

Homology of amino acid sequences of rat liver cathepsins B and H with that of papain

(thiol endopeptidase/molecular evolution/lysosomal enzymes)

KOJI TAKIO*, TAKAE TOWATARI†, NOBUHIKO KATUNUMA‡, DAVID C. TELLER‡, AND KOITI TITANI*

*Howard Hughes Medical Institute Laboratory and †Department of Biochemistry, SJ-70, University of Washington, Seattle, Washington 98195; and ‡Department of Enzyme Chemistry, Institute for Enzyme Research, Tokushima University, Tokushima 770, Japan

Communicated by Hans Neurath, March 21, 1983

ABSTRACT The amino acid sequences of rat liver lysosomal thiol endopeptidases, cathepsins B and H, are presented and compared with that of the plant thiol protease papain. The 252-residue sequence of cathepsin B and the 220-residue sequence of cathepsin H were determined largely by automated Edman degradation of their intact polypeptide chains and of the two chains of each enzyme generated by limited proteolysis. Subfragments of the chains were produced by enzymatic digestion and by chemical cleavage of methionyl and tryptophanyl bonds. Comparison of the amino acid sequences of cathepsins B and H with each other and with that of papain demonstrates a striking homology among their primary structures. Sequence identity is extremely high in regions which, according to the three-dimensional structure of papain, constitute the catalytic site. The results not only reveal the first structural features of mammalian thiol endopeptidases but also provide insight into the evolutionary relationships among plant and mammalian thiol proteases.

Cathepsin B was originally the name coined for the enzyme in bovine spleen that hydrolyzes benzoylarginine amide in the presence of cysteine. For a long time, it had been thought to be the only lysosomal thiol endopeptidase (1, 2). However, Otto *et al.* (3–5) have shown that more than one enzyme displays this activity. They isolated a major thiol endopeptidase from rat liver lysosomes with a M_r of 25,000 and renamed it cathepsin B₁. More extensive studies of thiol endopeptidases of rat liver lysosomes by Kirschke *et al.* (6–8) have led to the identification of two thiol proteases, cathepsins H and L. At the same time, Towatari *et al.* (9, 10) independently isolated a new thiol endopeptidase from cathepsin B₁ preparations, which was later established to be identical to cathepsin L of Kirschke *et al.* Cathepsin B was crystallized by Towatari *et al.* as spindle-form crystals (11). More recently, several workers have reported the presence of additional thiol endopeptidases in lysosomes—e.g., cathepsins T, N, P, and S (12–15)—but their amounts appear to be very low. These lysosomal thiol endopeptidases are considered to play an important role in intracellular protein degradation and possibly in post-translational processing of some biologically important substances such as endorphin (16) and insulin (14, 17).

We have recently isolated three major proteases from rat liver lysosomes, cathepsins B, H, and L, and purified them to a homogeneous state in terms of enzymatic activities, substrate specificities, and antigenic specificities (10, 11, 18). Crystalline cathepsin B is a mixture of a single-chain and a two-chain form of the enzyme (19). It is not known yet whether the proteolytic cleavage takes place before or after incorporation into lysosome or during its isolation. Some cathepsin B preparations contain

only the single-chain form of the enzyme, suggestive of artefactual limited proteolysis. The protease responsible for this limited proteolysis has not been identified. We also reported that the amino-terminal sequence of cathepsin B has a striking resemblance to that of the plant thiol endopeptidase papain (19).

We now present the complete amino acid sequences of rat liver cathepsins B and H and show that these two mammalian thiol proteases are homologous both to each other and to the plant thiol endopeptidases papain and actinidin, suggesting that all four enzymes have evolved from a common ancestral protein. This homology allows us to predict that the polypeptide chain folding of the mammalian enzymes will be found to be similar to that of papain (20) and actinidin (21).

The experimental details of the sequence analysis of cathepsins B and H will be published elsewhere.

MATERIALS AND METHODS

Crystalline rat liver cathepsin B was prepared as described (11). Cathepsin H was prepared by the method of Schwartz and Barrett (22). The intact chain and the two polypeptide chains of each enzyme, generated by proteolytic cleavage, were separated on a Sephacryl S-200 column after S-carboxymethylation of reduced disulfide bonds as reported (19). Trypsin treated with L-1-*p*-tosylamino-2-phenylethyl chloromethyl ketone (TPCK-trypsin) and *Staphylococcus aureus* V8 protease were obtained from Worthington and Miles, respectively. Gel filtration media were products of Pharmacia.

Chemical cleavages by cyanogen bromide (Eastman) and by 2-(2-nitrophenylsulphenyl)-3-methyl-3'-bromoindolenine (BNPS-skatole) (Pierce) were carried out by methods of Gross and Witkop (23) and of Omenn *et al.* (24), respectively.

Peptides were purified by gel filtration or by reversed-phase HPLC, or both, which was carried out with a Varian 5000 liquid chromatograph on columns of μ Bondapak C₁₈ or CN (Waters Associates) or SynChropak RP-P (SynChrom) by using a trifluoroacetic acid/acetonitrile elution system (25, 26).

Amino acid compositions were analyzed with a Dionex D500 amino acid analyzer. Edman degradations were performed with a Beckman Sequencer 890C according to Edman and Begg (27) by using a program adapted from Brauer *et al.* (28) in the presence of Polybrene (29). Phenylthiohydantoin derivatives were identified in a semiquantitative manner by two HPLC systems (30, 31).

A search for a homologous sequence was carried out with a VAX/VMS computer by using a "protein sequence database" of the *Atlas of Protein Sequence and Structure* (version 4, April 30, 1982) obtained from the National Biomedical Research

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviations: TPCK-trypsin, trypsin treated with L-1-*p*-tosylamino-2-phenylethyl chloromethyl ketone; BNPS-skatole, 2-(2-nitrophenylsulphenyl)-3-methyl-3'-bromoindolenine.

Foundation. The alignment of sequences was determined with a PDP-12 and a VAX/VMS computer by applying the sequence comparison program of de Haën *et al.* (32) along with scoring matrices of McLachlan (33) and Dayhoff *et al.* (34).

RESULTS AND DISCUSSION

Cathepsin H was isolated as an approximately 50:50 mixture of the single- and two-chain forms of the enzyme, as was crystalline cathepsin B. We took advantage of the limited proteolysis that occurred with both enzymes to determine their amino acid sequences. The heavy and light chains of each enzyme (and the intact chain also in the case of cathepsin H) were separated after cleavage of disulfide bonds that still held the two chains together in the native forms. The intact chain of cathepsin B was not completely separated from the heavy chain by gel filtration. The separated chains were then subjected to automated sequence analysis.

Analysis of the light chain of cathepsin B has been reported (19). From the cyanogen bromide digest of the heavy chain of cathepsin B, four major fragments (residues 48–129, 130–159, 160–194, and 195–252) and two minor overlapping fragments (residues 48–159 and 160–252) were isolated (Fig. 1). Cleavage with TPCK-trypsin after citraconylation (35) yielded five major fragments (residues 48–83, 84–200, 201–233, 234–250, and 251–252) and two minor fragments (residues 84–149 and 150–200), resulting from abnormal cleavage of a Tyr-Ser bond. To complete the sequence, these fragments were further digested with staphylococcal protease or cleaved with BNPS-skatole.

The amino acid sequence of cathepsin H was similarly determined from digests of both chains with cyanogen bromide,

trypsin, and staphylococcal protease. BNPS-skatole was also effectively utilized to prepare subfragments. In most cases, digests were separated by reversed-phase HPLC and then subjected to automated sequence analysis on a semimicro scale (1–10 nmol). These techniques have greatly facilitated sequence analysis of the two enzymes, which could only be obtained in rather small quantities (6 mg/kg of rat liver).

The 252-residue sequence of cathepsin B and the 220-residue sequence of cathepsin H are shown in Figs. 2 and 3, respectively. Though the site of limited proteolysis is near the amino terminus in cathepsin B, it is near the carboxyl terminus in cathepsin H. Both sites are found to involve an asparaginyl bond. Both enzymes are glycoproteins (11, 22), but cathepsin B contains less carbohydrate than the other cathepsins and does not bind to a concanavalin A column (22). With each enzyme, it is likely that the entire carbohydrate moiety is linked to a single asparaginyl residue: residue 111 in cathepsin B and residue 115 in cathepsin H. In the analysis of cathepsin B, only residue 111 failed to yield an identifiable phenylthiohydantoin derivative, although amino acid analysis of a small peptide (residues 108–120) generated by staphylococcal protease yielded 1 mol of aspartic acid per mol of peptide. Similarly, no phenylthiohydantoin derivative could be identified at residue 115 of cathepsin H, even though a small peptide (residues 112–121) clearly contained one extra aspartic acid residue by composition. However, due to the lack of direct evidence these glycosylation sites must be considered tentative and need to be confirmed. In the absence of direct evidence for overlaps between heavy and light chains (between residues 47 and 48 in cathepsin B and between residues 177 and 178 in cathepsin H), the sites of connections of the two chains must also be regarded as tentative. The M_r s

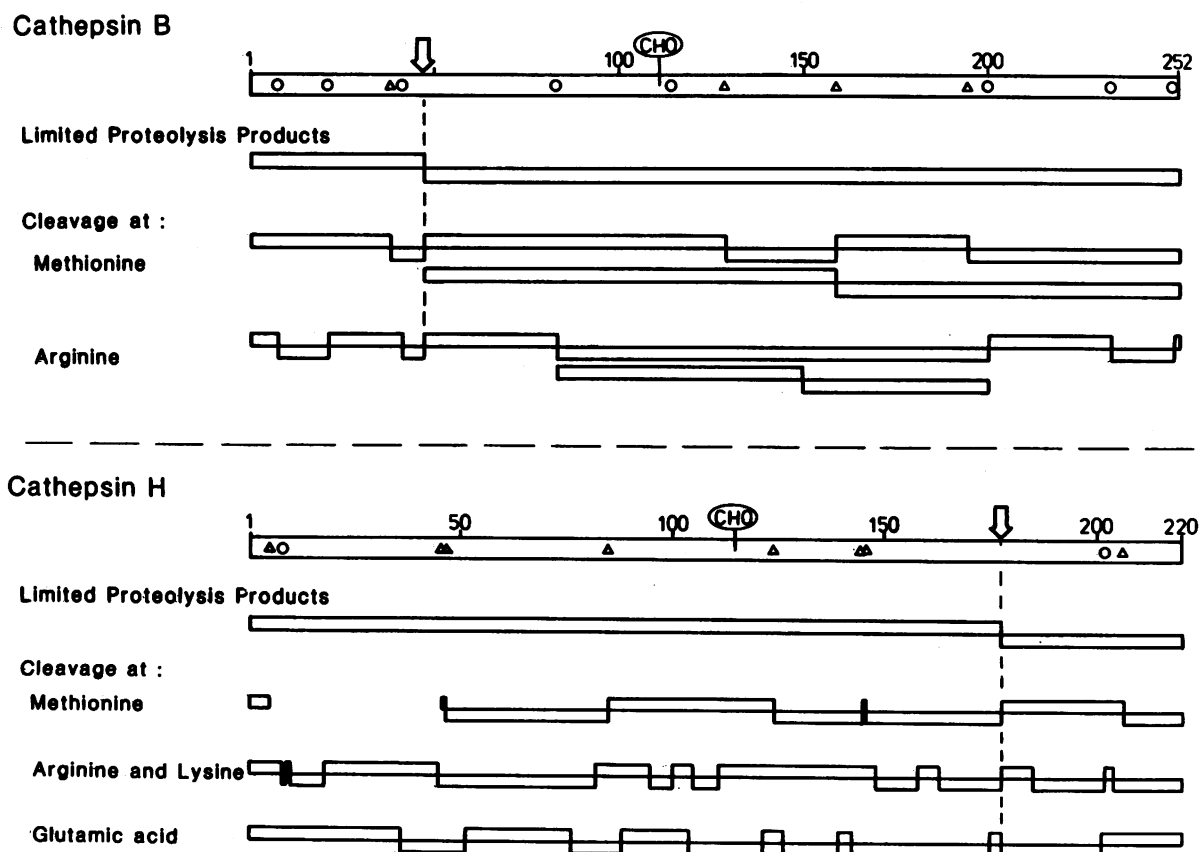


FIG. 1. Diagrams indicating the origins of the primary and secondary fragments of cathepsins B and H and their relationship to each other. The length of each bar is proportional to the length of the fragment. The top bars represent the single chain enzymes and indicate the location of arginine (○) and methionine (Δ) residues. Also indicated are the sites of glycosyl attachment (circled CHO) and of limited proteolysis (arrows).

10 20 30
 L P E S F D A R E Q W S N C P T I A Q I R D Q G S C G S C W
 40 50 60
 A F G A V E A M S D R I C I H T M V N V E V S A E D L L T C
 70 80 90
 C G I Q C G D G C N G G Y P S G A G N F W T R K G L V S G G
 100 110 120
 V Y N S H I G C L P Y T I P P C E H H V N G S R P P C T G E
 130 140 150
 G D T P K C N K M C E A G Y S T S Y K E D K H Y G Y T S Y S
 160 170 180
 V S D S E K E I M A E I Y K N G P V E G A F T V F S D F L T
 190 200 210
 Y K S G V Y K H E A G D V M G G H A I R I L G W G I E N G V
 220 230 240
 P Y W L V A N S W N V D W G D N G F F K I L R G E N H C G I
 250
 E S E I V A G I P R T Q

COMPOSITION

Ala A	14	Gln Q	5	Leu L	9	Ser S	20
Arg R	8	Glu E	18	Lys K	10	Thr T	12
Asn N	14	Gly G	33	Met M	4	Trp W	7
Asp D	12	His H	8	Phe F	8	Tyr Y	12
CySH C	14	Ile I	16	Pro P	12	Val V	16

FIG. 2. Complete amino acid sequence of rat liver cathepsin B. Circled letters C and H indicate by analogy with papain the active site cysteine and histidine residues, respectively. Boxed letter N indicates the site of glycosyl attachment and an arrow, the site of limited proteolysis. Molecular weights: without carbohydrate, M_r 27,411; including carbohydrate, M_r 29,000 (by NaDodSO₄/polyacrylamide gel electrophoresis).

of cathepsins B and H, omitting glycosyl groups, calculated from these sequences are 27,411 and 24,000, respectively.

In Fig. 4, the sequences of cathepsins B and H are compared with each other and also with that of papain (36, 37), which has already been shown to be homologous to that of another plant thiol proteinase, actinidin (38). The sequences are aligned so as to achieve maximal similarity by applying the comparison program of de Haën *et al.* (32).

It is obvious that the three proteins have a relatively high degree of identity in the amino- and carboxyl-terminal regions but rather less in the central region. The comparison of the three proteins in Fig. 4 is thus facilitated by arbitrarily dividing the sequences into three regions—an amino-terminal (or active site cysteinyl) region, a central, and a carboxyl-terminal (or active site histidyl) region. The amino-terminal region comprises 77 residues of the single-chain cathepsin B. In this region, the three proteins are aligned with only one insertion and two deletions to show 31–48% identity (Table 1). Significantly, an almost identical 11-residue sequence is found in the three enzymes in the vicinity of the active site cysteinyl residues (residue 29 in cathepsin B). It is of interest that, in this region, homology between cathepsin B and H is the lowest (31% identity), whereas it is highest when cathepsin H is compared to the plant enzyme papain (48% identity).

The sequences in the central region (residues 78–152 in cathepsin B) are more difficult to align because identity is low (<25% identity), in spite of rather liberal placement of gaps. In fact, the alignment in Fig. 4 in this region is statistically significant only by applying the mutation data matrix of Dayhoff *et al.* (34) and any alignment between cathepsin B and the other

10 20 30
 Y P S S M D W R K K G N V V S P V K N Q G A C G S C W T F S
 40 50 60
 T T G A L E S A V A I A S G K M M T L A E Q Q L V D C A Q N
 70 80 90
 F N N H G C Q G G L P S Q A F E Y I L Y N K G I M G E D S Y
 100 110 120
 P Y I G K N G Q C K F N P E K A V A F V K N V V N I T L N D
 130 140 150
 E A A M V E A V A L Y N P V S F A F E V T E D F M M Y K S G
 160 170 180
 V Y S S N S C H K T P D K V N H A V L A V G Y G E Q N G L L
 190 200 210
 Y W I V K N S W G S N W G N N G Y F L I E R G K N M C G L A
 220
 A C A S Y P I P Q V

COMPOSITION

Ala A	20	Gln Q	9	Leu L	12	Ser S	17
Arg R	2	Glu E	11	Lys K	14	Thr T	7
Asn N	20	Gly G	21	Met M	8	Trp W	5
Asp D	6	His H	3	Phe F	9	Tyr Y	12
CySH C	8	Ile I	8	Pro P	9	Val V	19

FIG. 3. Complete amino acid sequence of rat liver cathepsin H. Circled letters C and H indicate by analogy with papain the active site cysteine and histidine residues, respectively. Boxed letter N indicates the site of glycosyl attachment and an arrow, the site of limited proteolysis. Molecular weights: without carbohydrate, M_r 24,000; including carbohydrate, M_r 26,000 (by NaDodSO₄/polyacrylamide gel electrophoresis).

three thiol proteinases in this region is not statistically significant when the McLachlan scoring matrix (33) is applied. Perhaps a single large (*ca.* 30 residues) insertion resulted from incorporation of an intron into an exon during the long process of divergent evolution. The functional significance of the central region is not clear, even in papain, where the three-dimensional structure is established.

The degree of identity among the three sequences is very high in the carboxyl-terminal region (residues 153–252 in cathepsin B). Only in this region cathepsin H shows comparable identity to cathepsin B (40% identity) on the one hand and to papain (43%) on the other. In contrast to the regions surrounding the active site cysteinyl residues, the sequences around the active site histidyl residues (residue 197 in cathepsin B) have a lower degree of identity although homology is evident, particularly between cathepsin H and papain. Cathepsins B and H also share with papain and actinidin the three-residue sequence Asn-Ser-Trp (residues 217–219 in cathepsin B). According to the x-ray structure of papain, asparagine is hydrogen-bonded to the active site histidine and tryptophan acts to stabilize that bond (20). Glycine-205 in cathepsin B is found in cathepsin H and also in actinidin in the region corresponding to the only discrepancy between the reported chemical sequence (Asn-Pro-Gly) and the x-ray sequence (Gly-Pro-Asn) (20). Overall, the sequence of rat liver cathepsin H is more closely related to those of the plant enzymes papain and actinidin than to that of rat liver cathepsin B, suggesting that cathepsin B diverged from the common ancestral gene long before cathepsin H.

The amino acid sequences of cathepsins B and H were fitted to the known three-dimensional structure of papain (20) by using the alignments shown in Fig. 4. Almost all of the insertions and deletions were found to be in surface loops without af-

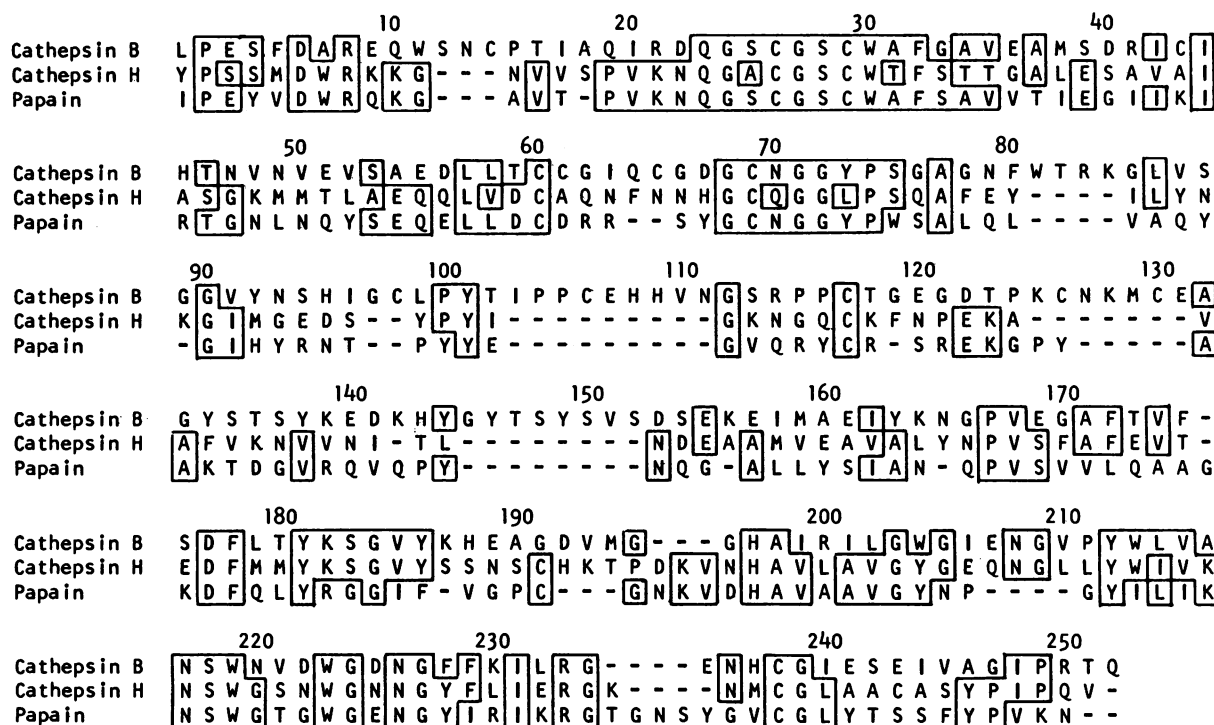


FIG. 4. Comparison of the amino acid sequences of cathepsins B and H and papain. Gaps are indicated by dashes. Residues common to two or more sequences are enclosed in boxes. The residue numbers are those of cathepsin B.

fecting the fundamental structure of papain. Fig. 5 compares in a schematic, simplified fashion the folding of the peptide chains, the sites of glycosyl attachment, active site residues, the sites of limited proteolysis, the location of the cysteinyl residues, and the assumed disulfide bonds of cathepsins B and H. It is impossible to draw a meaningful conformation for the central region of the cathepsin B from the present results because of the rather large and somewhat arbitrary insertions and the very limited homology to papain. Prediction of secondary structures by the method of Chou and Fasman (39) suggested that the central region of cathepsin B has rather extended structures. For this reason, the pairing of the disulfide bonds in this region is highly conjectural.

There is no evidence even that lysosomal cathepsins contain disulfide bonds. Most intracellular proteins contain free SH groups in the form of cysteinyl residues, but the intralysosomal environment may be more oxidative and favor the formation of disulfide bonds. Fittings of the cathepsin sequences to papain suggest that most of the cysteinyl residue pairs would be adjacent to each other and could form disulfide bonds without changing the native conformations of the proteins. Alternatively, if within the lysosome all the cysteinyl residues existed in the reduced form, they could have become oxidized to disulfide bonds during the isolation procedures. Pohl *et al.* (40) recently reported a disulfide bond in the light chain of bovine spleen cathepsin B although compelling evidence was not pre-

Table 1. Identity of the amino acid sequences of thiol endopeptidases

Region of comparison	Amino-terminal	Central	Carboxyl-terminal	Whole protein
Residues in cathepsin B	1-77	78-152	153-252	1-252
Cathepsin B/cathepsin H	31.1	13.6	39.8	31.5
Cathepsin B/papain	43.7	13.0	28.4	30.2
Cathepsin H/papain	47.9	21.4	42.9	40.2

Identity (%) was calculated from the sequence alignment of Fig. 4.

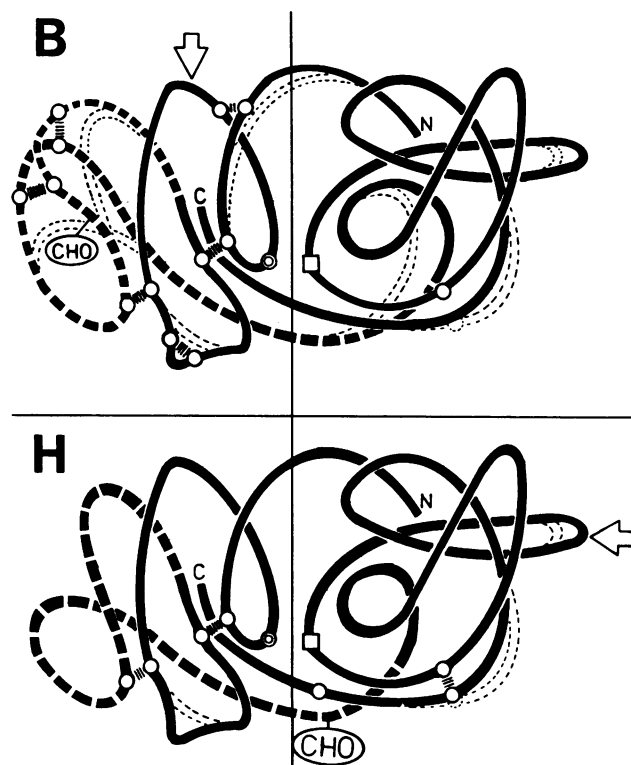


FIG. 5. Schematic comparison of the hypothetical polypeptide chain foldings of cathepsins B and H with that of papain (double broken line). Letters N and C indicate the amino and carboxyl termini, respectively. Circles, double circles, and squares indicate cysteinyl, active site cysteinyl, and active site histidyl residues, respectively. A circled CHO indicates the site of attachment of the glycosyl moiety. Arrows denote the sites of limited proteolysis. Hypothetical disulfide bonds are indicated by broken lines between adjacent circles. The central regions are indicated by bold broken lines.

sented. Their result agrees with one of our hypothetical disulfide bonds in rat liver cathepsin B (Cys-14-Cys-43). Of the 14 cysteinyl residues, 2 must exist as free SH groups, 1 at the active site and the other probably at the surface of the right lobe, which corresponds to 1 involved in disulfide pairing in papain. Also in cathepsin H, there must be two free SH groups in addition to the three conserved disulfide bonds that correspond to those of papain and actinidin. It is not clear yet whether free SH groups in regions other than the active site contribute to the mode of action of the enzymes.

The glycosylated residues appear to be located on a rather loose surface loop in the left lobe (cathepsin B) and on an extended structure connecting the right and left lobes (cathepsin H). The sites of limited proteolysis are also found at surface loops far from the catalytic sites; hence, the impact of limited proteolysis on catalytic activity may be minimal or absent. It is not known yet whether limited proteolysis is an endogenous process or whether it involves extraneous enzymes.

Cathepsins B and H not only act as endopeptidases but also as a peptidyl dipeptidase (41) and an amino peptidase (42), respectively. The positively charged guanidino group of arginine-200 in cathepsin B may conceivably be the carboxyl binding site for the peptidyl dipeptidase activity. This residue is separated by two residues from the active site histidine-197 and the corresponding alanyl residue in papain extends its side chain into the active site groove.

We thank Dr. Kenneth A. Walsh and Dr. Hans Neurath for their encouragement and valuable discussions. We are also indebted to Mrs. Santosh Kumar, Mr. Richard R. Granberg, Mr. Roger D. Wade, Mr. Lowell H. Ericsson, and Mr. Brad McMullen for their skillful assistance. This work was supported in part by grants from the National Institutes of Health (GM 15731 and GM 13401) and from the Ministry of Education, Science and Culture of Japan (Scientific Research Grant 56880028).

1. Tollan, H. H., Jones, M. E. & Fruton, J. S. (1952) *J. Biol. Chem.* **194**, 793-805.
2. Greenbaum, L. M. & Fruton, J. S. (1975) *J. Biol. Chem.* **226**, 173-180.
3. Otto, K. (1967) *Hoppe-Seyler's Z. Physiol. Chem.* **348**, 1449-1460.
4. Otto, K. & Bhakadi, S. (1969) *Hoppe-Seyler's Z. Physiol. Chem.* **350**, 1577-1588.
5. Otto, K. (1971) in *Tissue Proteinases*, eds. Barrett, A. L. & Dingle, J. T. (North-Holland, Amsterdam), pp. 1-28.
6. Kirschke, H., Langner, J., Wiederanders, B., Ansoerge, S., Bohley, P. & Hanson, H. (1976) *Acta Biol. Med. Ger.* **35**, 285-299.
7. Kirschke, H., Langner, J., Wiederanders, B., Ansoerge, S. & Bohley, P. (1977) *Eur. J. Biochem.* **74**, 293-301.
8. Kirschke, H., Langner, J., Wiederanders, B., Ansoerge, S., Bohley, P. & Hanson, H. (1977) *Acta Biol. Med. Ger.* **36**, 185-199.
9. Towatari, T., Tanaka, K., Yoshikawa, D. & Katunuma, N. (1976) *FEBS Lett.* **67**, 284-288.
10. Towatari, T., Tanaka, K., Yoshikawa, D. & Katunuma, N. (1978) *J. Biochem. (Tokyo)* **84**, 659-671.
11. Towatari, T., Kawabata, Y. & Katunuma, N. (1979) *Eur. J. Biochem.* **102**, 279-289.
12. Gohda, E. & Pitot, H. C. (1981) *J. Biol. Chem.* **256**, 2567-2572.
13. Evans, P. & Etherington, D. J. (1978) *Eur. J. Biochem.* **83**, 87-97.
14. Docherty, K., Carroll, R. J. & Steiner, D. F. (1982) *Proc. Natl. Acad. Sci. USA* **79**, 4613-4617.
15. Turnšek, T., Kregar, I. & Lebez, D. (1975) *Biochim. Biophys. Acta* **403**, 514-520.
16. Suhr, A. & Marks, N. (1979) *Eur. J. Biochem.* **101**, 23-30.
17. Smith, R. E. & Van Frank, R. M. (1975) in *Lysosomes in Biology and Pathology*, eds. Dingle, J. T. & Dean, R. T. (North-Holland, Amsterdam), pp. 193-249.
18. Katunuma, N., Towatari, T., Kominami, E., Hashida, S., Takio, K. & Titani, K. (1982) *Acta Biol. Med. Ger.* **40**, 1419-1425.
19. Takio, K., Towatari, T., Katunuma, N. & Titani, K. (1981) *Biochem. Biophys. Res. Commun.* **97**, 340-346.
20. Drenth, J., Jansonius, J. N., Koekoek, R. & Walthers, B. G. (1971) *Adv. Protein Chem.* **25**, 79-115.
21. Baker, E. N. (1977) *J. Mol. Biol.* **115**, 263-277.
22. Schwartz, W. N. & Barrett, A. J. (1980) *Biochem. J.* **191**, 487-497.
23. Gross, E. & Witkop, B. (1962) *J. Biol. Chem.* **237**, 1856-1860.
24. Ommen, G. S., Fontana, A. & Anfinsen, C. B. (1970) *J. Biol. Chem.* **245**, 1895-1902.
25. Mahoney, W. C. & Hermodson, M. A. (1980) *J. Biol. Chem.* **255**, 11199-11203.
26. Dunlap, C. E., III, Gentleman, S. & Lowney, L. I. (1978) *J. Chromatogr.* **160**, 191-198.
27. Edman, P. & Begg, G. (1967) *Eur. J. Biochem.* **1**, 80-91.
28. Brauer, A. W., Margolies, M. N. & Haber, E. (1975) *Biochemistry* **14**, 3029-3035.
29. Tarr, G. E., Beecher, J. F., Bell, M. & McKean, D. J. (1978) *Anal. Biochem.* **84**, 622-627.
30. Bridgen, P. J., Cross, G. A. M. & Bridgen, J. (1976) *Nature (London)* **263**, 613-614.
31. Ericsson, L. H., Wade, R. D., Gagnon, J., McDonald, R. M., Granberg, R. R. & Walsh, K. A. (1977) in *Solid Phase Methods in Protein Sequence Analysis*, eds. Previero, A. & Coletti-Previero, M. A. (Elsevier/North-Holland, Amsterdam), pp. 137-142.
32. de Haën, C., Swanson, E. & Teller, D. C. (1976) *J. Mol. Biol.* **106**, 639-661.
33. McLachlan, A. D. (1971) *J. Mol. Biol.* **61**, 409-424.
34. Dayhoff, M. O., Schwartz, R. M. & Orcutt, B. C. (1979) in *Atlas of Protein Sequence and Structure*, ed. Dayhoff, M. O. (Natl. Biomed. Res. Found., Silver Spring, MD), Vol. 5, Suppl. 3, pp. 345-352.
35. Habeeb, A. F. S. A. & Atassi, M. Z. (1970) *Biochemistry* **9**, 4939-4944.
36. Mitchel, R. E. J., Chaiken, I. M. & Smith, E. L. (1970) *J. Biol. Chem.* **245**, 3485-3492.
37. Husain, S. S. & Lowe, G. (1970) *Biochem. J.* **116**, 689-692.
38. Carne, A. & Moore, C. H. (1978) *Biochem. J.* **173**, 73-83.
39. Chou, P. Y. & Fasman, G. D. (1978) *Adv. Enzymol.* **47**, 45-148.
40. Pohl, J., Baudyš, M., Tomášek, V. & Kostka, V. (1982) *FEBS Lett.* **142**, 23-26.
41. Aronson, N. N. & Barrett, A. J. (1978) *Biochem. J.* **171**, 759-765.
42. Singh, H. & Kalnitsky, G. (1980) *J. Biol. Chem.* **255**, 369-374.