Intracellular Proteolysis (N. Katunuma & E. Kominami, eds.) Japan Scient. Press, Tokyo, 1989; pp. 27-34.

STRUCTURAL AND FUNCTIONAL ASPECTS OF HUMAN CATHEPSINS B, H AND L AND THEIR ENDOGENOUS PROTEIN INHIBITORS STEFINS*

V. Turk, R. Jerala, B. Lenarcic and A. Sali

Department of Biochemistry
J. Stefan Institute, Jamova 39
61000 Ljubljana, Yugoslavia

INTRODUCTION

Mammalian cysteine proteinases, lysosomal cathepsins B, H and L, belong to the group of closely related proteins of papain superfamily (1, 2). They participate in the intracellular breakdown of proteins and in the control of cellular functions by limited proteolysis. Recent studies show that these proteins are synthesized as larger glycosylated precursor forms (3, 4) and processed to the native enzymes. Cysteine proteinases have been implicated in several diseases including muscular distrophy (6), tumor metastasis (7, 8) and rheumatoid diseases (5).

Proteolytic activity in cells can also be regulated by endogenous protein inhibitors of cysteine proteinases, cystatins. They may protect the cells from inappropriate endogenous or external proteolysis and/or could be involved in the control mechanism responsible for intracellular or extracellular protein breakdown. The superfamily of cystatins comprises structurally homologous proteins. They are reversible and tight-binding inhibitors of cysteine proteinases of mammalian and plant origin (2, 9). The superfamily consists of three families with homologous sequences: stefins, cystatins and kininogens (10, 11). A model for their evolution has been constructed on the basis of their sequence homology (12 - 14).

In this report we present amino acid sequences of human cathepsins B, H and L. To investigate the inhibitors further, we synthesized and cloned the gene coding for human stefin B and expressed it in E. coli. We present evidence that cathepsin D inactivates human stefin A and stefin B, thus suggesting a new regulatory role for this enzyme. A phylogeny analysis of sequenced cystatins and structurally related proteins is described as well.

---

*This work was supported by a grant from the Research Council of Slovenia.

AMINO ACID SEQUENCES OF HUMAN CATHEPSINS B, H AND L

The lysosomal cysteine proteinases cathepsins B, H and L share several common characteristics. They are present in all mammalian cell types and synthesized in a precursor form. The mature enzymes are small proteins of Mr of about 25 - 30 kDa. They are optimally active at acidic pH and unstable at neutral and alkaline pH. We reported that human cathepsins B, H and L have a high degree of amino acid sequence homology with papain and actinidin (15, 16) (Fig. 1). The amino acid residues around the active site Cys-25 and His-159 (papain numbering) are highly conserved. Cathepsin B is composed of a light chain (residues 1 - 47) and a heavy chain (residues 48 - 252), which probably result from an autocatalytic cleavage. In comparison, the cDNA derived sequence (3) predicts that the two-chain form is generated by cleavage at two sites with the loss of dipeptide Ala-His which is located between Asn-48 and Val-50 (cathepsin B numbering). We conclude that single chain cathepsin B is composed of 254 amino acid residues. Active cathepsin H is composed of 230 amino acid residues. 222 of them form a single chain (see Fig. 1) and the other eight residues form a mini-chain Glu-Pro-Gln-Asn-Cys-Ser-Ala-Thr, which is disulfide-linked to the rest of the enzyme (16). The minor fraction, the two chain form, results from the proteolytic cleavage of a single peptide bond between Asp-177 and Gly-178. In contrast to human cathepsin B, which contains six additional amino acid residues at the carboxyl terminus of its precursor (3), the mature cathepsin H has the same C-terminal sequence as its precursor (Chan, S. J. et al., personal communication). Human cathepsin L has 217 amino acid residues and also exists as a two-chain enzyme (heavy chain, residues 1 - 175 and light chain, residues 176 - 217). The resulting protein sequence of human cathepsin L is in agreement with cDNA-derived sequence of human procathepsin L (17) except for the missing peptide Glu-Ser at the C-terminus of the heavy chain.

Cathepsins B, H and L have only one free SH-group while the other cysteines form disulfide bridges. We can conclude that all three cathepsins are homologous to papain, cathepsin L being the most closely related to papain.

NEW PROPERTIES OF HUMAN STEFIN A AND STEFIN B

These two protein inhibitors are the only members of stefin family that are of human origin. They occur both intracellularly and extracellularly. Their Mr is about 11 kDa and they lack disulfide bonds and carboxyhydrates. We determined the primary structure of stefin A isolated from human

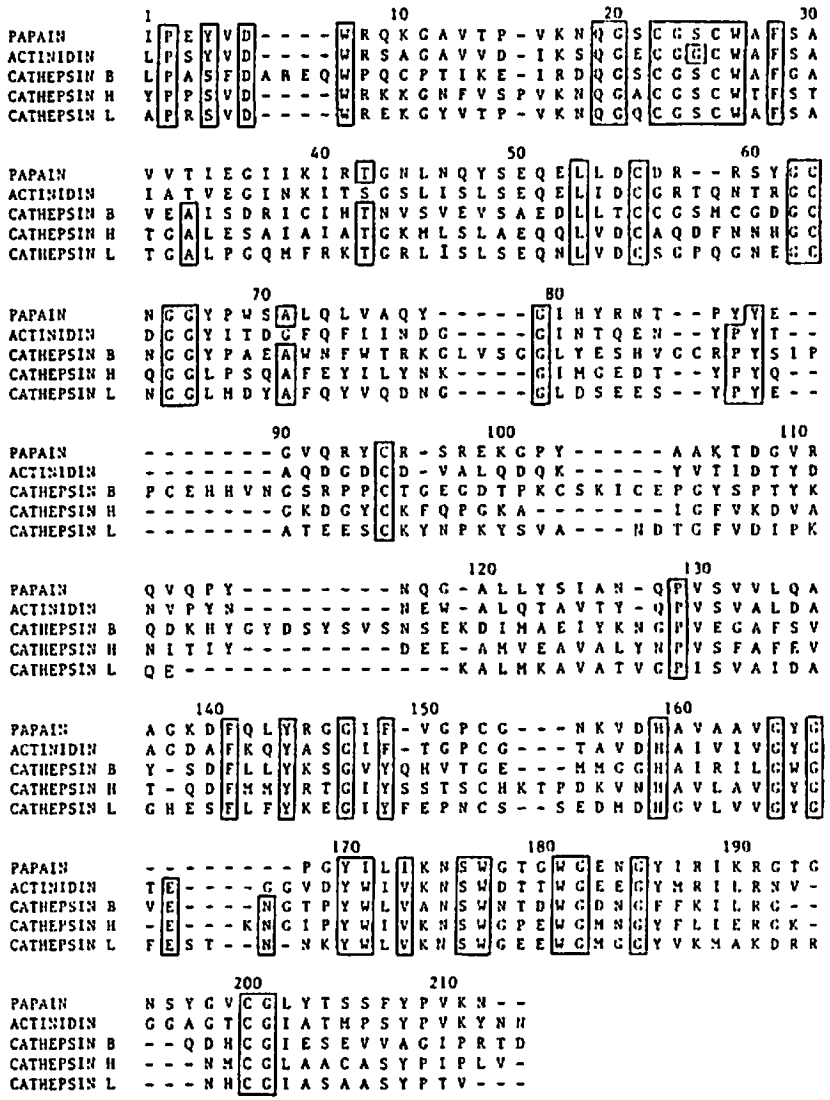

Fig. 1. Alignment of human cathepsin B (15), H and L (16) with papain and actinidin (29). Gaps are indicated by dashes. Residues common to all enzymes are boxed. Papain numbering is used.

```
a)  MIPGGLSEAKPATPEIQEIVDKVKPQLEEKTNETYGKLEAVQYKTQVVAGTNYYIK
b) acMMCGAPSATQPATAETQHIADQVRSQLEEKYNKKFPVFKAVSFKSQVVAGTNYFIK
    1         10        20        30        40        50

a)  VRAGDNKYMHLKVFKSLPGQNEDLVLTGYQVDKNKDDELTGF
b)  VHVGDEDFVHLRVFQSLPHENKPLTLSNYQTNKAKHDELTYF
      60        70        80        90      98
```

Fig. 2. The amino acid sequences of human stefin A (18) and human stefin B (19,20).

polymorphonuclear granulocytes (18) and stefin B from human liver (19, 20) (Fig. 2). Both inhibitors have an N-terminal methionine. The methionine residue probably represents the point of initiation of translation. Recently, a stretch of DNA containing the coding sequence for human stefin A was synthesized and expressed in E. coli (21). The recombinant stefin A exhibits similar biochemical properties as the native protein. In addition, a gene coding for human stefin B was synthesized and expressed (Jerala, R. et al., submitted for publication). The gene was assembled from 17 oligonucleotides (Fig. 3) and cloned in pUC8 cloning vector. The insert with the verified DNA sequence was subcloned into two expression vectors and expressed in E. coli both as a fusion protein with β-galactosidase and as a native protein. In contrast to the results reported for stefin A, stefin B could also be produced intracellularly in bacteria in high yields. Both the CNBr cleaved fusion protein and the native recombinant stefin B inhibited papain and reacted with antibodies against human stefin B.

Recently we investigated the interactions of both stefins with cathepsins H and L and papain. Incubation of these inhibitors with their target enzymes did not result in any detectable amounts of cleavage products. Therefore, the possible effect of other lysosomal proteinases on the stefins was investigated next. Surpisingly, aspartic proteinase cathepsin D cleaved both stefin A and stefin B inactivating them by limited proteolysis. The most susceptible bonds are Glu15-Ile16, Ala40-Val41, Tyr53-Tyr54 and Leu82-Thr83 in stefin A, Glu28-Glu29, Phe43-Lys44, Tyr53-Phe54 and Leu67-Arg68 in stefin B. It is evident that this cleavage specificity is similar to that of cathepsin D (for details see ref. 22 and this volume). The finding that cathepsin D inactivates stefins and possibly other cystatins may be of physiological importance. These results may explain increased cysteine proteinase activities and decreased levels of their inhibitors in distrophic muscles (23).

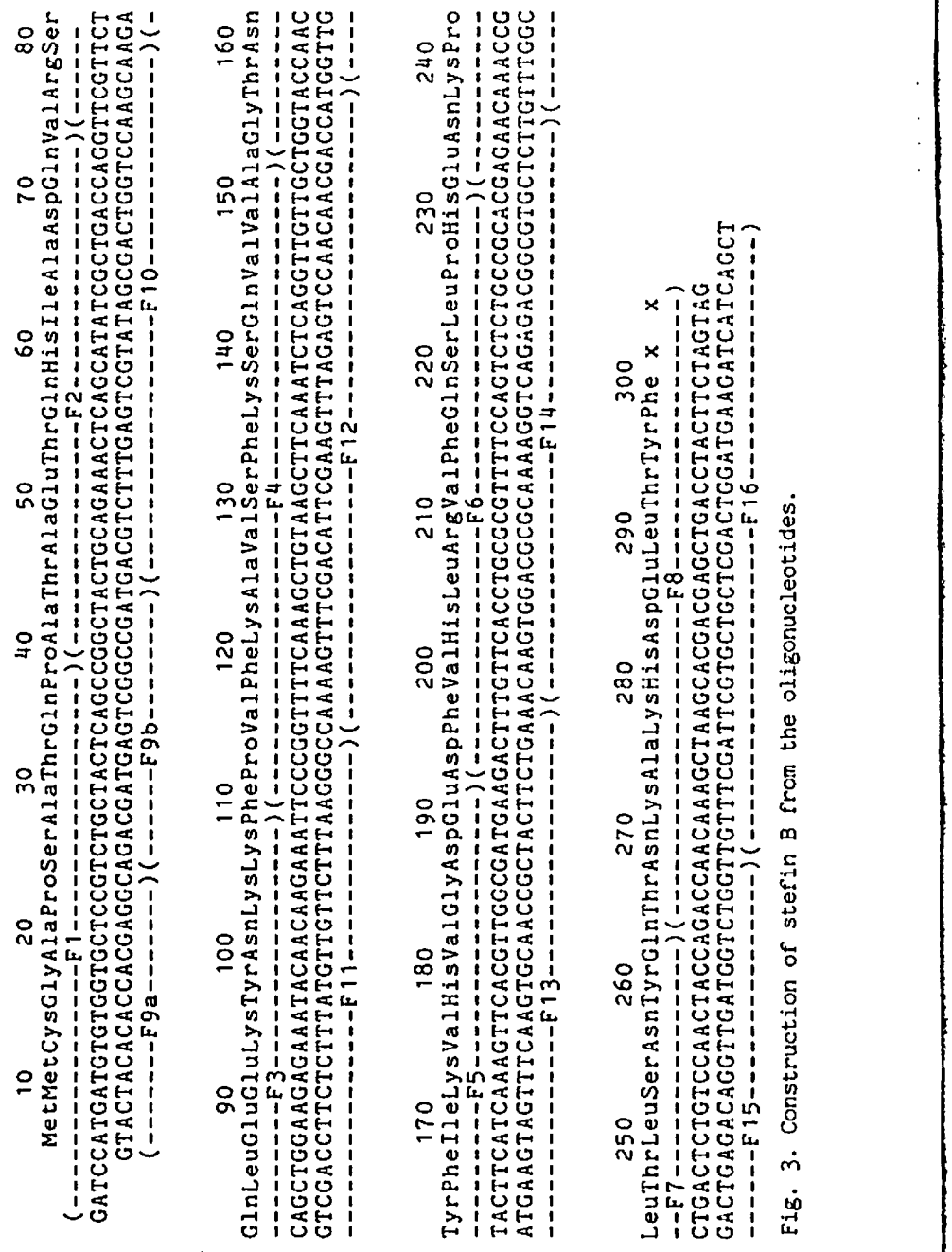

Fig. 3. Construction of stefin B from the oligonucleotides.

EVOLUTION OF CYSTATINS AND RELATED PROTEINS

The evolutionary tree (Fig. 4) is consistent with the clustering obtained in (14) and essentially the same as the dendrogram in (13). The description of evolutionary events similar to that in (13) can be adopted. The exceptions are the additions of oryzae-cystatin (27), puff adder venom cystatin (25) and histidine rich glycoprotein domains 1 and 2 (HRGD1 and HRGD2) (28). It can be inferred from the dendrogram that the orizae-cystatin belongs to the stefin family rather than

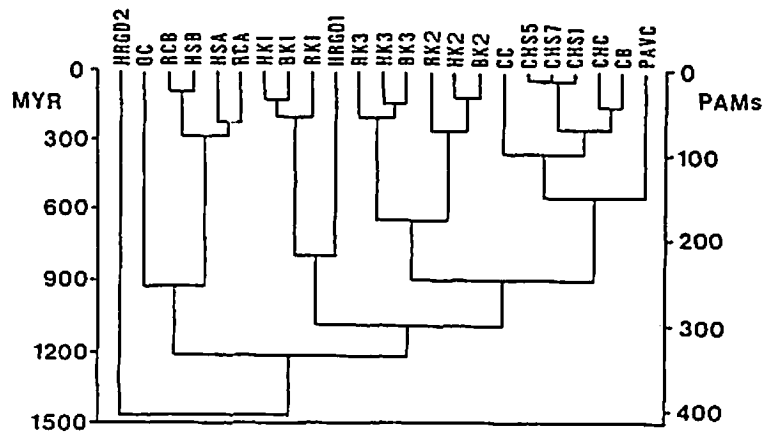

Fig. 4. Phylogeny of the cystatin superfamily. This dendrogram was constructed (using the method by Fitch & Margoliash (26)) from the corrected distance matrix obtained as described in (13). The scale on the right represents accepted point mutations (PAMs). The scale on the left indicates the postulated time flow in million of years (MYR). The sequences clustered are those of human stefin A (HSA), human stefin B (HSB), rat cystatin α (RCA), rat cystatin β (RCB), oryzae-cystatin (OC), human cystatin C (HCC), human saliva cystatin S7 (HSC7), human saliva cystatin S5 (HSC5), human saliva cystatin S1 (HSC1), bovine cystatin (BC), chicken cystatin (CC), puff adder venom cystatin (PAVC), histidine rich glycoprotein domain 1 (HRGD1), histidine rich glycoprotein domain 2 (HRGD2), human kininogen fragment 1 (HK1), bovine kininogen fragment 1 (BK1), rat kininogen fragment 1 (RK1), human kininogen fragment 2 (HK2), bovine kininogen fragment 2 (BK2), rat kininogen fragment 2 (RK1), human kininogen fragment 3 (HK3), bovine kininogen fragment 3 (BK3) and rat kininogen fragment 3 (RK3).

to the cystatin family and should be more appropriately named oryzae-stefin. In addition, HRGD1 is most similar to the kininogen domain 1 and might be a product of a gene duplication of the kininogen domain 1 gen in man. The puff adder venom cystatin clearly belongs to the cystatin family, although its pattern of insertions and deletions in the cystatin superfamily alignment significantly differs from the other inhibitors of the cystatin family (25). Finally, the dendrogram disputes the inclusion of the HRGD2 into the cystatin superfamily, since the similarities between HRGD2 and the members of the superfamily are not statistically significant.

In conclusion, it should be mentioned that the crystal structure of chicken egg white cystatin has recently been solved by X-ray difraction methods (24, this volume). The crystallographic analysis indicates the regions of stefins which may participate in the specific binding to the cysteine proteinases, most notably the Gln46-Gly50 region.

REFERENCES

1. Barrett, A. J. (1986) in Proteinase Inhibitors (Barrett, A. J. & Salvesen, G., eds.) pp. 3 - 22, Elsevier, Amsterdam
2. Turk, V. (1986) in Cysteine Proteinases and their Inhibitors (Turk, V., ed.), Walter de Gruyter, Berlin
3. Chan, S. J., San Segundo, B., McCormick, M. B. & Steiner, D. F. (1986) Proc. Natl. Acad. Sci. USA 86, 7721-7725
4. Kominami, E., Tsukahara, T., Hara, K. & Katunuma, N. (1988) FEBS Lett. 231, 225 - 228
5. Mort, J. S., Recklies, A. D. & Poole, A. R. (1984) Arthritis Rheum. 27, 509 - 515
6. Katunuma, N. & Kominami, E. (1983) Curr. Top. Cell. Regul. 22, 71 - 101
7. Sloane, B. F. & Honn, K. V. (1984) Cancer Metastasis Review 3, 249 - 263
8. Mullins, D. E. & Rifkin, D. B. (1985) in Developments in Cell Biology: Secretory Processes (Dean, R. T. & Stahl, P., eds.) pp. 159 - 177, Butterworths, London
9. Barrett, A. J., Rawlings, N. D., Davies, M. E., Machleidt, W. & Turk, V. (1986) in Proteinase Inhibitors (Barrett, A. J. & Salvesen, G., eds.) pp. 513 - 569, Elsevier, Amsterdam
10. Barrett, A. J., Fritz, H., Grubb, A., Isemura, S., Jarvinen, M., Katunuma, N., Machleidt, W., Mueller-

Esterl, W., Sasaki, M. & Turk, V. (1986) Biochem. J. 236, 311

11. Turk, V., Brzin, J., Kopitar, M., Kotnik, M., Lenarcic, B., Popovic, T., Ritonja, A., Babnik, J., Bode, W., & Machleidt, W. (1986) in Proteinases in Inflammation (Tschesche, H., ed.) pp. 77 - 92, Walter de Gruyter, Berlin
12. Mueller-Esterl, W., Fritz, H., Kellermann, J., Lottspeich, F., Machleidt, W. & Turk, V. (1985) FEBS Lett. 191, 221 - 226
13. Salvesen, G., Parkes, C., Rawlings, N. D., Brown, M. A., Barrett, A. J., Abrahamson, M & Grubb, A. (1986) in Cysteine Proteinases and their Inhibitors (Turk, V., ed.) pp. 413 - 428, Walter de Gruyter, Berlin
14. Sali, A. & Turk, V. (1987) Biol. Chem. Hoppe-Seyler 368, 493 - 499
15. Ritonja, A., Popovic, T., Turk, V., Wiedenmann, K. & Machleidt, W. (1985) FEBS Lett. 181, 169 - 172
16. Ritonja, A., Popovic, T., Kotnik, M., Machleidt, W. & Turk, V. (1988) FEBS Lett. 228, 341 - 345
17. Gal, S. & Gottesman, M. M. (1988) Biochem. J. 253, 303 - 306
18. Machleidt, W., Borchart, U., Fritz, H., Brzin, J., Ritonja, A. & Turk, V. (1983) Hoppe-Seyler's Z. Physiol. Chem. 364, 1481 - 1486
19. Ritonja, A., Machleidt, W., & Barrett, A. J. (1985) Biochem. Biophys, Res. Commun. 131, 1187 - 1192
20. Lenarcic, B., Ritonja, A., Sali, A., Kotnik, M., Turk, V. & Machleidt, W. (1986) in Cysteine Proteinases and their Inhibitors (Turk, V., ed.) pp. 473 - 487, Walter de Gruyter, Berlin
21. Straus, M., Bartsch, O., Stollwerk, J., Trstenjak, M., Boehming, A., Gassen, H. G., Machleidt, W. & Turk, V. (1988) Biol. Chem. Hoppe-Seyler 369, 209 - 218
22. Lenarcic, B., Kos, J., Dolenc, I., Lucovnik, P., Krizaj, I. & Turk, V. (1988) Biochem. Biophys. Res. Commun. in press
23. Gopalan, P., Dufresne, M. J. & Warner, A. H. (1987) Can. J. Pharmacol. 65, 124 - 129
24. Bode, W., Engh, R., Musil, Dj., Thiele, U., Huber, R., Karshikov, A., Brzin, J. Kos, J. & Turk, V. (1988) EMBO J. 7, 2593 - 2599
25. Barrett, A. J. (1987) TIBS 12, 193 - 1196
26. Fitch, W. M. & Margoliash, E. (1967) Science 155, 279 - 284
27. Abe, K., Emori, Y., Kondo, H., Suzuki, K. & Arai, S. (1987) J. Biol. Chem. 262, 16793 - 16797
28. Koide, T. & Odani, S. (1987) FEBS Lett. 216, 17 - 21
29. Kamphuis, I. G., Drenth, J. & Baker, E. N. (1985) J. Mol. Biol. 182, 317 -329