E*E*L**---DRRSY******Y-PWS*L*-

                    80        90        100       110       120       130       140
cath.S bov  IIDNN-GIDSEASYPYKAMD-GKCQYDVKNRA-A--TCS-RYIELPFGSEE-AL-KEAVANKGPVSVGIDASHSSF-FLY
cath.L hu   -VQD*G*L***E****E*TEE-S*K*NP*Y-SV*ND*-G--FVDI*K--Q*K**M*-***TV**I**A***G*E**L*-*
cath.L bov  .....................................................................
papain      -LVAQY**HYRNTPY*EGVQ-RY*RSRE*GPYA*K-*DGV*QVQ-*YN---*G**LY-SI**-Q****VLE*AGKD*Q-**

                    150       160       170       180       190       200       210
cath.S bov  KTGVYYDPSCT-QNVNHGVLVVGYG-HLDGKD---YWLVKNSWGLHFGDQGYIRMARNSGN---HCGIANYPSYPEI
cath.L hu   *E*I*FE*D*SSEDMD***L*****FESTES*NNK*********EEW*MG**VK**KDRR*---******SAA***TV
cath.L bov  ..............................NNKF*I****X*GEX*XG*.....................
papain      RG*IFVG*--*GNK-***A*AA****P*---------*I*I*****TGW*EN****IK*GT**SYGV**LYTSSF**VKN
```

Fig. 2. Alignment of the bovine cathepsin S sequence with those of other cysteine proteinases. Residues identical to those in cathepsin S are marked by an asterisk (*) in the other sequences and the numbering is according to the scheme for papain [26]. *Key:* (cath. S bov) bovine cathepsin S, (cath. L hu) human cathepsin L [5] and (cath. L bov) bovine cathepsin L. References are in the text.

The fragments generated by tryptic (R6) and cyanogen bromide (CN5) cleavages provided the necessary overlaps to establish the carboxyl-terminal sequence of the protein.

In contrast to single-chain cathepsin S, the other 3 lysosomal cysteine proteinases, human cathepsins B, H and L were isolated as two-chain enzymes [5,6]. It was therefore not surprising that bovine cathepsin L exists as a two-chain form of the enzyme (A. Čolić et al., in preparation). Both chains, the heavy chain with Mw of about 20000 Da and the light chain with Mw of about 5000 Da were separated after reduction and alkylation by gel chromatography. Their NH_2-terminal sequences were determined and compared with bovine cathepsin S, and some other members of the same superfamily as shown in Fig. 2. The alignment of bovine cathepsin S with bovine cathepsin L clearly shows the difference between both enzymes. The important difference is at position 12 where Cys was found in cathepsin S, but not in cathepsin L. This residue probably forms a disulfide bridge with Cys at position 108, and was not found in human cathepsin L. An additional difference between the enzymes was found at position 169, where the light chain of bovine cathepsin L begins. From Fig. 2 it is also evident that bovine cathepsin S is structurally homologous to cathepsin L and papain, and belongs to the papain superfamily [1].

In conclusion, we have determined the complete amino acid sequence of bovine cathepsin S and a partial sequence of bovine cathepsin L. Our data clearly demonstrate that these 2 enzymes are different. While this work was in progress, another group reported preliminary cDNA sequence of bovine cathepsin S [25].

Acknowledgements: We thank Dr K. Kramer (Manhattan, Kansas, USA) for the critical reading of the manuscript. This work was supported by a grant from the Research Council of Slovenia.

REFERENCES

[1] Barrett, A.J. (1986) in: Proteinase Inhibitors (Barrett, A.J. and Salvesen, G. eds) pp. 3–22, Elsevier, Amsterdam.
[2] Bond, J.S. and Butler, P.E. (1987) Ann. Rev. Biochem. 56, 333–364.
[3] Takio, K., Towatari, T., Katunuma, N., Teller, D.C. and Titani, K. (1983) Proc. Natl. Acad. Sci. USA 80, 3666–3670.
[4] Ritonja, A., Popovič, T., Turk, V., Widenmann, K. and Machleidt, W. (1985) FEBS Lett. 181, 169–172.
[5] Ritonja, A., Popovič, T., Kotnik, M., Machleidt, W. and Turk, V. (1988) FEBS Lett. 228, 341–345.
[6] Meloun, B., Baudys, M., Pohl, J., Pavlik, M. and Kostka, V. (1988) J. Biol. Chem. 263, 9089–9093.
[7] Turnšek, T., Kregar, I. and Lebez, D. (1975) Biochim. Biophys. Acta 403, 514–520.
[8] Turk, V., Kregar, I., Gubenšek, F. and Ločnikar, P. (1978) in: Protein Turnover and Lysosome Function (Segal, H.L. and Doyle, D.J. eds) pp. 353–361, Academic Press, New York.
[9] Ločnikar, P., Popovič, T., Lah, T., Kregar, I., Babnik, J., Kopitar, M. and Turk, V. (1981) in: Proteinases and their Inhibitors (Turk, V. and Vitale, L.J. eds) pp. 109–116, Mladinska knjiga – Pergamon press, Ljubljana – Oxford.
[10] Kregar, I., Ločnikar, P., Popovič, T., Suhar, A., Lah, T., Ritonja, A., Gubenšek, F. and Turk, V. (1981) Acta Biol. Med. Germ. 40, 1433–1438.
[11] Turk, V., Kregar, I., Popovič, T., Ločnikar, P., Kopitar, M. and Brzin, J. (1980) Period. biol. 82, 363–368.
[12] Kirschke, H., Schmidt, I. and Wiederanders, B. (1986) Biochem. J. 240, 455–459.
[13] Kirschke, H., Wiederanders, B., Bromme, D. and Rinne, A. (1989) Biochem. J. 264, 467–473.
[14] Maciewicz, R.A. and Etherington, D.J. (1988) Biochem. J. 256, 433–440.
[15] Turk, V., Brzin, J., Kopitar, M., Kregar, I., Ločnikar, P., Longer, M., Popovič, T., Ritonja, A., Vitale, Lj., Machleidt, W., Giraldi, J. and Sava, G. (1983) in: Proteinase Inhibitors (Katunuma, N., Umezawa, H. and Holzer, H. eds) pp. 125–134, Japan Sci. Soc. Press, Tokyo.
[16] Bromme, D., Steinert, A., Friebe, S., Fittkau, S., Wiederanders, B. and Kirschke, H. (1989) Biochem. J. 264, 475–481.
[17] Turk, V., Brzin, J., Lenarčič, B., Ločnikar, P., Popovič, T., Ritonja, A., Babnik, J., Bode, W. and Machleidt, W. (1985) in: Intracellular Protein Catabolism (Khairallah, E.A., Bond, J.S. and Bird, J.W.C. eds) pp. 91–103, Alan R. Liss, New York.
[18] Kotnik, M., Popovič, T. and Turk, V. (1986) in: Cysteine Proteinases and their Inhibitors (Turk, V. ed.) pp. 43–50, Walter de Gruyter, Berlin.
[19] Strop, P. and Cechova, D. (1981) J. Chromatogr. 207, 55–62.
[20] Buttle, D.J., Kembhavi, A.A., Sharp, S., Schute, R.E., Rich, D.H. and Barrett, A.J. (1989) Biochem. J. 261, 469–476.
[21] Henschen, A. (1986) in: Advanced Methods in Protein Microsequence Analysis, pp. 244–255, Springer, Berlin.
[22] Buttle, D.J., Ritonja, A., Pearl, L.H., Turk, V. and Barrett, A.J. (1990) FEBS Lett. 260, 195–197.
[23] Butler, P.J.G. and Hartley, B.S. (1972) Methods Enzymol. 25, 191–199.
[24] Hunkapiller, M.W. and Hood, L.E. (1983) Methods Enzymol. 91, 486–493.
[25] Wiederanders, B., Tootham, P., Bromme, D., Kalkkinen, N., Paquette, T., Kirschke, H. and Rinne, A. (1990) Abstracts ICOP Meeting 1990, 8th Conference on Proteolysis, P-46.
[26] Cohen, L.W., Coghlan, W.M. and Dihel, L.C. (1986) Gene 48, 219–227.