Volume 236, number 1, 57–61 FEB 06086 August 1988

Amino acid sequence of rat liver cathepsin L

Takae Towatari and Nobuhiko Katunuma

Division of Enzyme Chemistry, Institute for Enzyme Research, The University of Tokushima, Tokushima 770, Japan

Received 21 May 1988

The complete amino acid sequences of the heavy and light chains of rat liver cathepsin L (EC 3.4.22.15) were determined at the protein level. The heavy and light chains consisted of 175 and 44 amino acid residues, respectively, and their M_r values without glycosyl groups calculated from these sequences were 18941 and 5056, respectively. The amino acid sequence was also determined from the N-terminal sequences of the heavy and light chains, and the sequences of cleavage fragments of the heavy chain with lysylendopeptidase and cyanogen bromide. The fragments were aligned by comparison with the amino acid sequence deduced from the sequence of cDNA of rat preprocathepsin L. The sequence of rat liver cathepsin L determined at the protein level was identical with that deduced from the cDNA sequence except that in the heavy chain, residues 176–177 (Asp-Ser) were not present at the C-terminus and alanine was replaced by proline at residue 125. Asn-108 in the heavy chain is modified with carbohydrate.

Cathepsin L; Cysteine protease; Amino acid sequence; Lysosome

1. INTRODUCTION

Lysosomal cysteine proteinases such as cathepsin L, B and H are thought to be important in the degradation of intracellular proteins [1,2]. Kirschke et al. [3] and ourselves [4,5] have independently isolated cathepsin L and found that it has much higher endopeptidase activity than cathepsin B or H. Subsequently cathepsin L has been purified from mammalian and avian tissues and its properties have been studied in detail [6–9]. To understand the relationship between the structure and function of cathepsin L, its primary structure is of interest. The complete amino acid sequences of chicken [10,11] and human [12] cathepsin L and MCP [13] have been reported. Moreover, we have deduced the amino acid sequence of rat preprocathepsin L from the nucleotide sequence of cloned cDNA and shown that it is highly homologous (94%) with that of MCP and that it belongs to the papain family [14,15].

This paper reports the complete amino acid sequence of rat liver cathepsin L determined at the protein level, and its comparison with the amino acid sequence determined at the nucleotide level.

Correspondence address: N. Katunuma, Division of Enzyme Chemistry, Institute for Enzyme Research, The University of Tokushima, Tokushima 770, Japan

Abbreviations: HPLC, high-performance liquid chromatography; SDS-PAGE, SDS-polyacrylamide electrophoresis; Con A, concanavalin A; MCP, mouse cysteine proteinase; TFA, trifluoroacetic acid; PTH, phenylthiohydantoin

2. MATERIALS AND METHODS

Rat liver cathepsin L was purified to homogeneity by SDS-PAGE as described previously [5] with the additional purification steps of column chromatography on Con A-Sepharose and HPLC on TSK gel G3000 SW (Toyo Soda).

Lysylendopeptidase was obtained from Wako Pure Chemical Ind. (Tokyo, Japan). Carboxypeptidase Y was purchased from Oriental Yeast Co. (Tokyo, Japan). The reagents used for automated gas-phase sequencing were obtained from Applied Biosystem Inc. (Foster City, USA). All other chemicals were either HPLC grade or analytical grade.

Reduction and *S*-carboxymethylation of cathepsin L were carried out as reported [16]. The sample was dissolved in 1.5 M Tris-HCl, pH 8.6, containing 6 M guanidine hydrochloride and was reduced with dithiothreitol at room temperature for 2 h.

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

Thiol groups were carboxymethylated by incubation for 30 min in the dark with iodoacetamide. For isolation of the chemically modified heavy and light chains, the reaction mixture was applied directly to a HPLC column of TSK gel G3000SW equilibrated with 100 mM potassium phosphate buffer, pH 6.5, containing 6 M guanidine hydrochloride or subjected to reverse-phase HPLC on COSMOSIL 5C4-300 (Jasco) in a gradient of acetonitrile in 0.1% TFA. For cleavages of lysyl bonds [17], the *S*-carboxymethylated heavy chain was incubated at a molar ratio of 1:200 with lysylendopeptidase in 100 mM Tris-HCl, pH 9.0, for 15 h at 30°C. The reaction was stopped by addition of 10% formic acid (final concentration). For cleavage of methionyl bonds, the *S*-carboxymethylated heavy chain was dissolved in 70% formic acid containing a 100-fold molar excess of cyanogen bromide over methionyl residues and incubated for 24 h at room temperature [18]. The digestion was terminated by lyophilization. The resulting peptides were purified by reverse-phase HPLC on a column of TSK gel/ODS120T (Toyo Soda) or COSMOSIL 5C4-300 (Jasco) with a linear gradient of acetonitrile in 0.1% (v/v) TFA. Peptides were detected by their absorbance at 215 nm. For carboxypeptidase Y digestion, the *S*-carboxymethylated heavy chain or light chain was digested at a molar ratio of 1:100 with carboxypeptidase Y in 50 mM pyridine-acetate buffer, pH 6.5, containing 0.1% (w/v) SDS. Aliquots were removed after 0, 30, 60, and 90 min. The reaction was stopped by adding formic acid (10% final concentration) and the mixture was dried under vacuum. Amino acid analyses were performed with a Hitachi amino acid analyser (Hitachi 835). Amino acid sequences were determined with an Applied Biosystems 470A protein sequencer/Spectra Physics SP8100 HPLC system.

3. RESULTS AND DISCUSSION

Purified rat liver cathepsin L has been found to consist of two chains, a heavy chain (approx. 25 kDa) and a light chain (approx. 5 kDa). These two chains were separated by cleavage of disulfide bonds holding the two together, and the separated chains were subjected to automated sequence analysis. The complete amino acid sequence of rat liver cathepsin L obtained at the protein level and the overall strategy for analysis of the amino acid sequence are presented in fig.1.

The light chain was sequenced directly by automated Edman degradation. Two N-terminal sequences were obtained, a major one with N-terminal Asn and a minor one with N-terminal Lys. The ratio of the minor sequence to the major sequence was about 1/3 in mol of PTH amino acids. These sequences were attributed to residues 178–221 and residues 179–221, respectively, of the amino acid sequence predicted from the nucleotide sequence of cloned rat preprocathepsin L cDNA [14]. These results show that Asn-178 is the original N-terminal residue in the light chain of rat liver cathepsin L and that Lys-179 at the N-terminus is probably formed by an aminopeptidase attack on the light chain.

The sequence of the heavy chain of rat liver cathepsin L was determined by direct automated Edman degradation of the *S*-carboxymethylated heavy chain, and of fragments formed by digestion with lysylendopeptidase and by cleavage with cyanogen bromide. Direct automated Edman degradation of the *S*-carboxymethylated heavy chain stopped after 40 cycles. Five long fragments (residues 45–78, 79–99, 100–120, 125–159 and 160–175) and four short fragments (residues 11–17, 42–44, 95–99 and 121–124) in the digest of the heavy chain with the lysylendopeptidase were sequenced completely. As predicted from the presence of four methionine residues in the heavy chain of the amino acid sequence predicted from the cDNA of rat cathepsin L [14], five fragments were obtained by cyanogen bromide cleavage of the heavy chain and these were partially sequenced.

The peptides were aligned by comparison with the amino acid sequence predicted for rat preprocathepsin L. During Edman degradation, no amino acid could be identified at residue 108 of the heavy chain and the amino acid residue two steps after residue 110 in the sequence was found to be threonine. These results show that residue 108 of the heavy chain is a carbohydrate-linked asparagine residue. The sequence is essentially the same as that predicted from the cDNA sequence of rat cathepsin L reported previously [14]. The predicted sequence suggested that residue 155 in the heavy chain might be another glycosylation site, but asparagine residue was detected at position 155 of the heavy chain. Thus, Asn-108 was concluded to be the only glycosylation site in rat liver cathepsin L, as in the chicken enzyme. In human cathepsin L, two glycosylation sites have been identified at the protein level [12]. Digestions of the heavy and light chains with carboxypeptidase Y indicated that their C-terminal residues were threonine and asparagine, respectively. Additional results were consistent with the sequences Glu-Gly-Thr in the heavy chain and Ile-Val-Asn in the light chain.

The resulting sequence of rat liver cathepsin L shown in fig.1 was consistent with that deduced

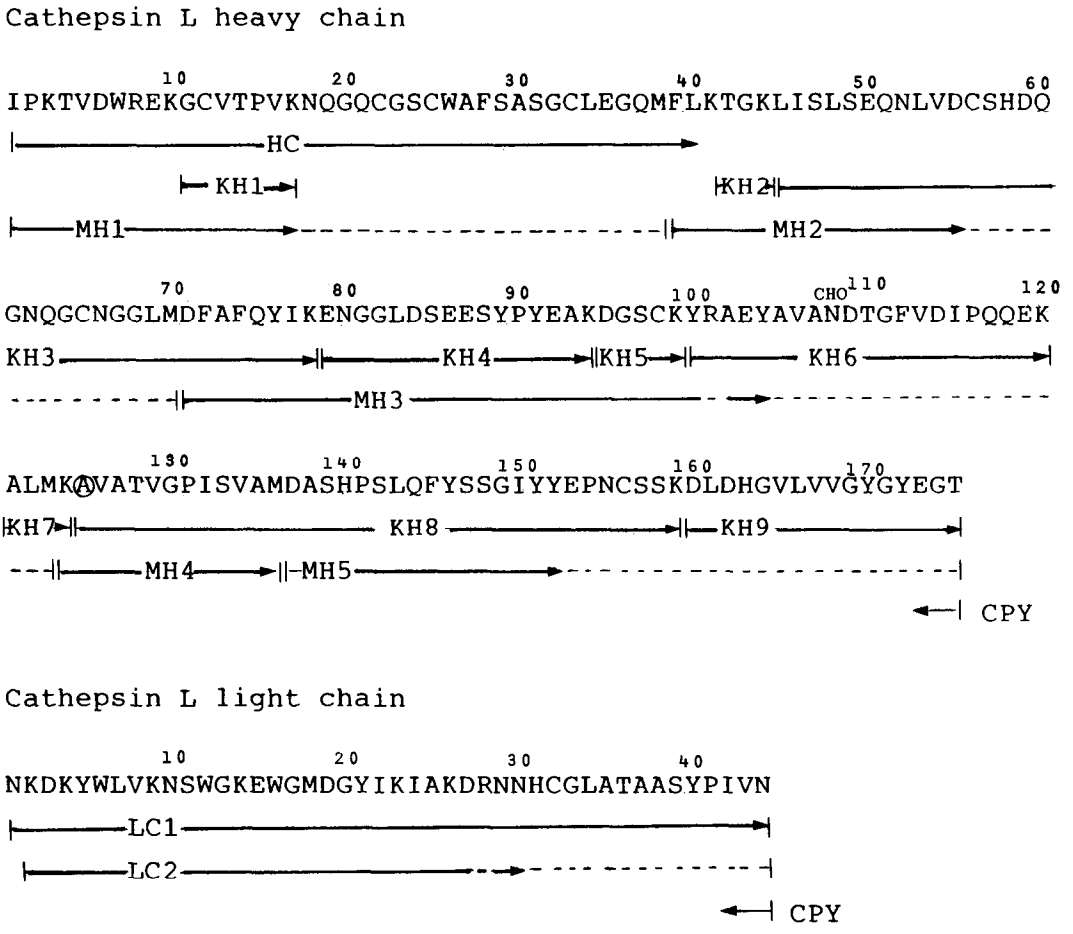

Fig.1. The amino acid sequence of rat liver cathepsin L and the strategy for sequence determination. The sequence of rat liver cathepsin L was obtained by N-terminal degradation of the heavy chain (HC) and the light chain (LC), and fragments obtained by cleavage with lysylendopeptidase (KH), cyanogen bromide (MH) and carboxypeptidase (CPY). The average repetitive yield was 94%. (⟶) Edman degradation step; (⟵) carboxypeptidase Y digestion (CPY); (---) not determined. Asn-108 carries a carbohydrate (CHO). The encircled residue differs from that in the cDNA-derived amino acid sequence [14].

from the sequence of cDNA of rat cathepsin L except that in the heavy chain residues 176–177 (Asp-Ser) were not present at the C-terminus and residue 125 was alanine, not proline. Cleavages at two sites between residues 175 and 177 with loss of a dipeptide to generate the two-chain form probably occur within lysosomes [19] and the substitution of an amino acid at residue 125 in the amino acid sequence of rat cathepsin L may be due to a cloning artifact. From these results, we conclude that purified rat liver cathepsin L is composed of 219 amino acid residues and consists of a heavy chain of 175 residues and a light chain of 44 residues. The M_r values of the heavy and light chains, without glycosyl groups, calculated from these sequences are 18941 and 5056, respectively.

In fig.2, the sequence of rat liver cathepsin L is compared with those of human [12] and chicken [10,11] cathepsin L and MCP [14] and also the sequences of rat cathepsins H and B. As already discussed, the deduced amino acid sequence of rat preprocathepsin L is highly homologos (94%) to that of MCP, whose sequence was determined by cDNA cloning. The amino acid sequences of rat cathepsin L and MCP can be aligned without any gaps and most of the differences between them are

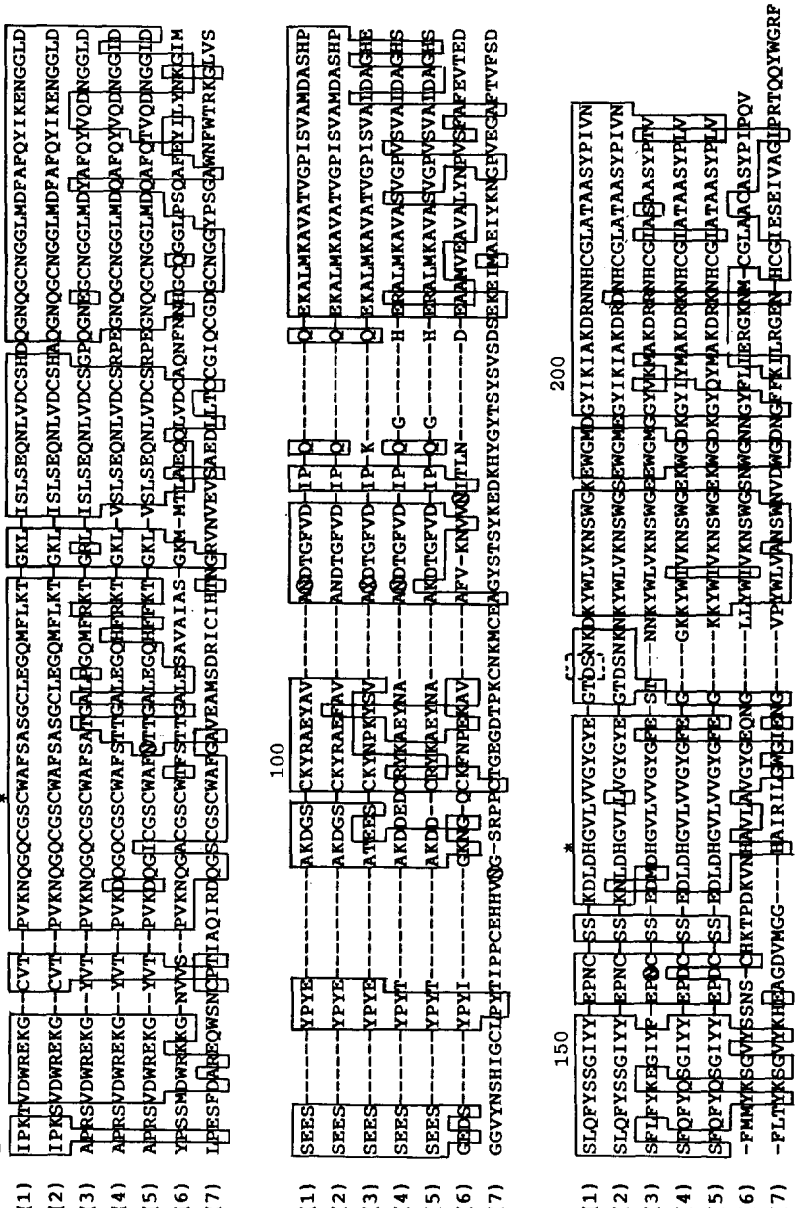

Fig.2. Comparison of the amino acid sequence of rat cathepsin L with those of MCP [13], human cathepsin L [12], chicken cathepsin L [10,11] and rat cathepsins H and B [20]. Identical amino residues to those in rat cathepsin L are boxed. Gaps (–) are inserted to obtain maximal homology. The stars denote the active-site residues cysteine and histidine. Glycosylated residues are encircled. Residues enclosed in dashed boxes are not present in the two-chain form of rat cathepsin L.

substitutions between conserved amino acids. Moreover, in the sequence of rat liver cathepsin L obtained at the protein level, residue 125 is the same amino acid (alanine) as in MCP. The sequence of rat liver cathepsin L shows 78% homology with that of human cathepsin L and 74–76% homology with that of chicken cathepsin L. The overall sequence homology of rat liver cathepsin L is 45% with rat cathepsin H and 25% with rat cathepsin B [20], when gaps introduced to maximize sequence homology are considered as mismatches. These sequences suggest that the structural similarities of cathepsin L with other cysteine proteinases are probably due to divergence of these enzymes from a common ancestral gene during evolution.

Acknowledgements: We would like to thank Mr H. Miyai for assistance in the amino acid analysis and Ms E. Inai for assistance in the preparation of this manuscript. This work was supported by a Grant-in-Aid for Scientific Research from the Ministry of Education, Science and Culture of Japan.

REFERENCES

[1] Katunuma, N., Towatari, T., Kominami, E. and Hashida, S. (1981) in: Proteinases and Their Inhibitors (Turk, V. and Vitale, L. eds) pp.83–92, Pergamon, Oxford.
[2] Barrett, A.J. and Kirschke, H. (1981) Methods Enzymol. 80, 535–561.
[3] Kirschke, H., Langner, J., Wiederander, B., Ansorge, S. and Bohldy, P. (1977) Eur. J. Biochem. 74, 293–301.
[4] Towatari, T., Tanaka, K., Yoshikawa, D. and Katunuma, N. (1976) FEBS Lett. 67, 284–288.
[5] Towatari, T., Tanaka, K., Yoshikawa, D. and Katunuma, N. (1978) J. Biochem. 84, 659–671.
[6] Mason, R.W., Taylor, M.A. and Etherington, D.J. (1984) Biochem. J. 217, 209–217.
[7] Okitani, A., Matsukura, U., Kato, H. and Fujimaki, M. (1980) J. Biochem. (Tokyo) 87, 33–43.
[8] Bando, Y., Kominami, E. and Katunuma, N. (1986) J. Biochem. (Tokyo) 100, 35–42.
[9] Wada, K. and Tanabe, T. (1986) FEBS Lett. 209, 330–334.
[10] Wada, K., Takai, T. and Tanabe, T. (1987) Eur. J. Biochem. 167, 13–18.
[11] Dufour, E., Obled, A., Valin, C. and Béchet, D. (1987) Biochemistry 26, 5689–5695.
[12] Ritonja, A., Popvić, T., Kotnik, M., Machleidt, W. and Turk, V. (1988) FEBS Lett. 228, 341–345.
[13] Portnoy, D.A., Erickson, A.H., Kochan, J., Ravetch, J.V. and Unkeless, J.C. (1986) J. Biol. Chem. 261, 14697–14703.
[14] Ishidoh, K., Towatari, T., Imajoh, S., Kawasaki, H., Kominami, E., Katunuma, N. and Suzuki, K. (1987) FEBS Lett. 223, 69–73.
[15] Drenth, J., Jansonius, J.N., Koekoek, R. and Walthers, B.G. (1971) Adv. Protein Chem. 25, 79–115.
[16] Takio, K., Towatari, T., Katunuma, N. and Titani, K. (1980) Biochem. Biophys. Res. Commun. 97, 340–346.
[17] Tsunasawa, S., Sugihara, A., Masaki, T., Sakiyama, F., Takeda, Y., Mawatani, T. and Narita, K. (1987) J. Biochem. 101, 111–121.
[18] Gross, E. and Witkop, B. (1962) J. Biol. Chem. 237, 1856–1860.
[19] Hara, K., Kominami, E. and Katunuma, N. (1988) FEBS Lett., in press.
[20] Takio, K., Towatari, T., Katunuma, N., Teller, D.C. and Titani, K. (1983) Proc. Natl. Acad. Sci. USA 80, 3666–3670.