*Proc. Natl. Acad. Sci. USA*
Vol. 84, pp. 3590–3594, June 1987
Biochemistry

# Endogenous inhibitor for calcium-dependent cysteine protease contains four internal repeats that could be responsible for its multiple reactive sites

YASUFUMI EMORI*, HIROSHI KAWASAKI*, SHINOBU IMAJOH*, KAZUTOMO IMAHORI†, AND KOICHI SUZUKI*

*Department of Molecular Biology, The Tokyo Metropolitan Institute of Medical Science, Honkomagome, Bunkyo-ku, Tokyo 113, Japan; and †Mitsubishi-Kasei Institute of Life Sciences, 11 Minamiooya, Machida-shi, Tokyo 194, Japan

*Communicated by H. A. Scheraga, February 2, 1987*

**ABSTRACT** **A cDNA encoding an endogenous inhibitor, termed calpastatin, for calcium-dependent cysteine protease (calpain, EC 3.4.22.17) was cloned by screening rabbit cDNA libraries with a synthetic oligodeoxynucleotide probe based on the partial amino acid sequence of the purified protein. The deduced amino acid sequence contains 718 amino acid residues ($M_r$, 76,964), and the mature protein corresponds to the deduced sequence from the 80th residue of the primary translation product (resultant $M_r$, 68,113). This deduced molecular weight is significantly lower than that determined by NaDodSO$_4$/polyacrylamide gel electrophoresis, suggesting the possibility that the inhibitor is post-translationally modified. The sequence of the mature inhibitor contains four consecutive internal repeats ≈140 amino acid residues long, each of which might be responsible for the inhibitory activity. Calpastatin is apparently different from a typical cysteine protease inhibitor (cystatin), suggesting that the mechanism of inhibition of calcium-dependent cysteine protease by the inhibitor might be different from that of other cysteine proteases by cystatin.**

Calcium-dependent cysteine protease (calpain, EC 3.4.22.17) is a widely distributed intracellular cysteine protease involved in a variety of cellular processes mediated by $Ca^{2+}$ (1–3). The enzyme catalyzes the limited proteolysis of various proteins and its activity, which is absolutely dependent on $Ca^{2+}$, is controlled by a specific endogenous inhibitor, calpastatin (1–3).

The structure of calpain has been well characterized; its large subunit has four domains, including a papain-like cysteine protease domain and a calmodulin-like calcium-binding domain (4–6), and the small subunit consists of two domains, a glycine-rich hydrophobic domain and a calmodulin-like calcium-binding domain (7–9). However, the structure of calpastatin has not been determined, and various values have been reported for its molecular weight (10–15). The molecular weights estimated by NaDodSO$_4$/polyacrylamide gel electrophoresis were ≈110,000 for the human liver (13), rabbit skeletal muscle (14), and porcine heart (15) inhibitor and ≈70,000 for the porcine erythrocyte inhibitor (15). The existence of multiple reactive sites for inhibition has also been reported: 1 mol of calpastatin inhibits several moles of calpain, although the precise mechanism of inhibition is not known (13, 15).

As a step toward elucidation of the inhibition mechanism of calpastatin, its structure was determined by analysis of its cDNA sequence. The complete derived amino acid sequence of the inhibitor is presented with a discussion on the structure–function relationship.

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.

## MATERIALS AND METHODS

**Purification of Calpastatin from Rabbit Liver and the Designing of an Oligodeoxynucleotide Probe.** Calpastatin was purified from rabbit liver essentially as described (13). The purified inhibitor gave a single band on NaDodSO$_4$/polyacrylamide gel electrophoresis, and its various properties were similar to those of the human liver inhibitor (13). Purified calpastatin was digested with trypsin or endopeptidase Arg-C (Boehringer Mannheim), and then the peptides generated were separated by HPLC on a Hitachi Gel 3063 column. Several peptides were subjected to amino acid sequence analysis with a gas-phase sequencer (model 470A; Applied Biosystems, Foster City, CA). An 18-base oligodeoxynucleotide mixed probe, 5′d[(G or A)TC(T or C)TG(T or C)TC(T,G,C, or A)CC(T or C)TG(G or A)TC]3′, corresponding to the sequence Asp-Gln-Gly-Glu-Gln-Asp, was synthesized with a DNA synthesizer (model 380B; Applied Biosystems).

**Construction of cDNA Libraries.** Total RNA was extracted from rabbit lung or heart by the guanidinium thiocyanate method, and poly(A)$^+$ RNA was selected by oligo(dT)-cellulose column chromatography (16). Double-stranded cDNA was synthesized essentially as described by Gubler and Hoffman (17). Flush ends of the cDNA were generated with T4 DNA polymerase, and the double-stranded cDNA was treated with *Eco*RI methylase. Phosphorylated *Eco*RI linker was ligated, digested with *Eco*RI, and then fractionated by electrophoresis on a low gelling temperature agarose gel. cDNA >2 kilobases was inserted into phage vector λgt10 and packaged *in vitro* (Stratagene Cloning Systems, San Diego, CA), and then the phages were grown on *Escherichia coli* C600Hfl (18). More than 95% of the phages were recombinant.

**Screening of cDNA Libraries.** About $3 \times 10^5$ recombinant phages were screened with the synthetic probe labeled at the 5′ terminus with $^{32}$P. Hybridization was performed overnight at 50°C in a solution containing 50 mM Tris·HCl (pH 8.0), 1 M NaCl, 10 mM EDTA, 0.2% bovine serum albumin, 0.2% NaDodSO$_4$, heat-denatured salmon sperm DNA at 200 μg/ml, and $^{32}$P-labeled probe at $10^6$ cpm/ml. The final washing was performed with 6× SSC (1× SSC is 0.15 M NaCl/0.015 M sodium citrate, pH 7.0), 0.1% NaDodSO$_4$ at 50°C.

**DNA Sequencing.** The nucleotide sequence was determined using pUC plasmid vectors and the dideoxy chain-termination method (19) from both directions of inserts.

**RNA Blot Hybridization.** RNA was denatured and electrophoresed on a formaldehyde-containing agarose gel (19). After the electrophoresis, the RNA was transferred to a nitrocellulose membrane and then hybridized with a nick-translated cDNA probe at 42°C in a solution containing 50 mM sodium phosphate (pH 7.0), 5× SSC, 50% (vol/vol) formamide, 0.1% bovine serum albumin, 0.1% polyvinylpyr-

rolidone, 0.1% Ficoll 400, 0.2% $NaDodSO_4$, and heat-denatured salmon sperm DNA at 200 μg/ml. The filter was finally washed in 0.1× SSC/0.1% $NaDodSO_4$ at 55°C.

## RESULTS

**Isolation and Identification of cDNA Clones for Calpastatin.** From $3 \times 10^5$ plaques of a rabbit lung cDNA library, three cDNA clones containing cDNA inserts for calpastatin were obtained. Since the nucleotide sequence analysis and RNA blot hybridization experiment revealed that the cDNA inserts were not full length and since the content of mRNA for calpastatin was the highest in the heart, other cDNA libraries were constructed using rabbit heart poly(A)$^+$ RNA and an oligo(dT) primer or a synthetic oligodeoxynucleotide primer corresponding to the sequence near 5′ terminus of the cDNA obtained (see Fig. 2). By screening these two cDNA libraries, 3 and 20 cDNA clones were isolated, respectively, from the oligo(dT)-primed library ($5 \times 10^4$ phages) and the primer-extended library ($10^5$ phages).

Nucleotide sequence analysis, as shown in Fig. 1, revealed the presence of a single long open reading frame of 2154 nucleotides that encoded 718 amino acid residues ($M_r$, 76,964) (Fig. 2). This open reading frame was preceded by a termination codon in phase (positions −135 to −133 in Fig. 2). The deduced amino acid sequence contained all of the determined partial amino acid sequences of the peptides derived from purified calpastatin, indicating that this open reading frame actually encodes calpastatin. Furthermore, on the basis of the amino acid composition and the results of RNA gel blot analyses described below, we concluded that these cDNA clones cover the total amino acid coding region and the 3′-noncoding region of calpastatin mRNA.

**RNA Blot Analysis.** Three mRNA species of calpastatin (I, 3.8 kilobases; II, 3.0 kilobases; III, 2.5 kilobases) were identified in rabbit heart on RNA blot analysis when a probe for the protein coding sequence was used (Fig. 3, lane 1). Since three poly(A) addition sites were detected in the nucleotide sequence (Fig. 2), these three mRNA species were thought to result from different poly(A) addition sites. We tested this by using probes specific for the 3′-noncoding sequence. The extreme 3′-noncoding sequence (probe 3) only hybridized with the longest mRNA species (mRNA I; Fig. 3, lane 3), and probe 2 hybridized with two mRNA species (mRNAs I and II; Fig. 3, lane 2). These results clearly indicate that the three mRNA species contain different lengths of the 3′-noncoding sequence, and that the cDNA clones obtained here cover nearly the full length of the mRNA. Similar results were obtained when RNA from other tissues of rabbit were used (data not shown).

**Amino Acid Sequence of Calpastatin.** The open reading frame deduced from the nucleotide sequence comprised 718 amino acid residues. The N terminus of purified calpastatin determined with a gas-phase sequencer was Glu-Lys-Thr-Ala-Ser-Arg-Ser-Lys-Glu-Pro-Val. This sequence completely matched that from the 80th residue of the deduced sequence. Consequently, the purified inhibitor should consist of 639 amino acid residues, if C-terminal trimming does not occur. The amino acid composition of the 639 residues coincided well with that of the purified inhibitor (Table 1). Thus, mature calpastatin consists of 639 amino acid residues with a molecular weight of 68,133.

In a search for domain structures that might be responsible for the multiple reactive sites for inhibition, dot matrix analysis was performed. As shown in Fig. 4*B*, several cross lines were detected, indicating the existence of four internal repeat structures (Fig. 4*A*). Homologous sequences of about 50 residues were observed at about every 140 amino acid residues (Fig. 4*A*). A highly homologous sequence, Glu-Lys-Leu-Gly-Glu-Xaa-Glu-Xaa-Thr-Ile-Pro-Pro-Xaa-Tyr-Arg, flanked by relatively conserved sequences was observed in the middle of each repeat. These repeating structures may be correlated with the multiple reactive sites for inhibition, and the highly homologous regions might be the reactive sites of calpastatin.

## DISCUSSION

We have cloned and sequenced cDNA clones for calpastatin that cover the total amino acid coding region. Important features of the deduced structure are that (*i*) the calculated molecular weight of 68,133 is significantly lower than that deduced on $NaDodSO_4$/polyacrylamide gel electrophoresis ($M_r$, 110,000), (*ii*) four internal sequence repeats, presumed reactive sites, are presented, and (*iii*) there is no sequence homology with cystatin, the most typical cysteine protease inhibitor.

A wide range of molecular weights for calpastatin has been reported from $M_r$ 68,000 to 170,000 on the basis of $NaDodSO_4$ gel electrophoresis (10–15). The reasons for the diversity of the values obtained on $NaDodSO_4$ gel electrophoresis and the difference between these values and those calculated from the amino acid sequence are not clear. One of the most probable explanations is post-translational modification.

The inhibitory activity of calpastatin must be explained in terms of the deduced primary structure. Since calpastatin has multiple reactive sites (11, 13), the four internal repeating sequences (Fig. 4*A*) are the most probable candidates for the reactive sites. The inhibitor inhibits 3 to 8 mol of calpain per

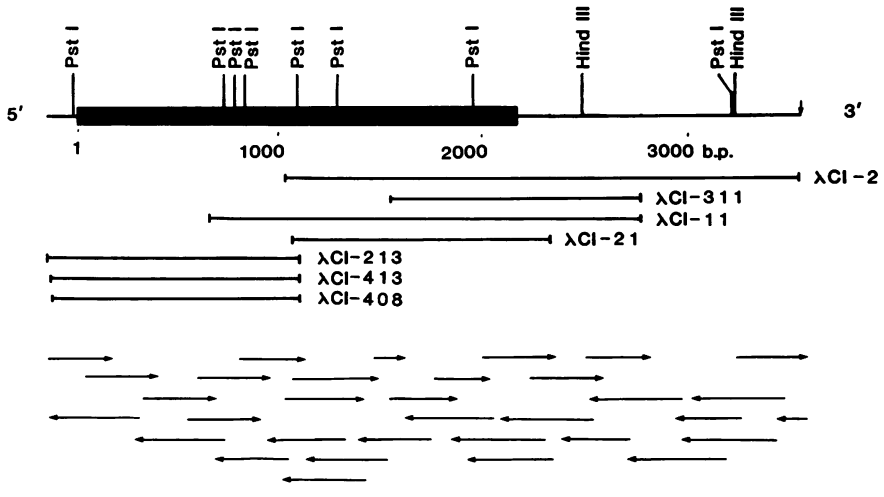


FIG. 1. Restriction map and sequencing strategy for cDNA clones of calpastatin. The amino acid coding region is shown by a solid box. Clone λCI-311 was obtained from a lung cDNA library, and the others were isolated from heart libraries. Clones λCI-2, 311, 11, and 21 had poly(A) tracts. Clones XCI-213, 413, and 402 were obtained from the primer-extended cDNA library. Arrows indicate the directions and lengths of sequencing.

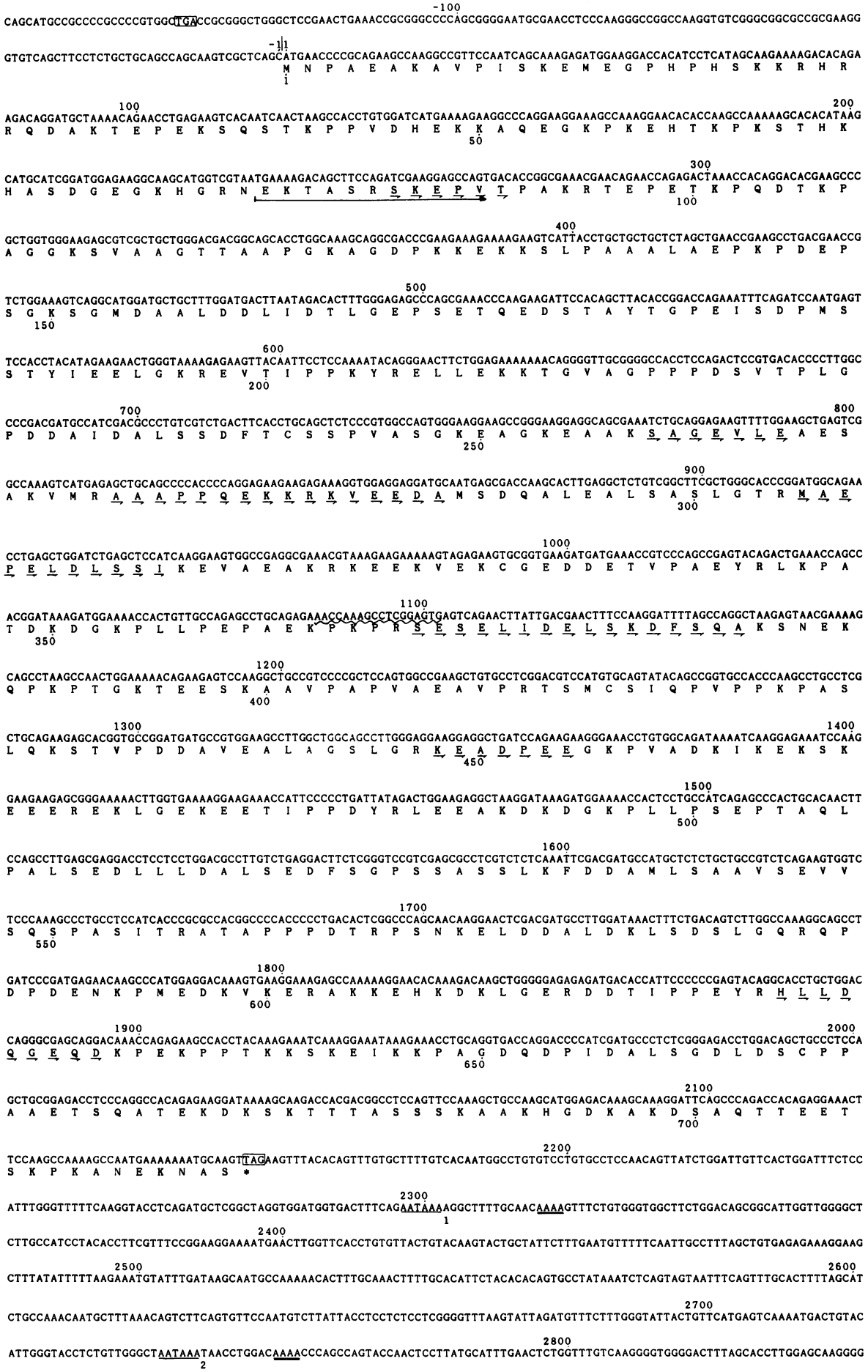

FIG. 2. *(Figure continues on the following page.)*

```
                                                                    2900
AAGAACAAAATGCAGGCGTGATAGCAAGTCACGGTACAGTGCATTCTGCGCATAAGTCTCAGCGTCTTGTCCTGTGGGGGCCCTGTGCTACACTCATGAACCTGCTTTCTGAGCTCTGCC

                                                          3000
CAGGAGCAGCTGTGGTTGCTTGGAGCCATGACCACAGTAGTCTGTGCACTCATGAAGCGGGATGCTAGACGGTGCAAAGTTGACCAGTCATCTACTTTAAAAACCGTACATACCTGAAAC

                    3100                                                                                          3200
TCAGAGTCTGCTTCCAGCACCCTGAGTGGTGTTCCGTGAGCATTGCACCGATGTTACGATGGGATTTTAATAAATCATCCGCTTCTGGGTGAATCCAGGCACGGAGGGCAGATGTGAGCT

                                                                                                    3300
GGAATTTGTCTGCAGAGAGAAGCTTTAATGAGGACTGAACTGACTTCACTACCGGAAAGTTAATAGCAACCTGTACTTCCTTGGAGGACAAGGTCAATGAAACGCCACACAGAACGGGCC

                                                                3400
CCAAATTTTATACTCCCTAAAGTTCATGCACAGAGCTACCTACGCAATTGCAAACTCCACCTGCGAAGTGCTGTGTCCCAGACACATTCTATGTTCTAGGCACCAACAGAACATGAGAAC

                                                          3500
ACTTTCCTAAAGGATTATACTATGTATTTAAGTCTCCCATTTGGACAGTACAGGTGTCTTCATTAAAAATCAATAAACTTGACATTGACAAAAAAAAAAAAAA
                                                                                                          3
```

FIG. 2. Nucleotide and predicted amino acid sequences of cDNA inserts for calpastatin. The deduced amino acid sequence is shown below the nucleotide sequence. Termination codons in phase are boxed. Poly(A) signal sequences, AATAAA, and the sites of poly(A) addition are underlined and double-underlined, respectively, and are identified as follows: 1, clone λCI-21; 2, clones λCI-311 and 11; 3, clone λCI-2. The N-terminal sequence of the purified 110-kDa calpastatin is indicated by a long arrow below the amino acid sequence. The determined amino acid sequences of tryptic or Arg-C peptides from the purified protein are indicated by short arrows below amino acid residues. A region corresponding to an oligodeoxynucleotide used for primer extension is also identified (nucleotide 1085–1102).

mol on the basis of the molecular weight estimated on $NaDodSO_4$ gel electrophoresis. Recalculation of the value for a molecular weight of 68,000 gives 1.8 to 4.9 mol, which roughly corresponds to the number of repeating units. The repeating structures of the four units most probably originated from an ancestral peptide of only one unit by means of a two-step gene duplication, although this must be confirmed by gene structural analysis (20).

A computer search using the National Biomedical Research Foundation protein data base* revealed that no other proteins including cystatin (21, 22), a typical member of the cysteine protease inhibitor family, have clear sequence homology to calpastatin. Moreover, the sequences of the presumed reactive sites (Fig. 4*A*) are completely different from cystatin. Calpastatin does not inhibit other cysteine proteases, and cystatin, except for the second domain of low-molecular-weight kininogen, does not inhibit calpain (22–24). Therefore, the structural dissimilarity between calpastatin and cystatin may not be surprising. It should be noted, however, that a sequence similar to -Thr(or Ser)-Ile-Pro-Pro- shown in Fig. 4*A* is found in the reactive sites of serine protease inhibitors (25).

The totally different structural properties of calpastatin from those of cystatin suggest that it is a new class of cysteine protease inhibitors and their mechanisms of inhibition might be different, although the most probable mechanism of inhibition is the direct interaction between calpastatin and the active site in the protease domain of calpain, as in the cases of most other protease inhibitors.

*National Biomedical Research Foundation (1986) Protein Sequence Data Base of the Protein Identification Resource (Washington, DC), Release No. 8.0.

This work was supported in part by research grants from the Ministry of Education, Science and Culture, and from National Center of Neurology and Psychiatry (NCNP) of the Ministry of Health and Welfare of Japan.

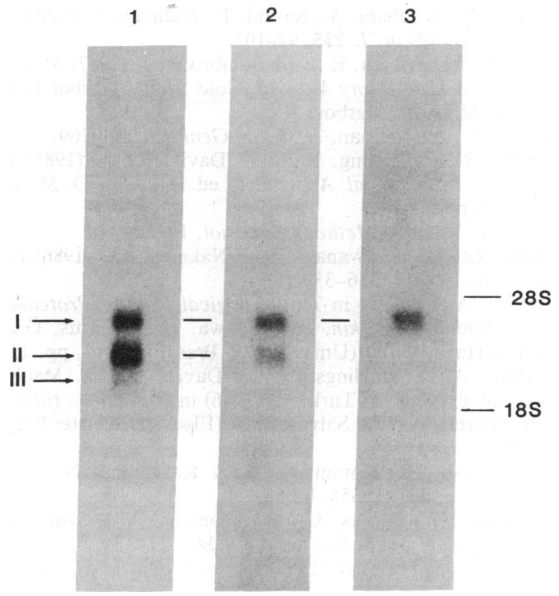


FIG. 3. RNA gel blot analysis. Poly(A)$^+$ RNA from rabbit heart was electrophoresed on a 1% agarose gel, blotted, and then hybridized with three probes. Lanes: 1, probe 1 covers the total amino acid coding region; 2, probe 2 is an 841-bp *Hin*dIII fragment (nucleotides 2479–3220 in Fig. 2); 3, probe 3 is a *Hin*dIII/*Eco*RI fragment of clone λCI-2 corresponding to nucleotides 3221–3544. The mRNA species (I, II, and III) and the locations of 28S and 18S rRNA markers are shown.

Table 1. Amino acid composition of rabbit calpastatin

| Residue | Amino acid composition: Deduced from cDNA, %* (numbers of residues) | Experimental, % |
|---|---|---|
| Asx (Asp) | 9.5 (56) | 10.3 |
| (Asn) | (5) | |
| Thr | 6.0 (38) | 4.8 |
| Ser | 9.9 (63) | 8.7 |
| Glx (Glu) | 14.8 (77) | 15.7 |
| (Gln) | (17) | |
| Pro | 10.8 (69) | 10.2 |
| Gly | 5.0 (32) | 6.5 |
| Ala | 11.7 (75) | 12.7 |
| Cys | 0.6 (4) | ND |
| Val | 3.6 (23) | 3.3 |
| Met | 1.3 (8) | 1.6 |
| Ile | 2.2 (14) | 1.4 |
| Leu | 7.0 (45) | 7.7 |
| Tyr | 0.9 (6) | 0.9 |
| Phe | 0.6 (4) | 1.1 |
| Lys | 12.5 (80) | 11.3 |
| His | 0.5 (3) | 0.8 |
| Arg | 3.1 (20) | 3.0 |
| Trp | 0.0 (0) | ND |
| Total | 100.0 (639) | 100.0 |

Predicted composition of purified calpastatin (amino acid residues 80–718, in Fig. 2). Amino acid composition of purified calpastatin was determined with a Hitachi 855 amino acid analyzer equipped with a fluorescent detection system. ND, not determined.

A

```
170  TQEDSTAYTGPEISDPMSSTYIEELGKREVTIPPKYRELLE-KKTGVAGPPPDS  222
307  EPELDLSSIKEVAEAKRKEEKVEKCGEDDETVPAEYRLKPATDKDGKPLLP-EP  359
450  ADPEEGKPVADKIKEKSKEEEREKLGEKEETIPPDYRLEEAKDKDGKPLLPSEP  503
587  PDPDENKPMEDKVKERAKKEHKDKLGERDDTIPPEYRHLLDQGEQDKPEKPPTK  640
                    * *  *****   **** **      * ***  *
                           *    * *  **              *
```

B

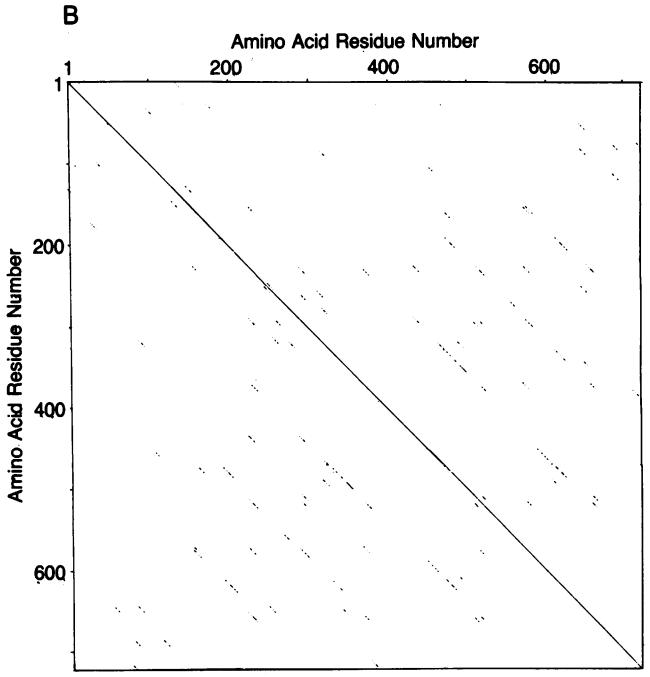


FIG. 4. (*A*) Amino acid sequences of the four internal repeats of calpastatin. The four repeats are aligned for maximal homology. Amino acid residues are shadowed when more than two residues are identical. Single stars indicate residues conserved in three of the four repeats. Double stars indicate residues conserved in all of the four repeats. The numbers on the left and right indicate the residue numbers of the starting and terminating residues, respectively (Fig. 2). (*B*) Dot matrix analysis of the amino acid sequence of calpastatin. Both the horizontal and vertical lines denote the amino acid sequence of calpastatin (718 residues). The possibility count was 5/8; dots were drawn only when more than 5 out of 8 residues were matched.


1. Imahori, K. (1982) in *Calcium and Cell Function*, ed.Cheung, W. Y. (Academic, New York), Vol. 3, pp. 473–485.
2. Murachi, T. (1983) in *Calcium and Cell Function*, ed. Cheung, W. Y. (Academic, New York), Vol. 4, pp. 377–410.
3. Kay, J. (1983) in *Proteases: Potential Role in Health and Disease*, eds. Horl, W. H. & Heidland, A. (Plenum, New York), pp. 519–532.
4. Ohno, S., Emori, Y., Imajoh, S., Kawasaki, H., Kisaragi, M. & Suzuki, K. (1984) *Nature (London)* **312,** 566–570.
5. Emori, Y., Kawasaki, H., Sugihara, H., Imajoh, S., Kawashima, S. & Suzuki, K. (1986) *J. Biol. Chem.* **261,** 9465–9471.
6. Aoki, K., Imajoh, S., Ohno, S., Emori, Y., Koike, M., Kosaki, G. & Suzuki, K. (1986) *FEBS Lett.* **205,** 313–317.
7. Sakihama, T., Kakidani, H., Zenita, K., Yumoto, N., Kikuchi, T., Sasaki, T., Kannagi, R., Nakanishi, S., Ohmori, M., Takio, K., Titani, K. & Murachi, T. (1985) *Proc. Natl. Acad. Sci. USA* **85,** 6075–6079.
8. Emori, Y., Kawasaki, H., Imajoh, S., Kawashima, S. & Suzuki, K. (1986) *J. Biol. Chem.* **261,** 9472–9476.
9. Ohno, S., Emori, Y. & Suzuki, K. (1986) *Nucleic Acids Res.* **14,** 5559.
10. Nakamura, M., Inomata, M., Hayashi, M., Imahori, K. & Kawashima, S. (1984) *J. Biochem. (Tokyo)* **96,** 1339–1407.
11. Ishiura, S., Tsuji, S., Murofushi, M. & Suzuki, K. (1982) *Biochim. Biophys. Acta* **701,** 216–223.
12. Lepley, R. A., Pampusch, M. & Dayton, W. R. (1985) *Biochim. Biophys. Acta* **828,** 95–103.
13. Imajoh, S., Kawasaki, H., Kisaragi, M., Mukai, M., Sugita, H. & Suzuki, K. (1984) *Biomed. Res.* **5,** 481–488.
14. Nakamura, M., Inomata, M., Hayashi, M., Imahori, K. & Kawashima, S. (1985) *J. Biochem. (Tokyo)* **98,** 757–765.
15. Takano, E., Kitahara, A., Sasaki, T., Kannagi, R. & Murachi, T. (1986) *Biochem. J.* **235,** 97–102.
16. Maniatis, T., Fritsch, E. F. & Sambrook, J. (1982) *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY).
17. Gubler, U. & Hoffman, B. (1983) *Gene* **25,** 263–269.
18. Huynh, T. K., Young, R. A. & Davis, R. W. (1985) *DNA Cloning. A Practical Approach*, ed. Glover, D. M. (IRL, Oxford), pp. 49–78.
19. Messing, J. (1983) *Methods Enzymol.* **101,** 20–78.
20. Müller-Esterl, W., Iwanaga, S. & Nakanishi, S. (1986) *Trends Biochem. Sci.* **11,** 336–339.
21. Barrett, A. J. (1985) in *The Biological Role of Protease and Their Inhibitor in Skin*, eds. Ogawa, H., Lazarus, G. S. & Hopsu-Hava, V. K. (Univ. Tokyo Press, Tokyo), pp. 13–26.
22. Barrett, A. J., Rawlings, N. D., Davies, M. E., Machleidt, W., Salvesen, G. & Turk, V. (1986) in *Proteinase Inhibitors*, eds. Barrett, A. J. & Salvesen, G. (Elsevier, Amsterdam), pp. 515–569.
23. Wakamatsu, N., Kominami, E. & Katunuma, N. (1982) *J. Biol. Chem.* **257,** 14653–14656.
24. Salvesen, G., Parkes, C., Abrahamson, M., Grubb, A. & Barrett, A. J. (1986) *Biochem. J.* **234,** 429–434.
25. Carrell, R. & Travis, J. (1985) *Trends Biochem. Sci.* **10,** 20–24.