The *N*-terminal amino acid sequences of the heavy and light chains of human cathepsin L

Relationship to a cDNA clone for a major cysteine proteinase from a mouse macrophage cell line

Robert W. MASON,* John E. WALKER† and Frederick D. NORTHROP†

*Department of Biochemistry, Strangeways Research Laboratory, Worts Causeway, Cambridge CB1 4RN, U.K., and

†Medical Research Council, Laboratory of Molecular Biology, Hills Road, Cambridge CB2 2QH, U.K.

Human liver cathepsin L consists of a heavy chain and a light chain with M_r values of 25000 and 5000 respectively. The chains have been purified and their *N*-terminal amino acid sequences have been determined. The 40 amino acids determined from the heavy chain and 42 amino acids sequenced in the light chain are homologous with the *N*-terminal and *C*-terminal regions respectively of the superfamily of cysteine proteinases. Therefore it is likely that the two chains of cathepsin L are derived by proteolysis of a single polypeptide precursor. Of the amino acids sequenced, 81% are identical with the homologous portions of a protein sequence for a major cysteine proteinase predicted from a cDNA clone from a mouse macrophage cell line. This is the closest relative amongst the known sequences in the superfamily and strongly indicates that the protein encoded by this mRNA is cathepsin L. The mouse protein is also probably the major excreted protein of a transformed cell line [Gal & Gottesman (1986) *Biochem. Biophys. Res. Commun.* **139**, 156–162]. The heavy chain is identical in only 71% of its residues with the sequence of ox cathepsin S, providing further evidence that this latter enzyme is probably not a species variant of cathepsin L. The relationship with a second unidentified cathepsin cDNA clone from a bovine library is much weaker (41% identity), and so this clone remains unidentified.

INTRODUCTION

Cathepsins B, H and L and calpains I and II are the best characterized mammalian cysteine proteinases (Barrett & Kirschke, 1981; Murachi, 1984). Cathepsins B and H and calpain II have all been sequenced and are homologous with papain (Takio *et al.*, 1983; Ohno *et al.*, 1984; Ritonja *et al.*, 1985). Two different cDNA clones have been characterized that encode proteins that are homologous with this family of enzymes (Gay & Walker, 1985; Portnoy *et al.*, 1986). The latter encodes a cysteine proteinase that is a major secreted protein from a mouse macrophage cell line (Portnoy *et al.*, 1986).

Cathepsin L has been purified from livers of rats, rabbit and man. All of these enzymes have similar catalytic properties and always contains a heavy and light chain with M_r values of 25000 and 5000 respectively (Kirschke *et al.*, 1977; Mason *et al.*, 1984, 1985). In addition, an enzyme from ox spleen, called cathepsin S, has a specific activity for azocasein similar to that of cathepsin L but has a lower specific activity for the synthetic substrate Z-Phe-Arg-NHMec and is less sensitive to inhibition by the active-site-directed inhibitor Z-Phe-Phe-CHN₂ (Kirschke *et al.*, 1984). However, it is not clear from published data, including its *N*-terminal sequence (Turk *et al.*, 1983), whether cathepsin S is a species variant of cathepsin L or whether it is a different enzyme.

In experiments described below we have determined the *N*-terminal sequences of the two chains of human

cathepsin L in order to examine the structural similarity of this enzyme to cathepsin S and to attempt to identify the cDNA clones encoding the hitherto-unidentified cysteine proteinases. The sequences determined are closely related to the sequence of a hitherto-unidentified cathepsin encoded in a mouse cDNA, strongly suggesting that this major mouse cysteine proteinase is cathepsin L.

METHODS

Preparation of the enzyme

Human liver cathepsin L was purified as described previously and SDS/polyacrylamide-gel electrophoresis under reducing conditions showed that the enzyme consists of two chains, heavy and light, with M_r values of 25000 and 5000 respectively (Mason *et al.*, 1985).

Protein chemical methods

S-[¹⁴C]Carboxymethylated enzyme was prepared as follows. The enzyme (150 µg), which was dissolved in 20 mM-sodium acetate buffer (2 ml), pH 5.0, containing 1 mM-EDTA and 400 mM-NaCl, was transferred into a solution containing 0.1 M-Tris/HCl, pH 8.3, 6 M-guanidinium chloride and 1 mM-EDTA by chromatography on Sephadex G-25 (fine grade). N₂ was bubbled through the protein solution. Dithiothreitol (4 µmol) was added and thereby the protein was reduced, at 37 °C for 1 h. The protein was then allowed to react in the dark with iodo[2-¹⁴C]acetic acid (8.5 µmol, 25 µCi) at 37 °C

Abbreviations used: NHMec, 7-(4-methyl)coumarylamide; Ep-459, 1-3-carboxy-*trans*-2,3-epoxypropyl-leucylamido-(4-amino)butane; Z, benzyloxycarbonyl; -CHN₂, diazomethane.

for a further 1 h period. Excess reagent was inactivated by addition of 2-mercaptoethanol to a final concentration of 10 mM. The *S*-carboxymethylated heavy and light chains of cathepsin L and unincorporated iodoacetic acid were separated from each other by chromatography on a column (30 cm × 1 cm int. diam.) of Sephadex G-50 (superfine grade) in 100 mM-Tris/HCl, pH 8.3, containing 6 M-guanidinium chloride. Fractions were examined by SDS/polyacrylamide electrophoresis. Samples were loaded in guanidinium chloride. A white precipitate formed in the slots of the gel, but this did not adversely affect subsequent electrophoresis. Pooled fractions containing heavy and light chains were then equilibrated with 100 mM-ammonium bicarbonate by chromatography on a column (30 cm × 1 cm int. diam.) of Sephadex G-25 (fine grade), and then freeze-dried.

The *N*-terminal sequences of the two chains were determined with the aid of an Applied Biosystems 470A gas-phase protein sequencer. Phenylthiohydantoin from the heavy chain were identified by offline h.p.l.c. (Downward *et al.*, 1984); those from the light chain were identified by online chromatography with an Applied Biosystems 120A liquid chromatograph. The total radioactivity in each amino acid derivative released by the sequencer was estimated in order to determine the positions of *S*-[¹⁴C]carboxymethylcysteine residues.

RESULTS AND DISCUSSION

Purification of the heavy and light chains of human cathepsin L

The reduced and carboxymethylated heavy and light chains of human cathepsin L were separated by chromatography on Sephadex G-50 (superfine grade) as described in Fig. 1(a) and their purity was confirmed by gel electrophoresis (Fig. 1b). The apparent M_r values estimated by SDS/polyacrylamide electrophoresis of the two chains were 30000 and 5000. Thus *S*-carboxymethylation of the heavy chain results in the protein migrating more slowly than reduced, non-carboxymethylated, protein in SDS/polyacrylamide-gel electrophoresis.

N-Terminal sequences of the heavy and light chains

By protein-sequence analysis of the heavy chain, residues 1–40 were identified (see Table 1). Initial sequence analysis experiments with samples of the light chain indicated that it was heterogeneous at its *N*-terminus, presumably because of autolysis. To circumvent this problem, the purified enzyme was freeze-dried before reduction and carboxymethylation as this inactivates human cathepsin L (R. W. Mason, unpublished work). The *S*-carboxymethylated light chain, isolated from inactivated enzyme, was found to be homogeneous, and 42 amino acids were determined from the *N*-terminus (Table 2). The M_r corresponding to this sequence is 4776. This, and the alignment with the mouse macrophage protein up to the penultimate residue (Fig. 2), suggest that the sequence represents all, or almost all, of the sequence of the light chain, although further experimentation is needed to be certain of this.

Alignment of these sequences with members of the mammalian cysteine proteinases superfamily (Fig. 2) demonstrates that the heavy chain is likely to be derived from the *N*-terminal region and the light chain from the

Fig. 1. Purification of H- and L-chains from reduced and *S*-[¹⁴C]carboxymethylated human cathepsin L

(a) Chromatography on Sephadex G-50 (superfine grade). Fractions (1 ml) were collected. Δ , A_{280} ; $\circ$, radioactivity (d.p.m./ml). Fractions 6–8 (heavy chain) and 18–20 (light chain) were pooled and employed in sequencing experiments. For further details, see the Methods section. (b) Polyacrylamide-gel electrophoresis of fractions in (a). A 12% acrylamide gel with 1.3% bisacrylamide was used (Bury, 1981). A 25 μ l portion of fraction was loaded on to the gel.

C-terminal part of a single-chain precursor. It also shows that the active-site cysteine residue, at position 25, of this superfamily of proteinases (Barrett *et al.*, 1984) is conserved in the heavy chain of cathepsin L. This is in keeping with the finding that the radiolabelled active-site-directed inhibitor *N*-[³H]acetyl-Ep-459 binds to the heavy chain of the intact enzyme (Mason *et al.*, 1985).

Comparison with the mouse cDNA clone

Of the 82 amino acids identified in human cathepsin L, 81% are identical with the protein sequence predicted from a mouse cDNA clone (Portnoy *et al.*, 1986). This suggests that the two proteins are in the same subfamily according to the terminology of Dayhoff *et al.* (1983), and that the cDNA clone encodes mouse cathepsin L. A similar level of identity is found between cathepsin B of

Table 1. N-Terminal sequence of human cathepsin L heavy chain

About 4 nmol of protein were employed. Abbreviations used: PTH, phenylthiohydantoin; CmC, S-carboxymethylcysteine; —, not quantified.

Cycle	PTH identity	PTH yield (nmol)	Radioactivity released (c.p.m.)
1	Ala	> 3.9	604
2	Pro	2.53	280
3	Arg	—	247
4	Ser	—	744
5	Val	2.09	674
6	Asp	0.90	498
7	Trp	1.46	358
8	Arg	—	347
9	Glu	1.19	1253
10	Lys	1.57	939
11	Gly	0.97	657
12	Tyr	1.24	570
13	Val	1.79	545
14	Thr	—	585
15	Pro	0.90	685
16	Val	0.77	607
17	Lys	1.12	605
18	Asn	0.44	973
19	Gln	0.48	724
20	Gly	0.56	720
21	Gln	0.41	709
22	CmC	—	2159
23	Gly	0.27	1979
24	Ser	—	1402
25	CmC	—	2128
26	Trp	0.20	1763
27	Ala	0.34	1011
28	Phe	0.38	995
29	Ser	—	964
30	Ala	0.26	936
31	Thr	—	836
32	Gly	0.14	536
33	Ala	0.106	641
34	Leu	0.23	738
35	Pro	0.086	453
36	Gly	0.055	583
37	Gln	0.096	566
38	Met	0.124	98
39	Phe	0.027	129
40	Tyr	0.015	214

rat and human liver (Takio *et al.*, 1983; Ritonja *et al.*, 1985). It should also be noted that, providing the mouse and human cathepsin L do not differ extensively in the unsequenced region of the human chain, the determined sequence represents approx. 37% of the human enzyme.

The human cathepsin L and the proposed mouse homologue display one interesting difference: the human enzyme does not contain the cysteine residues present in positions 12 and 33 of the mouse protein. Cysteine residues are also present at the equivalent positions in cathepsin B and may form a disulphide bridge (Takio *et al.*, 1983).

C-Terminal processing has been observed for a number of lysosomal enzymes, including cathepsin D (Erickson & Blobel, 1983; Faust *et al.*, 1985) and cathepsin B (San Segundo *et al.*, 1985). Comparison of

the human cathepsin L protein sequences with the mouse cDNA sequence suggests that C-terminal processing of this enzyme is limited at most to a single residue, although the exact extent of C-terminal processing requires the cDNA sequence for the human enzyme and confirmation of the C-terminal amino acid.

A major excreted protein from a transformed mouse fibroblast cell line has been identified as an activatable cysteine proteinase that has a similar specificity of action on the oxidized β -chain of insulin to that of rat cathepsin L (Gal & Gottesman, 1986a,b). This is probably the same

Table 2. N-Terminal sequence of human cathepsin L light chain

About 10 nmol of protein were employed. Although substantial radioactivity was detected in some cycles and, in particular, cycle 4, no peak corresponding to S-carboxymethylcysteine (CmC) was evident in h.p.l.c. traces. However, a peak for S-carboxymethylcysteine was present in h.p.l.c. analysis of cycle 31, and so this residue was assigned as S-carboxymethylcysteine.

Cycle	Identity	Yield (nmol)	Radioactivity released (c.p.m.)
1	Asn	7.31	403
2	Asn	5.15	450
3	Lys	8.12	288
4	Tyr	7.16	851
5	Trp	5.86	718
6	Leu	6.71	697
7	Val	5.79	645
8	Lys	6.49	442
9	Asn	4.93	463
10	Ser	3.56	360
11	Trp	1.85	300
12	Gly	2.95	312
13	Glu	2.06	191
14	Glu	1.97	400
15	Trp	0.98	242
16	Gly	1.70	166
17	Met	2.63	303
18	Gly	1.74	151
19	Gly	1.91	127
20	Tyr	1.91	130
31	Val	1.47	121
22	Lys	1.72	85
23	Met	2.00	182
24	Ala	1.85	145
25	Lys	1.241	157
26	Asp	0.843	188
27	Arg	1.106	115
28	Arg	1.476	69
29	Asn	1.010	127
30	His	0.305	179
31	CmC	0.508	1000
32	Gly	0.717	475
33	Ile	0.586	282
34	Ala	0.597	166
35	Ser	0.380	79
36	Ala	0.597	166
37	Ala	0.570	63
38	Ser	0.133	79
39	Tyr	0.154	97
40	Pro	0.087	157
41	Thr	0.184	106
42	Val	0.094	100

Fig. 2. Homology between human cathepsin L and other members of the cysteine proteinase superfamily

The sequences are numbered according to the mouse cDNA sequence (Portnoy *et al.*, 1986). Gaps are indicated by '-' and boxed residues are identical with those of human cathepsin L. '?' indicates unidentified cathepsins. Sequence data were obtained from San Segundo *et al.* (1985; human cathepsin B), Takio *et al.* (1983; rat cathepsins B and H), Takahashi *et al.* (1984; pig cathepsins B and H), Pohl *et al.* (1982; ox cathepsin B), Turk *et al.* (1983; ox cathepsin S) Portnoy *et al.* (1986; unknown mouse cathepsin) and Gay & Walker (1985; unknown ox cathepsin).

enzyme as that expressed in mouse macrophages which has been identified here as cathepsin L [see Gal & Gottesman (1986b) for comparisons]. The identification of secreted forms of cathepsin L suggests an extracellular function for this powerful proteolytic enzyme, which is capable of hydrolysing both collagen and elastin (Kirschke *et al.*, 1982; Mason *et al.*, 1986).

Relationship of cathepsin L to other cathepsins

The human cathepsin L is more distantly related to the sequences of human cathepsin B (45% identity) and rat

cathepsin H (57% identity) and a number of gaps have to be inserted to enable these proteins to be aligned (Fig. 2).

A close relationship (71% identity) is found between human cathepsin L and the ox cathepsin S. However, many of the *N*-terminal 30 amino acids are conserved in all cysteine proteinases and, for example, human cathepsin L is identical in 70% of its amino acids with rat cathepsin H over this region. From residues 31 onwards none of the amino acids in the partial sequence of ox cathepsin S and only four out of ten of those in cathepsin

H are identical with those of human cathepsin L. Cathepsins L and S can be differentiated by catalytic criteria, as they have different specific activities with Z-Phe-Arg-NHMec, and cathepsin S is inhibited by Z-Phe-Phe-CHN₂ more slowly than cathepsin L (Kirschke *et al.*, 1984). However, they have similar activity against the protein substrate azocasein. Furthermore, cathepsin S has only been purified from ox spleen and lymph nodes, tissues that do not seem to contain a distinct form of cathepsin L (Turk *et al.*, 1980, 1983). The sequence data presented here supports the suggestion that ox cathepsin S is not a species variant of cathepsin L, but is a different enzyme.

The L-chain of human cathepsin L is rather distantly related to the sequence derived from an ox cDNA clone (< 50% identity; Gay & Walker, 1985), suggesting strongly that the enzyme encoded by the clone is not a cathepsin L. Its weak relationship to cathepsin B and H led to the earlier proposal that it is most unlikely to be cathepsin B or H (Gay & Walker, 1985). Thus its identity remains unknown.

We thank Dr D. A. Portnoy for providing sequence data before their publication and for a preprint, and Dr. S. Gal and Dr. M. M. Gottesman for providing a preprint of their paper. We are grateful to our colleagues for valuable discussions.

REFERENCES

- Barrett, A. J. & Kirschke, H. (1981) *Methods Enzymol.* **80**, 535–561
- Barrett, A. J., Nicklin, M. J. H. & Rawlings, N. D. (1984) *Symp. Biol. Hung.* **25**, 203–217
- Bury, A. F. (1981) *J. Chromatogr.* **213**, 419–500
- Dayhoff, M. O., Barker, W. C. & Hunt, L. T. (1983) *Methods Enzymol.* **91**, 524–545
- Downward, J., Yarden, Y., Mayes, E., Scrace, G., Totty, N., Stockwell, P., Ullrich, A., Schlessinger, J. & Waterfield, M. D. (1984) *Nature (London)* **307**, 521–527
- Erickson, A. H. & Blobel, G. (1983) *Biochemistry* **22**, 5201–5205
- Faust, P. L., Kornfeld, S. & Chirgwin, J. M. (1985) *Proc. Natl. Acad. Sci. U.S.A.* **82**, 4910–4914
- Gal, S. & Gottesman, M. M. (1986a) *J. Biol. Chem.* **261**, 1760–1766
- Gal, S. & Gottesman, M. M. (1986b) *Biochem. Biophys. Res. Commun.* **139**, 156–162
- Gay, N. J. & Walker, J. E. (1985) *Biochem. J.* **225**, 707–712
- Kirschke, H., Langner, J., Wiederanders, B., Ansorge, S. & Bohley, P. (1977) *Eur. J. Biochem.* **74**, 293–301
- Kirschke, H., Kumbhavi, A. A., Bohley, P. & Barrett, A. J. (1982) *Biochem. J.* **201**, 367–372
- Kirschke, H., Locnikar, P. & Turk, V. (1984) *FEBS Lett.* **174**, 123–127
- Mason, R. W., Taylor, M. A. J. & Etherington, D. J. (1984) *Biochem. J.* **217**, 209–217
- Mason, R. W., Green, G. D. J. & Barrett, A. J. (1985) *Biochem. J.* **226**, 233–241
- Mason, R. W., Johnson, D. A., Barrett, A. J. & Chapman, H. A. (1986) *Biochem. J.* **233**, 925–927
- Murachi, T. (1984) *Biochem. Soc. Symp.* **49**, 149–167
- Ohno, S., Emori, Y., Imajoh, S., Kawasaki, H., Kasaragi, M. & Suzuki, K. (1984) *Nature (London)* **312**, 566–570
- Pohl, J., Baudys, M., Tomasek, V. & Kostka, V. (1982) *FEBS Lett.* **142**, 23–26
- Portnoy, D. A., Erickson, A. H., Kochan, J., Ravetch, J. V. & Unkeless, J. C. (1986) *J. Biol. Chem.*, in the press
- Ritonja, A., Popovic, T., Turk, V., Wiedenmann, K. & Machleidt, W. (1985) *FEBS Lett.* **181**, 169–172
- San Segundo, B., Chan, S. J. & Steiner, D. F. (1985) *Proc. Natl. Acad. Sci. U.S.A.* **82**, 2320–2324
- Takahashi, T., Dehdarani, A. H., Schmidt, P. G. & Tang, J. (1984) *J. Biol. Chem.* **259**, 9874–9882
- Takio, K., Towatari, T., Katunuma, N., Teller, D. C., Titani, K. (1983) *Proc. Natl. Acad. Sci. U.S.A.* **80**, 3666–3670
- Turk, V., Kregar, I., Popovic, T., Locnikar, P., Kopitar, M. & Brzin, J. (1980) *Period. Biol.* **82**, 363–368
- Turk, V., Brzin, J., Kopitar, M., Kregar, M., Locnikar, P., Longer, M., Popovic, T., Ritonja, A., Vitale, L., Machleidt, W., Giraldi, T. & Sava, G. (1983) in *Proteinase Inhibitors* (Katunuma, N., Umezawa, H. & Holzer, H., eds.), pp. 125–134, Japan Scientific Societies Press, Tokyo

Received 6 May 1986/4 August 1986; accepted 4 September 1986