Volume 228, number 2, 341-345 FEB 05580 February 1988

Amino acid sequences of the human kidney cathepsins H and L

Anka Ritonja*+, Tatjana Popović*, Matjaž Kotnik*, Werner Machleidt+ and Vito Turk*

*Department of Biochemistry, J. Stefan Institute, Jamova 39, 61000 Ljubljana, Yugoslavia and +Institut für Physiologische Chemie, Physikalische Biochemie und Zellbiologie der Universität München, Goethestrasse 33, D-8000 München 2, FRG

Received 28 December 1987

The complete amino acid sequences of human kidney cathepsin H (EC 3.4.22.16) and human kidney cathepsin L (EC 3.4.22.15) were determined. Cathepsin H contains 230 residues and has an M_r of 25116. The sequence was obtained by sequencing the light, heavy and mini chain and the peptides produced by cyanogen bromide cleavage of the single-chain form of the enzyme. The glycosylated mini chain is a proteolytic fragment of the propeptide of cathepsin H. Human cathepsin L has 217 amino acid residues and an M_r of 23720. Its amino acid sequence was deduced from N-terminal sequences of the heavy and light chains and from the sequences of cyanogen bromide fragments of the heavy chain. The fragments were aligned by comparison with known sequences of cathepsins H and L from other species. Cathepsins H and L exhibit a high degree of sequence homology to cathepsin B (EC 3.4.22.1) and other cysteine proteinases of the papain superfamily.

Cathepsin H; Cathepsin L; Amino acid sequence; Procathepsin; Cysteine proteinase; (Human)

1. INTRODUCTION

Tissue cysteine proteinases cathepsins B, H, L and S are generally thought to play an important role in intracellular protein degradation. Their properties and the first available structural data suggested that these enzymes belong to the papain superfamily [1]. The amino acid sequences of different animal cysteine proteinases such as rat cathepsin B [2,3], rat cathepsin H [2], bovine cathepsin B [4], mouse cysteine proteinase [5], chicken cathepsin L [6,7] and rat cathepsin L [8] support this classification. In addition amino acid sequences of human cathepsin B obtained at both the protein [9] and nucleic acid level [10,11] show high sequence homology to the cysteine proteinases mentioned above, confirming that further evolutionary divergence has occurred within this superfamily.

Here we report for the first time the complete amino acid sequences of enzymatically active cathepsins H and L of human origin.

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

2. MATERIALS AND METHODS

Active human kidney cathepsin H was purified as in [12]. Cathepsin L was isolated from human kidney as described in [13].

All chemicals used were of analytical grade. Iodo[^{3}H]acetic acid was from Amersham. Standards used for the molecular mass determinations were from Serva.

Apparent molecular masses of cathepsins H and L and their pre-existing proteolytic fragments were determined by gel chromatography on a calibrated Sephacryl S-200 column eluted with 80% formic acid as described [16].

For reduction and alkylation, the samples were dissolved in 0.5 M Tris, pH 8.0, containing 5 mM EDTA and 6 M guanidine hydrochloride, reduced with 10 mM 2-mercaptoethanol and then alkylated with 10 mM iodo[^{3}H]acetic acid.

For the determination of free SH groups, native and active cathepsin H or L were alkylated with iodo[^{3}H]acetic acid without prior reduction.

Amino acid analyses were performed with a Kontron Liquimat II analyser using fluorescence detection after a post-column reaction with *o*-phthalaldehyde. The samples were hydrolyzed with 5.7 N HCl in evacuated sealed tubes for 24 and 72 h. Peptides were hydrolyzed for only 24 h. For CNBr cleavage, the alkylated samples were dissolved in 70% formic acid and incubated with a 100-fold molar excess of CNBr over methionyl residues for 48 h in the dark at room temperature. The resulting peptides were separated by gel chromatography on a Sephacryl S-200 column (0.6 × 150 cm) in 80% formic acid and by the reverse-phase HPLC on a Gynkotek RP C18 column using gradients of acetonitrile in 0.1% trifluoroacetic acid.

Published by Elsevier Science Publishers B.V. (Biomedical Division)
00145793/88/$3.50

Amino acid sequences were determined by automated solid-phase Edman degradation using a non-commercial sequencer. The amino acid phenylthiohydantoin derivatives were identified and quantified by the reverse-phase HPLC operating on-line to the sequencer [14]. Prior to sequencing the peptides were covalently coupled to porous glass supports as described [15]. Cysteine was identified as *S*-[^{3}H]carboxymethylcysteine phenylthiohydantoin by liquid-scintillation counting.

3. RESULTS AND DISCUSSION

3.1. *Amino acid sequence of cathepsin H*

Active cathepsin H eluting as a single symmetrical peak on a Sephacryl S-200 column (M_r about 27000) contained four different N-terminal sequences. When chromatographed on the same column after reduction and *S*-carboxymethylation, the sample separated into three peaks [16]. The molecular mass of the first peak remained 27 kDa and yielded two N-terminal sequences in a ratio at 3:2. The higher amount represents the polypeptide chain starting with Tyr and the lower portion corresponds to a chain with two more residues, Gly and Pro, at the N-terminal (fig.1). Probably these two residues have been cleaved off by the aminopeptidase activity of cathepsin H giving rise to the shorter form. Further degradation of the enzyme was stopped by two successive proline residues at the second and third position. The second chromatographic peak was a 42 residues long peptide amounting to approx. 20% of the total enzyme as determined from the yields of the N-terminal sequences. This peptide represents the light chain corresponding to the C-terminal part of the enzyme (fig.1). 80% of the enzyme was present as the single chain. The difference in molecular mass due to the light chain was too small to separate the single-chain form from the heavy chain by gel chromatography. The third peak was a glycosylated octapeptide containing one cysteine residue which was disulfide linked to the rest of the native cathepsin H molecule, as we reported earlier [16]. Using DNA sequence data, this octapeptide, named mini chain, has been found to be part of the propeptide region of the enzyme (Chan, S.J. et al., personal communication). All preparations of the active enzyme tested by N-terminal sequencing contained equimolar amounts of the mini chain and the rest of the enzyme.

The sample from the 27 kDa peak was used for

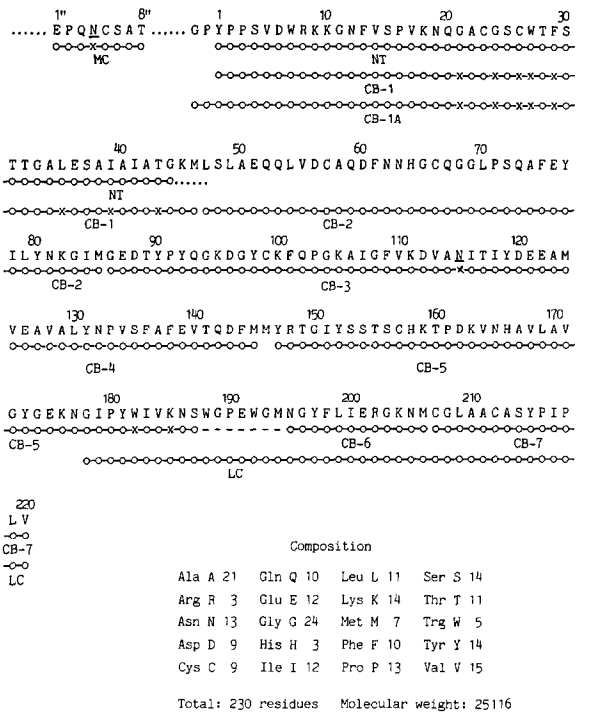

Fig.1. Amino acid sequence and composition of human cathepsin H and strategy of sequence determination. The sequence of the human cathepsin H protein was obtained by N-terminal degradation of the whole enzyme (NT), its light chain (LC), a complete set of cyanogen bromide fragments (CB-1–CB-7) of the *S*-[^{3}H]carboxymethylated protein and of the mini chain (MC) which was separated from the rest of the enzyme after reduction and *S*-carboxymethylation. Numbering of the protein sequence refers to the major component, a minor form was found to be N-terminally extended by two residues (CB-1A). Bars indicate the fragments used for sequence determination; (○) residues that have been directly identified by solid-phase Edman degradation of the individual fragments, (×) residues that have not been identified. Cysteine was detected by radioactivity of its *S*-[^{3}H]carboxymethyl derivative. Glycosylated asparaginyl residues were tentatively assigned to position 115 and position 4″ of the mini chain (underlined) where unknown polar PTH-derivatives were found to yield aspartic acid after hydrolysis.

chemical cleavage at methionyl bonds. The cyanogen bromide peptides (CB-1–CB-7) were sequenced. The strategy for amino acid sequence analysis is outlined in fig.1. N-terminal degradation of the active enzyme resulted in 46 residues which were confirmed by the sequence of the CB-1 peptide. The sequence of the CB-2 peptide comprises the residues 49–87 and that of CB-3 the residues 88–126. Residue 117 is presumably

glycosylated. The peptide CB-5 overlaps the C-terminal part of the heavy chain and the N-terminal part of the light chain in the single-chain form enzyme. CB-7 is the C-terminal part of the enzyme molecule identified by the lack of C-terminal homoserine. Its sequence is identical to the C-terminal sequence of the light chain.

The non-overlapped fragments were aligned by their homology to the corresponding parts of the known sequence of cathepsin H from the rat [2]. The obtained sequence of the human cathepsin H protein was found to be identical with the amino acid sequence predicted from the cDNA sequence of human preprocathepsin H (Chan, S.J. et al., personal communication), with the exception that we found a glutamic acid residue in position 191 instead of glutamine.

Active cathepsin H is composed of 230 amino acid residues, 222 of which form a single chain and an 8 residue long mini chain which is disulfide-linked to the rest of the enzyme. The minor portion of the two chain form results from proteolytic cleavage of a single peptide bond between Asp-177 and Gly-178. In contrast to human cathepsin B [11] and L (see below) the mature cathepsin H has the same C-terminal sequence as its precursor.

3.2. *Amino acid sequence of cathepsin L*

Active cathepsin L with a molecular mass of about 22 kDa yielded two N-terminal sequences in a 1 : 1 molar ratio, as we have reported [16]. After reduction and *S*-carboxymethylation, the sample separated into a 17 kDa heavy chain and 5 kDa light chain on a Sephacryl S-200 column. The heavy chain was cleaved at methionyl residues. Five peptides (CB-1–CB-5) were obtained, purified and sequenced (fig.2). The sequence of the N-terminal peptide CB-1 was known from the N-terminal sequence of the heavy chain. The peptide CB-5 is the C-terminal peptide of the heavy chain because its C-terminal residue is Thr and not the methionyl derivative homoserine.

Alignment of the peptides CB-2, CB-3 and CB-4 was done by comparison with the cDNA-derived amino acid sequences of mouse cysteine proteinase [5] and of rat preprocathepsin L [8]. The resulting sequence of the human cathepsin L protein (fig.2) was found to be consonant with the cDNA-derived amino acid sequence of human preprocathepsin L (Gal, S. et al., personal communication) except for

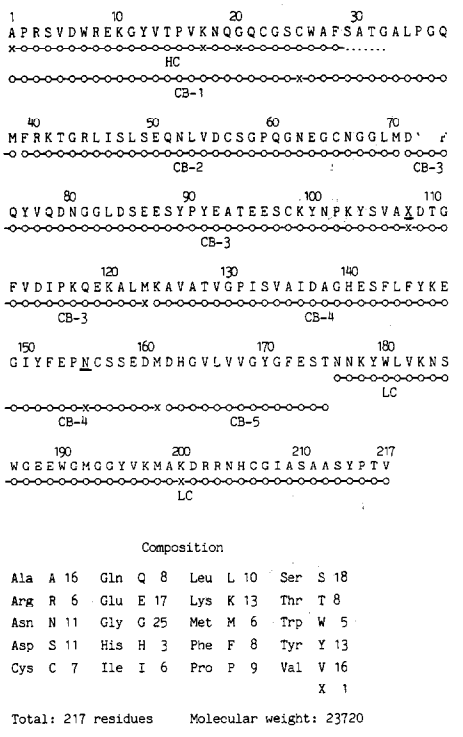

Fig.2. Amino acid sequence and composition of human cathepsin L and strategy of sequence determination. The amino acid sequence of *S*-[^{3}H]carboxymethylated human cathepsin L was determined by the automated solid-phase Edman degradation of the heavy chain (HC), the light chain (LC) and of five cyanogen bromide fragments of the heavy chain (CB-1–CB-5). The fragments used for sequence analysis are indicated by bars; (○) residues that were directly identified by Edman degradation of the individual fragments, (×) residues which were not directly identified. Cysteine was determined by radioactivity of its *S*-[^{3}H]carboxymethyl derivative. Two glycosylation sites of the type N-X-T/S are tentatively assumed at positions 108 and 155 (underlined) where no residue and a low amount of asparagine were found, respectively.

residues 176–178 which are missing at the C-terminus of the heavy chain. From these results we conclude that mature human cathepsin L is composed of 217 amino acid residues forming a heavy chain of 175 residues and a light chain of 42 residues.

In both active cathepsins H and L we have found one free SH-group. We expect that active cathepsin H from the total content of cysteine residues (figs 1,2) contains 4 disulfide bridges, whereas three are present in the active cathepsin L molecule. In human cathepsin B we have found

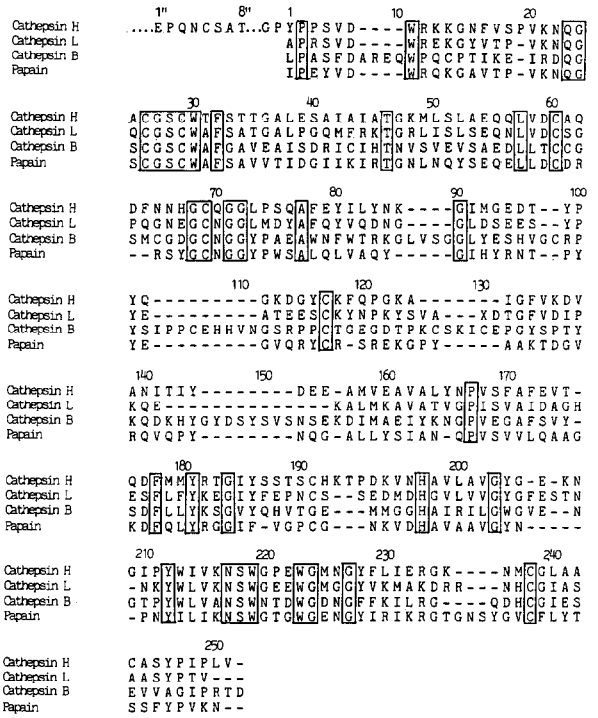

Fig.3. Sequences of the three human cathepsins H, L and B are compared with that of the plant proteinase papain. The sequences were aligned to achieve maximal homology. Residues identical in all four proteinases are boxed. Numbering of residues refers to cathepsin B.

one free SH-group and 13 half cystines [9]. Cathepsin B probably forms 7 disulfide bridges; the missing half cystine should be positioned in the propeptide region. It seems that the 'mini chain' of cathepsin B containing this cysteine residue was lost after reduction of the starting samples.

3.3. *Sequence homology*

The sequences of human cathepsins H and L are compared with those of human cathepsin B and of papain in fig.3. For comparison the sequences are arbitrarily divided into three regions [2] as shown in table 1. In the active site cysteinyl region the three human cathepsins show 31–56% of identity, the highest (56%) being between cathepsins L and H and the lowest (31%) when cathepsin H is compared to cathepsin B. 24% of the residues are identical in all four proteinases. These structural similarities of cathepsins and papain are probably due to divergence from a common ancestral gene

Table 1

Homology of amino acid sequences of human cathepsins B, H and L and papain

	Residues in CB			
	1–252 (whole protein)	1–77 (active site cysteinyl region)	78–152 (central region)	153–252 (active site histidyl region)
CH/CB	30.0	31.1	15.5	35.6
CL/CB	29.9	34.2	19.5	31.6
CL/CH	45.1	56.1	30.4	43.8
Ident. resid.[a]	13.8	24.3	9.3	11.1

[a] Identical residues in all four enzymes as indicated in fig.3

Identity in (%) was calculated from the sequence alignment shown in fig.3. (CH) cathepsin H, (CB) cathepsin B, (CL) cathepsin L

during evolution. The degree of identity in the central region is low (9.3%) in spite of a rather liberal placement of the gaps. The functional significance of the central region is not clear even in papain, where the three-dimensional structure is established. In the C-terminal parts where the active site hystidyl regions are located, a lower degree of identity is observed compared to the cysteinyl regions (31–43%). Also, in this region cathepsin L is closely related to cathepsin H (43% identity) whereas it shows a lower homology with cathepsin B (31% identity). The sequences in the vicinity of the active site hystidinyl residue may reflect the different peptidase specificities of the cathepsins as was suggested for cathepsin B [2]. Whereas cathepsin L shows only endopeptidase activity [22], cathepsins B and H act as endopeptidases as well as dipeptidyl peptidase [17–19] and aminopeptidase [20,21], respectively.

Acknowledgements: This work was supported by the Research Council of Slovenia (grant C1-0515-106 to V.T.), by the Sonderforschungsbereich 207 of the University of Munich (grant C-2 to W.M.) and by the International Bureau of the Kernforschungsanlage Jülich GmbH.

REFERENCES

[1] Barrett, A.J., Nicklin, M.J.H. and Rowlings, N.D. (1984) Symp. Biol. Hung. 25, 203–217.
[2] Takio, K., Towatari, T., Katunuma, N., Teller, D.C. and Titani, K. (1983) Proc. Natl. Acad. Sci. USA 80, 3666–3670.

[3] San Segundo, B., Chan, S.J. and Steiner, D.F. (1985) Proc. Natl. Acad. Sci. USA 82, 2320–2324.
[4] Meloun, B., Pohl, J. and Kostka, V. (1986) in: Cysteine Proteinases and their Inhibitors (Turk, V. ed.) pp.19–29, Walter de Gruyter, Berlin.
[5] Portnoy, D.A., Erickson, A.H., Kochan, J., Ravetch, J.V. and Unkeless, J.C. (1986) J. Biol. Chem. 261, 14697–14703.
[6] Wada, K., Takai, Y. and Tanabe, T. (1987) Eur. J. Biochem. 167, 13–18.
[7] Dufour, E., Obled, A., Valin, S., Bechet, D., Ribadeau-Dumas, B. and Huet, J.C. (1987) Biochemistry 26, 5689–5695.
[8] Ishidoh, K., Towatari, T., Imajoh, S., Kawasaki, S., Kominami, I., Katunuma, N. and Suzuki, K. (1987) FEBS Lett. 223, 69–73.
[9] Ritonja, A., Popović, T., Turk, V., Wiedermann, K. and Machleidt, W. (1985) FEBS Lett. 181, 169–172.
[10] Fong, D., Calhoun, D.H., Hsieh, W.T., Lee, B. and Wells, R.D. (1986) Proc. Natl. Acad. Sci. USA 83, 2909–2913.
[11] Chan, S.J., San Segundo, B., McCormick, M.B. and Steiner, D.F. (1986) Proc. Natl. Acad. Sci. USA 83, 7721–7725.
[12] Popović, T., Brzin, J., Kos, J., Lenarčič, B., Machleidt, W., Ritonja, A. and Turk, V. (1988) Biol. Chem. Hoppe-Seyler, in press.
[13] Kotnik, M., Popović, T. and Turk, V. (1986) in: Cysteine Proteinases and their Inhibitors (Turk, V. ed.) pp.43–50, Walter de Gruyter, Berlin.
[14] Machleidt, W. (1983) in: Modern Methods in Protein Chemistry (Tschesche, H. ed.) pp.267–302, Walter de Gruyter, Berlin.
[15] Machleidt, W., Borchart, U. and Ritonja, A. (1986) in: Advanced Methods in Microsequence Analysis (Wittmann-Liebold, B. et al. eds) pp.91–107, Springer, Heidelberg.
[16] Machleidt, W., Ritonja, A., Popović, T., Kotnik, M., Brzin, J., Turk, V., Machleidt, I. and Müller-Esterl, W. (1986) in: Cysteine Proteinases and their Inhibitors (Turk, V. ed.) pp.3–18, Walter de Gruyter, Berlin.
[17] Aronson, N.N. and Barrett, A.J. (1978) Biochem. J. 171, 759–765.
[18] Nakai, N., Wada, K., Kobashi, K. and Hase, J. (1978) Biochem. Biophys. Res. Commun. 83, 881–885.
[19] Bond, J.S. and Barrett, A.J. (1980) Biochem. J. 189, 17–25.
[20] Kirschke, H., Langner, J., Wiederanders, B., Ansorge, S., Bohley, P. and Hanson, H. (1977) Acta Biol. Med. Germ. 36, 185–199.
[21] Singh, H. and Kalnitsky, G. (1980) J. Biol. Chem. 255, 369–374.
[22] Barrett, J.A. and Kirschke, H. (1981) Methods Enzymol. 80, 535–561.