Inhibition of Mammalian Legumain by Some Cystatins Is Due to a Novel Second Reactive Site*

Marcia Alvarez-Fernandez‡, Alan J. Barrett§, Bernd Gerhardt‡, Pam M. Dando§, Jian Ni‡, and Magnus Abrahamsson‡‡

From the §Department of Clinical Medicine, Institute of Laboratory Medicine, Lund University Hospital, S-221 85 Lund, Sweden, the ¶Medical Research Council Molecular Enzymology Laboratory, Babraham Institute, Babraham, Cambridge CB2 4AT, United Kingdom, and ¶¶Human Genome Sciences, Inc., Rockville, Maryland 20850-3338

We have investigated the inhibition of the recently identified family C13 cysteine peptidase, pig legumain, by human cystatin C. The cystatin was seen to inhibit enzyme activity by stoichiometric 1:1 binding in competition with substrate. The $K_i$ value for the interaction was 0.20 nM, i.e. cystatin C had an affinity for legumain similar to that for the papain-like family C1 cysteine peptidase, cathepsin B. However, cystatin C variants with alterations in the N-terminal region and the “second hairpin loop” that rendered the cystatin inactive against cathepsin B, still inhibited legumain with $K_i$ values 0.2–0.3 nM. Complexes between cystatin C and papain inhibited legumain activity against benzoyl-Asn-phenylalanine (BAPNA) as efficiently as did cystatin C alone. Conversely, cystatin C inhibited papain activity against benzoyl-Arg-phenylalanine (BARPA) whether or not the cystatin had been incubated with legumain, strongly indicating that the cystatin inhibited the two enzymes with non-overlapping sites. A ternary complex between legumain, cystatin C, and papain was demonstrated by gel filtration supported by immunoblotting. Screening of a panel of cystatin superfAMILY members showed that type 1 inhibitors (cystatins A and B) and low $M_r$ kininogen (type 3) did not inhibit pig legumain. Of human type 2 cystatins, cystatin D was non-inhibitory, whereas cystatin E/M and cystatin F displayed strong ($K_i$ 0.0016 nM) and relatively weak ($K_i$ 10 nM) affinity for legumain, respectively. Sequence alignments and molecular modeling led to the suggestion that a loop located on the opposite side to the papain-binding surface, between the $\alpha$-helix and the first strand of the main $\beta$-pleated sheet of the cystatin structure, could be involved in legumain binding. This was corroborated by analysis of a cystatin C variant with substitution of the Asn193 residue in this loop (N193K-cystatin C); this variant showed a slight reduction in affinity for cathepsin B ($K_i$ 1.5 nM) but $>>5,000$-fold lower affinity for legumain ($K_i$ $>>1,000$ nM) than wild-type cystatin C.

The activities of cysteine peptidases of the papain family (C1) such as cathepsins B, H, L, S, and K in and around mammalian cells are regulated by reversible, tight-binding protein inhibitors of the cystatin superfAMILY (1). The cystatins constitute a superfAMILY of evolutionarily related proteins that are all composed of at least one 100–120-residue domain with conserved sequence motifs (2). The single-domain human members of this superfAMILY are of two major types. The type 1 cystatins (or stefins) A and B contain approximately 100 amino acid residues, lack disulfide bridges, and are synthesized without signal peptides. Cystatins of type 2 are secreted proteins of approximately 120 amino acid residues ($M_r$ 13,000–14,000) and contain at least two characteristic intrachain disulfide bonds. The type 2 cystatins include the human cystatins C, D, S, SN, and SA, which are all products of genes located in the cystatin multigene locus on chromosome 20 (3). Two recently identified type 2 cystatins, cystatin E/M and cystatin F (also called leukocystatin), are also secreted low $M_r$ proteins but are more atypical in that they are glycoproteins and show only 30–35% sequence identity in alignments with the classical type 2 cystatins. They are, however, still functional inhibitors of family C1 cysteine peptidases (4–7). It has been shown that the cystatin inhibition of cysteine peptidases of the papain family is due to a tripartite wedge-shaped structure with very good complementarity to the active site clefts of such enzymes (8). The three parts of the cystatin polypeptide chain included in the enzyme-binding domain are the N-terminal segment, a central loop-forming segment with the motif Gln-Xaa-Val-Xaa-Gly, and a second C-terminal hairpin loop typically containing a Pro-Trp pair (8–10).

Legumain (EC 3.4.22.34) is a cysteine endopeptidase that was until recently known only from plants (11, 12) and Schisotosoma (13). In plants there is abundant evidence that legumain performs a protein-processing function, causing limited proteolysis of precursor proteins and protein splicing (12, 14). Following the discovery of the enzyme in mammalian cells, it was cloned and sequenced from human (15) and mouse (16). The amino acid sequences of legumains show that they belong to a distinct family of cysteine endopeptidases (C13). Mammalian legumain is predominantly lysosomal in distribution (16), but its strict specificity for the hydrolysis of bonds on the carboxyl side of asparagine is very different from that of any cathepsin and adapts it particularly for limited proteolysis (17). Human legumain may have an important physiological function as a key enzyme in antigen presentation (18).

It was recently reported that pig legumain is inhibited by human cystatin C and chicken cystatin with $K_i$ values below 5 nM (15). This finding was unexpected, since the cystatins are already known as potent inhibitors of the papain-like cysteine peptidases in the unrelated family C1. The legumain family members are believed to have a protein fold quite unlike that of papain, and to be much more closely related to the caspases and gingipain (19). Although the active site cysteine residue could seem to be a common factor, it is known not to be required...
for the interaction of papain with cystatins (20). The present investigation was undertaken to elucidate the mechanism of inhibition of mammalian legumain by cystatins.

**EXPERIMENTAL PROCEDURES**

**Proteins**—Recombinant human cystatin D (the natural Arg Variant) and wild-type human cystatin C were produced in *Escherichia coli* and purified as earlier described (21, 22). Recombinant wild-type human cystatin E/M and F were isolated after baculovirus expression (4, 5). Human cystatin A and bovine cystatin B were obtained from Dr. I. Björk (Department of Veterinary Medical Chemistry, Swedish Agricultural University, Uppsala, Sweden) as isolated *E. coli* produced proteins (23). Human L-kininogen-1 was isolated from blood plasma (24).

Papain (EC 3.4.22.2) was purchased from Sigma (catalog no. P4762) and used directly or after affinity purification on Sepharose-coupled Gly-Gly-Tyr-Tyr, which resulted in 70–75% activable enzyme (25, 26). Papain was carboxymethylated using a previously described procedure (27). Mammalian legumain (EC 3.4.22.14) was purified from pig kidney and characterized as recently reported (15). Human cathepsin B (EC 3.4.22.1) was obtained from Calbiochem (La Jolla, CA).

**Production of Cystatin C Variants**—Cystatin C devoid of the N-terminal Gly-Gly-Tyr-Arg, which resulted in 70–75% activable enzyme (25, 26). Dimeric cystatin C was obtained from wild-type recombinant human cystatin C with native cystatin C, were isolated after mutagenesis in an W106G-, (R8G,L9G,V10G)-, and (R8G,L9G,V10G,W106G)-cystatin C, cystatin C variants with Gly replacements for one, three, or four of the residues involved in papain interactions, W106G-, (R8G,L9G,V10G), and (R8G,L9G,V10G,W106G)-cystatin C, were isolated after mutagenesis in an *E. coli* expression system (28). Dimeric cystatin C was obtained from wild-type recombinant human cystatin C variant for 30 min at 2°C, and purified from trace amounts of monomeric cystatin C by SEC (29).

A cystatin C variant with Lys substitution for residue Asn 39 was obtained by oligonucleotide-directed mutagenesis of the cystatin C cDNA gene in pH313 (22, 30), using a PCR protocol. Taking advantage of a unique PstI recognition site located 24 base pairs downstream from the Asn 39 codon, a downstream primer was designed to introduce a C

A nucleotide substitution in the Asn 39 codon to result in a AAA codon for Lys (5′-CCACCTGCGACGCCGCGCTGTGGTGATCATGTTTGGCT-3′; -strand, with PstI site and mutation underlined). This oligonucleotide was used together with the upstream primer vector MA206 (26), 5′-GTTTCGCCTGTCTGTTTTGC-3′, (both at 0.4 μM final concentration) to amplify a 360-base pair fragment by PCR using 0.1 ng of pH313 DNA as template. DNA polymerase and PCR buffer were from the AmpliTaq kit (Perkin-Elmer-Cetus), and the PCR was accomplished by 30 incubation cycles of (95° C, 1 min; 57° C, 1 min; 72° C, 1 min) in a Perkin-Elmer-Cetus 2400 thermocycler. The PCR product was purified (PCR Purification Kit; Genomed, Oeynhausen, Germany), digested with Clal and PstI (Life Technologies, Inc., Paisley, United Kingdom) and ligated into Clal-cut and dephosphorylated pH313, to generate plasmid pCmut39K. The plasmid was introduced into Es coli MC1061 as described in detail elsewhere (26). That the plasmid was correctly mutated was verified by complete nucleotide sequencing of the cystatin C insert, as described (26). The conditions for culturing and induction of expression in bacteria containing pCmut39K were as described previously for wild-type cystatin C production using pH313 (22). Periplasmic extracts containing the recombinant cystatin C variant were obtained by cold osmotic shock (31) and directly applied to a Superdex 75 (Amershall Pharmacia Biotech, Uppsala, Sweden) SEC column (1.6 × 100 cm; in 50 mM ammonium bicarbonate buffer, pH 7.8, containing 100 mM NaCl). Fractions containing N39K-cystatin C were identified by PAGE analysis of a highly purified enzyme sample, by standard methods.

**Immunoblotting**—To verify the identity of protein bands after SDS-PAGE separation (above), transfer to PVDF membranes (Immobilon-P; Millipore, Bedford, MA) was performed using electrophoresis (Trans-Blot;SDS; Bio-Rad). Immunodetection of cystatin C was done exactly as described before (36). The same procedure was followed for Cm-papain detection with polyclonal rabbit anti-papain antibodies (produced by standard immunization procedures, using Cm-papain as antigen). Legumain was detected by use of horseradish peroxidase-conjugated monoclonal anti-canavalin A (Sigma; catalog no. L6397) at a final concentration of 5 μg/ml. Antibody- or concanavalin A-detected protein bands were visualized using chemiluminescence (ECL Plus; Amersham Pharmacia Biotech). Band intensities were assessed by densitometric scanning, using a Bio-Rad Imaging Densitometer GS-670 and Molecular Analyst software (Bio-Rad).

**Enzyme Inhibition Assays**—The methods used for active site titration of papain (with Bz-n-AArg-NHPhNO3 as substrate; Bachem Feinchemikalien, Bubendorf, Switzerland) and for titration of the molar papain-inhibitory concentration in cystatin preparations have been reviewed (1). Active inhibitor concentrations determined in this way were used for calculation of Ki values, as this is the method traditionally used. A cystatin C preparation derived from freshly isolated cystatin C with active pig legumain (from a 30 min cystatin C with active pig legumain and further incubating the mixture, at room temperature, for 30–60 min. Mixtures analyzed by SEC contained Cm-papain, and legumain at final concentrations of 81, 16, and 10 μM, or 9, 24, and 14 μM, respectively.

Separation and size estimation of the different enzyme-inhibitor complexes was performed by SEC on a Superdex 75 HR 10/30 column (Amersham Pharmacia Biotech) equilibrated in 50 mM sodium citrate buffer, pH 5.6, containing 150 mM NaCl. The column was operated at a flow rate of 0.5 ml/min using an HPLC system (Waters) equipped with multiple wavelength detector and an integration system (Waters 990). Ovalbumin (Mr, 43,000), bovine serum albumin (Mr, 67,000), chymotrypsinogen (Mr, 23,400), and carbonic anhydrase (Mr, 30,000) were used for construction of a calibration curve.

The fractions corresponding to each SEC peak were pooled and concentrated by precipitation, by addition of nine volumes of 20% (w/v) trichloroacetic acid. Precipitated proteins in pellets obtained after centrifugation were resuspended in a minimal volume of SDS sample buffer, and analyzed by SDS-PAGE in 16.5% gels using the buffer system described by Schägger and von Jagow (35).

The total cystatin and enzyme concentrations determined in this way are compared with total protein concentration determined by A280 measurement (22, 24). The apparently lower inhibitor concentration is likely due to some of the papain molecules being catalytically inactive (possibly due to oxidation of the catalytic Cys residue) but still capable of binding cystatin. The stoichiometry of papain-cystatin C binding is indeed 1:1 viewed by molar total protein concentrations, according to fluorescence titration (33). Therefore, total protein concentrations are used in the text when describing SEC experiments with papain-cystatin C mixtures.

The fluorogenic substrate used for determination of equilibrium constants for dissociation (Kd) of complexes between cystatins and family C1 cysteine peptidases (1) was Z-Phe-Arg-NHMec (10 μM; from Bachem Feinchemikalien) and the assay buffer was 100 mM sodium phosphate buffer (adjusted to pH 6.5 for papain, pH 6.0 for cathepsin B), containing 1 mM dithiothreitol and 1 mM EDTA. Steady state velocities were measured before and after addition of the cystatin variant under study in assays at 37°C, and Ki values were calculated according to Henderson (37). Corrections for substrate competition were made using Kd, with inhibitor (protein) batch used for titration. The assay conditions employed (60 and 55 μM for papain and cathepsin B, respectively).

Essentially the same procedures were used for legumain inhibition assays. Pig legumain was titrated with a cystatin C solution of known total protein concentration in a microtiter plate format using Bz-n-AArg-NHPhNO3 as substrate (Bachem Feinchemikalien), in sodium citrate/
phosphate buffer (15), pH 5.8 (39.5 mM citric acid, 121 mM Na₂HPO₄), containing 1 mM dithiothreitol, 1 mM EDTA, and 0.1% (w/v) CHAPS. The same buffer was used for fluorogenic continuous rate assays at 37°C with 10 μM Z-Ala-Ala-Asn-NHMec, prepared as described by Kemhavi et al. (11), as substrate. The legumain concentration used for Ki determinations in such assays was 0.1–0.5 nM. Results from such assays with substrate concentration in the range 5–50 μM were used to assess whether the cystatin interaction was competitive with substrate binding, by standard methods. The Km value for legumain hydrolysis of this substrate under the assay conditions, used for corrections of apparent Kᵢ values, was 30 μM.

Computer Modeling—Sequence alignments were carried out using programs in the GCG package (38), taking into account the structural alignment (39) of the known structures of chicken cystatin (Protein Data Bank (PDB) identification codes ICEW and 1A67; Refs. 8 and 40), cystatin A (PDB codes 1DVC and 1DVD; Ref. 41), and cystatin B (PDB code ISTF; Ref. 10), all obtained from the Brookhaven Protein Data Bank (42, 43). Graphic illustrations were produced using the program Swiss-PdbViewer (44, 45) and then rendered with QuickDraw3D (Apple Computer, Inc.).

RESULTS

Cystatin C Is a Tight-binding Legumain Inhibitor—To clarify whether cystatins are efficient legumain inhibitors or not, the interaction between pig legumain and human cystatin C was initially investigated. Cystatin C was able to completely inhibit legumain activity against Bz-Asn-NHPhNO₂ in a time- and dose-dependent manner. Titration curves drawn from experiments with varying cystatin concentrations in such assays were linear (see Fig. 1 for example). As the concentrations of enzyme and inhibitor in the assay were in the order of 1 μM, this indicated a relatively tight complex between enzyme and inhibitor with a Kᵢ value below 10 nM. Using Z-Ala-Ala-Ala-Asn-NHMec in different concentrations as substrate in a continuous-rate legumain assay at lower enzyme concentration (0.1–0.5 nM), it was observed that the cystatin C interaction with the enzyme was reversible and competing with substrate binding (results not shown). Corrected for substrate competition, the Kᵢ value for the cystatin C-legumain complex was 0.20 nM.

Studies by SEC gave evidence for a 1:1 interaction between inhibitor and enzyme, with a faster eluting peak (9.0 ml) appearing upon mixing of the proteins, corresponding to a Mr of 46,300 (Fig. 2A). This size agrees very well with a theoretical Mr of 46,000–47,000 calculated as the sum of the Mr for cystatin C (13,343, from the amino acid sequence of recombinant cystatin C; Ref. 30) and the Mr for pig legumain according to SDS-PAGE of the glycosylated native enzyme (33,600 and 33,100 estimated from the SDS-polyacrylamide gels in Figs. 3C and 2B, respectively). The retention volumes of 9.65 and 12.4 ml for the individual components on the calibrated SEC column (Fig. 2A, arrows) equaled Mr values of 35,300 and 13,400 for legumain and cystatin C, respectively. Analysis of the peak at 9.0 ml by SDS-PAGE agreed with the expected protein staining for legumain and cystatin C if they were present in equimolar amounts in a complex (Fig. 2B). In addition, the dissociated complex in the SDS gel demonstrated a cystatin C band size identical to that of the native inhibitor with no signs of degradation products, indicating that the inhibitor is not cleaved as a result of the enzyme interaction.

The Papain-Inhibitory Site on Cystatin C Is Not Responsible for Legumain Inhibition—To elucidate which parts of the cystatin structure are involved in binding and inhibition of mammalian legumain, the interactions of cystatin C variants with alterations in the N-terminal region and the second hairpin loop were studied (Table 1). The (des1–10)-cystatin C variant, devoid of the N-terminal decapeptide as a result of neutrophil elastase cleavage and with seriously compromised affinities for cathepsins B, H, and L (28), showed the same affinity for legumain as native cystatin C. Three cystatin C variants with Gly replacements for up to four critical residues with side chains participating in the high affinity binding between the inhibitor and papain-like cysteine peptidases (26) also displayed virtually unaltered affinity for legumain. Additionally, dimerized cystatin C, which has been shown to lose papain-inhibitory activity completely and to be due to intermolecular binding via the papain-inhibitory reactive site regions of two monomeric units (46, 47), was essentially as efficient as the monomeric cystatin in the inhibition of legumain. Thus, the cystatin C surface responsible for the inhibition of papain-like enzymes seemed not to be involved in legumain binding.

The strict substrate specificity of legumain, with a requirement for an Asn residue in the P₁ position, allowed studies of the papain-cystatin C interaction in the presence of legumain by use of Bz-Arg-NHPhNO₂ as substrate (Fig. 1A). It was observed that the dose-dependent inhibition of papain by dilu-
tions of a cystatin C solution was virtually unaffected when a portion of the same cystatin solution had been preincubated with an approximately equimolar amount of legumain (under conditions favoring stoichiometric interactions between enzyme and inhibitor). Essentially identical cystatin C-enzyme mixtures could be analyzed for the presence of legumain-inhibitory sites, as papain showed very slow hydrolysis of the legumain substrate, Bz-Asn-NHPhNO₂ (Fig. 1B). The dose-dependent inhibition of legumain by dilutions of the cystatin C solution was largely unaffected when a portion of the same cystatin solution was preincubated with papain (at the highest concentration possible given the concentrations of the stock solutions used). In the fluorogenic legumain assay with Z-Ala-Ala-Asn-NHMec as substrate, cystatin C preincubated with papain in 1:1 and 1:10 molar ratio displayed $K_i$ values of 0.26 and 0.33 nM, respectively, i.e. very similar to cystatin C alone (0.20 nM).

Demonstration of a Ternary Complex—Taken together, the results described above left little doubt that cystatin C inhibits legumain by a site that is distinct from that inhibiting papain and related peptidases. Consistent with this, the kinetic experiments indicated that the cystatin could simultaneously bind both legumain and papain, despite the small size of the cystatin molecule (M, 13,343) and the 2- and 3-fold larger papain and legumain molecules, respectively. Attempts were therefore made to detect a ternary complex between the cystatin and the two peptidases. For experiments with excess papain, Cm-pa-
Legumain Inhibition by Cystatins

Equilibrium constants for dissociation of pig legumain and human cathepsin B complexes with cystatin C variants were determined by steady-state kinetics in continuous rate assays at 37 °C with Z-Ala-Ala-Asn-NHMec and Z-Phe-Arg-NHMec, respectively, as substrate. The standard deviation (S.D.) and number of measurements (n) used to calculate the mean $K_i$ values are indicated. The $K_i$ values were corrected for substrate competition as described under "Experimental Procedures."

| Cystatin C variant | Legumain | Cathepsin B |
|-------------------|----------|-------------|
|                   | $K_i$    | S.D.        | $K_i$ | S.D. | n |
| Wild type         | 0.20     | ±0.014      | 0.25  |      | 6 |
| (des1–10)-        | 0.26     | ±0.056      | >1000 |      | 4 |
| (R9G,L9G,V10G,W106G)- | 0.22     | ±0.011      | >1000 |      | 3 |
| Dimeric$^c$       | 0.39     | ±0.027      | >1000 |      | 2 |
| N39K-             | >1000$^b$|      | 1.5    | ±0.14| 3 |

$^a$ $\Delta V_{nm} > 0.95$ at [I] = 670 nM.

$^b$ Results for W106G- and (R8G,L9G,V10G)-cystatin C were, as that for the (R8G,L9G,V10G,W106G)- variant, not significantly different from the wild type cystatin $K_i$ value.

$^c$ Calculated from active papain-inhibitory concentration of monomeric cystatin C, determined prior to dimerization, and assuming no protein loss during the following SEC purification.

$^d$ $\Delta V_{nm} > 0.95$ at [I] = 510 nM.

$^e$ $\Delta V_{nm} > 0.95$ at [I] = 380 nM.

Legumain Inhibitory Activity of Cystatin Superfamily Members—A panel of different cystatins was tested for inhibitory activity against pig kidney legumain. The results (Table II) showed that the type 2 inhibitors analyzed, human cystatin A and bovine cystatin B, had no significant inhibitory activity for pig legumain. Similarly, the type 3 inhibitor human L-kininogen, containing two cystatin repeats with inhibitory activity against papain-like cysteine peptidases, did not show any legumain-inhibitory activity. Of human type 2 cystatins, cystatins D was also non-inhibitory. Cystatin E/M demonstrated tight-binding affinity for pig legumain, with a $K_i$ value almost 100-fold lower than that of cystatin C, whereas cystatin F showed a significant but lower affinity for the enzyme, 50-fold lower than that of cystatin C.

Possible Location of the Legumain Inhibitory Site on Cystatins—An alignment of the amino acid sequences for the different cystatins (Fig. 4) was inspected for similarities between the proteins showing activity against legumain. Of those segments known to be present on the surfaces of the cystatin molecules (8, 10, 40, 41, 47), the loop starting at Asn$^{39}$, following the conserved a-helix, seemed like a good candidate for inclusion in a legumain-binding surface. This was because: 1) it is located on the opposite side to the papain-inhibitory surface (Fig. 5), which would allow the simultaneous binding of legumain and papain observed for cystatin C; and 2) it contains residues that are conserved among the cystatins with legumain-inhibitory activity, including an Asn residue, which could be directly involved in interactions with the substrate specificity pocket of legumain, given the strict specificity of legumain for asparaginyl bonds (17). To try to verify the possibility that the Asn$^{39}$ residue is intimately involved in the legumain-inhibitory site, this residue was mutated in cystatin C. The substitution selected was Asn$^{39}$ → Lys, guided by the notion that the non-inhibitory cystatin D, as well as cystatin B, have a positively charged residue in the loop segment. The N39K cystatin C variant was produced by E. coli expression and purified to homogeneity. The variant showed a 5-fold decreased affinity for cathepsin B as compared with wild-type cystatin C (Table I), but total loss of inhibitory activity for pig legumain ($K_i$, >1,000 nM, equaling >5,000-fold lower legumain affinity than wild-type cystatin C). This strongly indicated that the loop between the a-helix and the first strand of the main $\beta$-pleated sheet of the cystatin structure, and its Asn$^{39}$ residue, is part of a novel second reactive site of some cystatins.

DISCUSSION

The aim of the present investigation was to study the mechanism of inhibition of mammalian legumain by cystatins, to clarify how the inhibitor structure can result in tight-binding inhibition of enzymes belonging to two entirely different enzyme families, namely the papain family (C1) and the legumain family (C13). The different arrangements of catalytic residues

---

2 A. J. Barrett, unpublished observation.
Equilibrium constants for dissociation of pig legumain complexes with recombinant bovine cystatin B, recombinant human cystatins A, C, D, E/M, and F, and low Mr, kininogen purified from human blood plasma were determined by steady-state kinetics in continuous rate assays at 37 °C with Z-Ala-Ala-Asn-NHMec as substrate. The standard deviation (S.D.) and number of measurements (n) used to calculate the mean Ki values given are indicated. The Ki values were corrected for substrate competition as described under “Experimental Procedures.” Kd, rate of substrate hydrolysis in absence of inhibitor; v, rate of substrate hydrolysis in presence of inhibitor. The quality of the cystatin preparations used was checked by determination of Ki values for their interaction with papain or cathepsin B (published values are reviewed in Refs. 1 and 5).

| Cystatin  | Kd (nM) | S.D. | n | Cystatin activity control |
|-----------|---------|------|---|--------------------------|
| Cystatin A| >1000   |      |   | Cathepsin B: Kd = 1.5 ± 0.58 nM (n = 2) |
| Cystatin B| >1000   |      |   | Papain: Kd = 0.037 ± 0.010 nM (n = 4) |
| Cystatin C| 0.20    | ±0.014 | (6) | Cathepsin B: Kd = 0.25 nM (n = 1) |
| Cystatin D| >1000   |      |   | Papain: Kd = 1.9 ± 1.4 nM (n = 4) |
| Cystatin E/M| 0.0016 | ±0.00052 | (5) | Papain: Kd = 0.39 ± 0.13 nM (n = 12) |
| Cystatin F| 10      | ±1.0 | (4) | Papain: Kd = 1.1 ± 0.26 nM (n = 5) |
| L-kininogen| >1000   |      |   | Papain: Kd = 0.030 ± 0.012 nM (n = 3) |

Inhibition of legumain by cystatins of type 1, 2, and 3

Fig. 4. Sequence alignment of human members of the cystatin superfamily. The alignment was done based on secondary structure elements from the NMR models of cystatin A and chicken cystatin (not represented in the figure) (8, 39–41). Arrows represent β-strands (yellow and blue for type 1 and 2, respectively) and red cylinders represent α-helices. The region possibly involved in a legumain-inhibitory site (the back-side loop) is shaded in yellow. This loop-forming segment is magnified below; residues with similar chemical properties present in the three cystatins showing legumain-inhibitory activity (cf. Table II) but not in the others, and thus possibly constituting a consensus sequence for legumain binding, are boxed. Residue Asn39 in cystatin C is shown in orange for cystatin A and in light blue in cystatin C, in accordance with the colors used in Fig. 6. The numbering refers to the cystatin C sequence as deduced from its cDNA, starting from the first residue of the mature protein (30, 48). For the other cystatins, the naturally occurring forms with longest N-terminal segments are shown (4, 5, 21, 49–51). Dots indicate gaps introduced to optimize the alignment.

and different active site motifs show that the two families are evolutionarily unrelated, and that their peptidases have different protein folds (19). Moreover, legumain is not inhibited by the general inhibitor of enzymes belonging to family C1, E-64 (15), which supports the theory that the general topography of the active site clefts of legumains are probably entirely different from those of family C1 enzymes. Our present SEC, electrophoretic and enzyme kinetic results show that cystatin C can inhibit mammalian legumain, as cystatins inhibit family C1 enzymes (52), by high affinity reversible binding (Ki, 0.20 nM), in a bimolecular reaction that is competitive with substrate, and with no detectable cleavage of the cystatin in the legumain complex. Despite these similarities, our present results demonstrate that the mechanisms of inhibition of legumain and family C1 endopeptidases must be completely different.

From structural studies of several cystatins (8–10), it is well known that the N-terminal segment together with the “first and second hairpin loops” in cystatins are responsible for the inhibition of the C1 enzymes (Fig. 5). Consequently, removal of the N-terminal segment or substitution of any of the conserved amino acids in the N-terminal segment or the hairpin loops by Gly/Ala residues abolishes or seriously affects the inhibition of papain-like enzymes (26, 53). Of four such variants analyzed in the present study, all displayed virtually unaltered binding of legumain. Dimeric cystatin C, which is completely inactive...
against papain-like enzymes and by NMR studies has been shown to be a result of intermolecular interactions between the papain-binding surfaces of two cystatin C molecules (47), still showed legumain inhibition. We believe that this, together with the enzyme kinetic results presented and the direct demonstration of a ternary complex by SEC, proves that the binding sites for papain and legumain on cystatin C likely are completely independent of each other.

Where then is the legumain reactive site? Our investigation of a set of other mammalian cystatin superfamily members indicated that the capacity to inhibit legumain is a property of only some cystatins (Table II). Guided by this result and amino acid sequence comparisons, we propose that the side of cystatins directly opposite to the papain-binding surface is responsible for the legumain binding and inhibition. The loop segment connecting the main α-helix of the cystatin structure to the first long β-strand contains a conserved Asn residue (residue 39 in cystatin C) and seems quite conserved in sequence in those cystatins that show inhibitory activity: cystatins C, E/M, and F. An importance of the Asn39 residue was confirmed by construction of the N39K cystatin C variant, which was seen to lack legumain inhibitory activity. A correctly positioned Asn residue on the cystatin surface could possibly result in an initial substrate-like interaction between the inhibitor and legumain. Besides the requirement for an Asn residue in the P1 position, legumains have no clear preferences for residues in other subsites (17, 54). There are therefore few obvious structural possibilities for specific legumain inhibition besides interaction with the S1 pocket. Still, assuming that the “back-side loop” containing Asn39 interacts with the enzyme in a mode resembling substrate binding, it appeared from the inhibition data obtained that a loop segment preferentially containing polar amino acids is compatible with legumain interaction. The consensus sequence found in the three inhibitory cystatins is Ser(Thr)-Asn39-Asp(Ser)-Met(Ile). Strikingly, a Ser38-Asn39-sensus sequence found in the three inhibitory cystatins is compatible with legumain interaction. The obtained results suggest a loop segment preferentially containing polar amino acids is responsible for legumain binding and inhibition. The loop segment connecting the main α-helix of the cystatin structure to the first long β-strand contains a conserved Asn residue (residue 39 in cystatin C) and seems quite conserved in sequence in those cystatins that show inhibitory activity: cystatins C, E/M, and F. An importance of the Asn39 residue was confirmed by construction of the N39K cystatin C variant, which was seen to lack legumain inhibitory activity. A correctly positioned Asn residue on the cystatin surface could possibly result in an initial substrate-like interaction between the inhibitor and legumain. Besides the requirement for an Asn residue in the P1 position, legumains have no clear preferences for residues in other subsites (17, 54). There are therefore few obvious structural possibilities for specific legumain inhibition besides interaction with the S1 pocket. Still, assuming that the “back-side loop” containing Asn39 interacts with the enzyme in a mode resembling substrate binding, it appeared from the inhibition data obtained that a loop segment preferentially containing polar amino acids is compatible with legumain interaction. The consensus sequence found in the three inhibitory cystatins is Ser(Thr)-Asn39-Asp(Ser)-Met(Ile). Strikingly, a Ser38-Asn39-Asp40 sequence is completely conserved in mouse, rat, and bovine cystatin C, as well as in chicken cystatin, which also inhibits pig legumain (15). The positively charged Lys residue in this segment, present in the non-inhibitory cystatins B and D, may be unfavorable for inhibition.

The suggested binding loop must be able to adopt a conformation to allow legumain interaction, but at the same time not expose the Asn39-Xaa bond to cleavage. A different back-side loop conformation may be one reason why the type 1 cystatins studied, with a loop sequence largely containing the proposed consensus sequence for legumain inhibition, although being two residues smaller, do not show inhibitory activity (Fig. 6). In the case of the inhibitory cystatins, the loop might be partially restrained in cystatin F as Cys37 likely is involved in a disulfide bridge (5), which can explain why cystatin F is a poorer inhibitor than cystatin C and E/M. The size and conformation of the loop could also be one reason why cystatin D does not inhibit legumain, because of an amino acid insertion in this loop (Fig. 4). For the type 3 cystatin studied, human L-kininogen, the lack of legumain-inhibitory activity may be due to steric reasons, as both legumain and kininogen are bulky molecules. Two of the three cystatin domains of kininogens are clearly able to inhibit papain-like peptidases (55), which demonstrates that the papain-binding surfaces of these domains are exposed and accessible to protein interactions. Whether the kininogen structure is sufficiently flexible to also allow exposure of the back-side loops on the opposite sides of these domains is presently unclear, as a three-dimensional model for type 3 cystatins is unfortunately not yet available. For the individual kininogen domains, the sequence requirements for a legumain-binding back-side loop suggested above seem to be fulfilled for domain 3, but not for domain 2 of human kininogen. Clearly, more studies are needed to clarify whether perhaps some variants of low or high Mr kininogens, resulting from proteolytic cleavages to release the kinin portion or individual cystatin domains of the protein, display legumain-inhibitory activity.

Although our initial studies indicate that the back-side loop around Asn39 is important for the ability of some cystatins to efficiently inhibit legumain, other cystatin segments may also be involved in interactions with the enzyme, just as several segments are involved in the cystatin inhibition of papain.
presentation of certain antigens to the immune system, but no efficient inhibitors to this activity are presently known (18), continued studies to elucidate and explore the mechanism of legumain inhibition by the novel cystatin site may prove valuable.

Acknowledgments—We gratefully acknowledge the skilled technical assistance of Anne-Cathrine Lüsström, Inger Nilsson, and Lorraine Smith. We thank Drs. Ingemar Björk and Anders Grubb for support and gifts of cystatin preparations used to generate the data presented in Table II.

REFERENCES

1. Abrahamson, M. (1994) Methods Enzymol. 244, 685–700
2. Rawlings, N. D., and Barrett, A. J. (1996) J. Mol. Biol. 30, 60–71
3. Schnittger, S., Rao, V. V., Abrahamson, M., and Hansmann, I. (1993) Genomics 16, 50–55
4. Ni, J., Abrahamsson, M., Zhang, M., Fernandez, M. A., Grubb, A., Su, J., Yu, G.-L., Li, Y., Parmelee, D., Xing, L., Coleman, T. A., Mentz, S., Tsubakura, R., Nguyen, N., Hesselberg, M., and Gentz, R. (1997) J. Biol. Chem. 272, 10853–10858
5. Ni, J., Fernandez, M. A., Danielsson, L., Chilakkuru, R. A., Zhang, J., Grubb, A., Su, J., Gentz, R., and Abrahamsson, M. (1998) J. Biol. Chem. 273, 24797–24804
6. Sotiropoulou, G., Anisowicz, A., and Sager, R. (1997) J. Biol. Chem. 272, 963–969
7. Halfon, S., Ford, J., Foster, J., Dowling, L., Lucian, L., Sterling, M., Xu, Y., Weiss, M., Ikeda, M., Liggett, D., Helms, A., Caux, C., Lebecque, S., Honnoum, C., Menon, S., McClanahan, T., Gorman, D., and Zurawski, G. (1998) J. Biol. Chem. 273, 16400–16408
8. Bode, W., Engh, R., Musil, D., Thiele, U., Huber, R., Karashik, A., Brzin, J., Kos, J., and Turk, V. (1988) EMBO J. 7, 2593–2596
9. Abrahamsson, M., Ritonja, A., Brown, M. A., Grubb, A., Machleidt, W., and Barrett, A. J. (1987) J. Biol. Chem. 262, 9688–9694
10. Stubbs, M. T., Laber, B., Bode, W., Huber, R., Jeralda, R., Lenarcic, B., and Turk, V. (1990) EMBO J. 9, 1939–1947
11. Kembhavi, A. A., Buttle, D. J., Knight, C. G., and Barrett, A. J. (1993) Arch. Biochem. Biophys. 303, 208–213
12. Hara-Nishimura, I. (1998) in Handbook of Proteolytic Enzymes (Barrett, A. J., Rawlings, N. D., and Woessner, J. F., eds) pp. 746–749, Academic Press, London
13. Dalton, J. F., and Brindley, P. J. (1998) in Handbook of Proteolytic Enzymes (Barrett, A. J., Rawlings, N. D., and Woessner, J. F., eds) pp. 749–754, Academic Press, London
14. Min, W., and Jones, H. (1994) Struct. Biology 1, 502–504
15. Chen, J.-M., Dando, P. M., Rawlings, N. D., Stevens, R. A. W., and Woessner, J. F., eds) pp. 749–754, Academic Press, London
16. Chau, C. T., Laurell, C. B., and Franzen, B. (1979) Anal. Biochem. 94, 311–317
17. Abrahamsson, M., Grubb, A., Olafsson, I., and Lundwall, A. (1987) FEBS Lett. 216, 229–233
18. Abrahamsson, M., Barrett, A. J., Salvesen, G., and Grubb, A. (1986) J. Biol. Chem. 261, 11262–11269
19. Blumberg, S., Schechter, I., and Berger, A. (1970) Eur. J. Biochem. 15, 97–102
20. Hall, A., Hakansson, K., Mason, R. W., and Grubb, A. (1995) J. Biol. Chem. 270, 5115–5121
21. Asthanat, A., Brown, M. A., Kembhavi, A. A., Nicklin, M. J., Sayers, C. A., Sanner, D. C., and Barrett, A. J. (1993) Biochem. J. 311, 129–138
22. Abrahamsson, M., Mason, R. W., Hansson, H., Buttle, D. J., Grubb, A., and Ohlsson, K. (1991) Biochem. J. 273, 621–626
23. Garnbret, L., Ekiel, I., and Abrahamsson, M. (1998) Biochemistry 37, 17309–17317
24. Abrahamsson, M., Grubb, A., Olafsson, I., and Lundwall, A. (1987) FEBS Lett. 216, 229–233
25. Daiu, H., Jensen, E. B., Tetttrup, H., Grubb, A., Abrahamsson, M., Olafsson, I., and Carlsen, S. (1989) Gene (Amst.) 79, 325–332
26. Jeppsson, J. O., Laurell, C. B., and Fransen, B. (1979) Clin. Chim. Acta 91, 373–383
27. Lindahl, P., Abrahamsson, M., and Björk, I. (1992) Biochem. J. 281, 49–55
28. Glaser, A. N., and Smith, E. L. (1971) in The Enzymes (Boyer, P., ed) pp. 501–546, Academic Press, New York
29. Schagger, H., and von Jagow, G. (1987) Anal. Biochem. 166, 368–379
30. Bjarnadottir, M., Wulff, B., Sameni, M., Sloane, B. F., Keppler, D., Grubb, A., and Abrahamsson, M. (1998) Mol. Pathol. 51, 317–326
31. Hendersen, P. J. (1972) Biochem. J. 127, 321–333
32. Devereux, J., Haeberli, P., and Smithies, O. (1984) Nucleic Acids Res. 12, 387–395
33. Turk, V., and Bode, W. (1991) FEBS Lett. 285, 213–219
Legumain Inhibition by Cystatins

40. Dieckmann, T., Mitschang, L., Hofmann, M., Kos, J., Turk, V., Auerswald, E. A., Jaenicke, R., and Oschkinat, H. (1993) J. Mol. Biol. 234, 1048–1059
41. Martin, J. R., Craven, C. J., Jerala, R., Kraus-Zitko, L., Zerovnik, E., Turk, V., and Waltho, J. P. (1995) J. Mol. Biol. 246, 331–343
42. Abola, E. E., Bernstein, F. C., Bryant, S. H., Koetzle, T. F., and Weng, J. (1987) in Crystallographic Databases: Information Content, Software Systems, Scientific Applications (Allen, F. H., Bergerhoff, G., and Sievers, R., eds) pp. 107–132, Data Commission of the International Union of Crystallography, Bonn
43. Abola, E. E., Sussman, J. L., Prilusky, J., and Manning, N. O. (1997) Methods Enzymol. 277, 556–571
44. Guex, N. (1996) Experientia (Basel) 52, A26
45. Guex, N., and Peitsch, M. C. (1996) Protein Data Bank Q. Newsletter 77, 7
46. Ekiel, I., and Abrahamson, M. (1996) J. Biol. Chem. 271, 1314–1321
47. Ekiel, I., Abrahamson, M., Fulton, D. B., Lindahl, P., Storer, A. C., Levadoux, W., Lafrange, M., Labelle, S., Pomerleau, Y., Groleau, D., LeSauteur, L., and Gehring, K. (1997) J. Mol. Biol. 271, 266–277
48. Grubb, A., and Lofberg, H. (1982) Proc. Natl. Acad. Sci. U. S. A. 79, 3024–3027
49. Ritonja, A., Machleidt, W., and Barrett, A. J. (1985) Biochem. Biophys. Res. Commun. 131, 1187–1192
50. Machleidt, W., Borchert, U., Fritz, H., Brzin, J., Ritonja, A., and Turk, V. (1983) Hoppe Seylers Z. Physiol. Chem. 364, 1481–1486
51. Ohkubo, I., Kurachi, K., Takasawa, T., Shiokawa, H., and Sasaki, M. (1984) Biochemistry 23, 5691–5697
52. Nicklin, M. J. H., and Barrett, A. J. (1984) Biochem. J. 223, 245–253
53. Auerswald, E. A., Genenger, G., Assaf-Machleidt, I., Machleidt, W., Engh, R. A., and Fritz, H. (1992) Eur. J. Biochem. 209, 837–845
54. Abe, Y., Shirane, K., Yokusawa, H., Matsushita, H., Mita, M., Kato, I., and Ishii, S. (1993) J. Biol. Chem. 268, 3525–3529
55. Sylvesen, G., Parkes, C., Abrahamson, M., Grubb, A., and Barrett, A. J. (1996) Biochem. J. 234, 429–434