Cystatin F Is a Glycosylated Human Low Molecular Weight Cysteine Proteinase Inhibitor*

(Received for publication, December 1, 1997, and in revised form, June 23, 1998)

Jian Niš, Marcia Alvarez Fernandez‡, Lena Danielsson¶, Rajeev A. Chillakuru‡, Junli Zhang‡, Anders Grubbs§, Jeffrey Su‡, Reiner Gentz‡, and Magnus Abrahamson§

From ‡Human Genome Sciences, Inc., Rockville, Maryland 20850-3338, the §Department of Clinical Chemistry, Institute of Laboratory Medicine, Lund University Hospital, S-221 85 Lund, Sweden, and the ¶Department of Immunotechnology, Lund University, S-220 07 Lund, Sweden

A previously undescribed human member of the cystatin superfamily called cystatin F has been identified by expressed sequence tag sequencing in human cDNA libraries. A full-length cDNA clone was obtained from a library made from mRNA of CD34-depleted cord blood cells. The sequence of the cDNA contained an open reading frame encoding a putative 9-residue signal peptide and a mature protein of 126 amino acids with two disulfide bridges and enzyme-binding motifs homologous to those of Family 2 cystatins. Unlike other human cystatins, cystatin F has 2 additional Cys residues, indicating the presence of an extra disulfide bridge stabilizing the N-terminal region of the molecule. Recombinant cystatin F was produced in a baculovirus expression system and characterized. The mature recombinant protein processed by insect cells had an N-terminal segment 7 residues longer than that of cystatin C and displayed reversible inhibition of papain and cathepsin L (Kᵢ = 1.1 and 0.31 nM, respectively), but not cathepsin B. Like cystatin E/M, cystatin F is a glycoprotein, carrying two N-linked carbohydrate chains at positions 36 and 88. An immunoblot for quantification of cystatin F showed that blood contains low levels of the inhibitor (0.9 ng/ml). Six B cell lines in culture secreted barely detectable amounts of cystatin F, but several T cell lines and especially one myeloid cell line secreted significant amounts of the inhibitor. Northern blot analysis revealed that the cystatin F gene is primarily expressed in peripheral blood cells and spleen. Tissue expression clearly different from that of the ubiquitous inhibitor, cystatin C, was also indicated by a high incidence of cystatin F clones in cDNA libraries from dendritic and T cells, but no clones identified by expressed sequence tag sequencing in several B cell libraries and in >600 libraries from other human tissues and cells.

Cysteine proteinase inhibitors of the cystatin superfamily are ubiquitous in the body and are generally tight-binding inhibitors of papain-like cysteine proteinases, such as cathespins B, H, L, S, and K (for review, see Ref. 1). They should therefore serve a protective function to regulate the activities of such endogenous proteinases, which otherwise may cause uncontrolled proteolysis and tissue damage. Cysteine proteinase activity can not normally be measured in body fluids, but can be detected extracellularly in conditions like endotoxin-induced sepsis (2) and metastasizing cancer (3) and at local inflammatory processes such as in rheumatoid arthritis (4), purulent bronchiectasis (5), and periodontitis (6), which indicates that a tight enzyme regulation by cystatins is a necessity in the normal state. A deficiency state in which the levels of the intracellular cystatin, cystatin B, are lowered due to mutations has recently been shown to segregate with a form of progressive myoclonus epilepsy (7), which points to additional specialized functions of cystatins. Moreover, cell culture results showing that chicken cystatin, human cystatin C, and human cystatin D inhibit the replication of polio (8), herpes simplex (9) and corona (10) viruses, respectively, and that human cystatin A inhibits rhadovirus-induced apoptosis (11) indicate that cystatins play additional roles in the human defense system. The cystatins constitute a superfamily of evolutionary related proteins, all composed of at least one 100–120-amino acid residues domain with conserved sequence motifs (12). The previously well characterized single-domain human members of this superfamily could be grouped in two protein families. The Family 1 members, cystatins (or stefins) A and B, contain ~100 amino acid residues, lack disulfide bridges, and are not synthesized as preproteins with signal peptides. The Family 2 cystatins, cystatins C, D, S, SN, and SA, are secreted proteins of ~120 amino acid residues (Mr, 13,000–14,000) and have two characteristic intrachain disulfide bonds. Recently, we identified an additional human cystatin superfamily member by EST1 sequencing in epithelial cell-derived cDNA libraries that we named cystatin E (13). The same cystatin was independently discovered by differential display experiments as an mRNA species down-regulated in breast tumor tissue, but present in the surrounding epithelium and reported under the name cystatin M (14). Cystatin E/M is an atypical secreted low Mr cystatin in that it is a glycoprotein and shows only 30–35% sequence identity in alignments with the human Family 2 cystatins, which demonstrates that additional cystatin families are yet to be identified (13). The cystatin E/M gene has been localized to chromosome 11 (15), whereas all human Family 2 cystatin genes are clustered on the short arm of chromosome 20 (16), which further stresses that cystatin E/M is only distantly related to the other secreted human low Mr cystatins.

In this investigation, we continued our search for novel human proteins with distant relationship to the cystatin super-
family. We report the characteristics of a cystatin identified in cDNA libraries from immune cells, cystatin F, which is only 30–34% homologous in overall sequence to the human Family 2 cystatins, but also only 29% homologous to cystatin E/M. Like cystatin E/M, cystatin F is a glycoprotein and a potent cysteine proteinase inhibitor.

**EXPERIMENTAL PROCEDURES**

**Identification of cDNAs Encoding Cystatin F**—A database containing >1,000,000 ESTs obtained from >850 different cDNA libraries has been generated by the Human Genome Sciences, Inc. and the Institute for Genomic Research using high-throughput automated DNA sequencing. Gene analysis of randomly selected human cDNA clones (17, 18). Sequence homology comparisons of each EST were performed against the GenBank™ Data Bank using the BLAST and BLASTN algorithms (19). ESTs having homology to previously identified sequences (p = 0.01) and motif search using the known amino acid sequence of human cystatin C (M27891) against this database revealed several ESTs having translated sequences >30% homologous to that of cystatin C. One clone (HCUDE60) encoding an intact N-terminal signal peptide was identified in a human CD34-depleted cord blood cell library and selected for further investigation. The complete cDNA sequence of both strands of this clone was determined by sequencing the cDNA to the clone end.

**Baculovirus Expression for Production of Cystatin F**—The entire coding sequence of the cystatin F cDNA was amplified using standard polymerase chain reaction techniques with primers corresponding to the 5′- and 3′-sequences of the gene (upstream primer (with tailing BamHI site underlined), 5′-CGC GGA TCC GCG ATG CCG GCT-GGA-3′; and downstream primer (with Asp718 site), 5′-CGG GCC AGG GCG GCG TCA-3′). The amplified fragment was purified, digested with BamHI and Asp718, and again purified. The baculovirus expression vector pA2, derived from pNR704 (20, 21), was digested with BamHI and Asp718, followed by agarose gel purification. The open and purified pA2 vector was ligated with the amplified cystatin cDNA sequence encoding the T4 DNA ligase (Life Technologies, Inc.).

**Recombinant baculoviruses encoding cystatin F** were generated by co-transfection of SF9 cells with 5 μg of transfer vector containing the HCUDE60 cDNA and 1 μg of BaculGold viral DNA (Pharmingen) using Lipofectin (Life Technologies, Inc.). Details on screening and cloning of the recombinant baculoviruses have been reported earlier (13).

Recombinant cystatin F was purified from 10 liters of SF9 cell supernatant, after infection of the insect cells, using an IC Cel401 medium (JRH Biosciences) with 1% (v/v) fetal bovine serum. The harvested supernatants were pooled and clarified by centrifugation at 18,000 × g in a continuous flow centrifuge. The supernatant was loaded onto a strong cation-exchange column (Poros HS50, PerSeptive Biosystems) at pH 6.2. The column had previously been equilibrated in 20 mM sodium acetate buffer, pH 6.0, containing 100 mM NaCl, and then loaded onto a weak cation-exchange column (Poros CM20, PerSeptive Biosystems). The fractions containing cystatin F were identified by SDS-PAGE analysis and pooled for the next step. The pooled fractions were dialyzed against 40 mM Tris-HCl buffer, pH 8.0, containing 100 mM NaCl. After loading the supernatant onto the HS50 column, the column was washed with the equilibration buffer, followed by successive elutions with the same buffer containing 250, 400, 600, and 1000 mM NaCl. The recombinant cystatin F eluted in the 400 mM NaCl fraction. This fraction was then loaded onto a size-exclusion chromatography column (Superdex S-75, Amersham Pharmacia Biotech), which had previously been equilibrated in 20 mM sodium acetate buffer, pH 5.5, containing 100 mM NaCl. The fractions containing cystatin F were identified by SDS-PAGE analysis and pooled for the next step. The pooled fractions were dialyzed against 40 mM Tris-HCl buffer, pH 8.0, containing 100 mM NaCl. This pool was then loaded onto a pre-equilibrated strong anion-exchange column (Poros HQ50, PerSeptive Biosystems) at pH 6.2. The column had previously been equilibrated in 20 mM sodium acetate buffer, pH 5.5, containing 100 mM NaCl and then loaded onto a weak cation-exchange column (Poros CM20, PerSeptive Biosystems). The fractions containing cystatin F were eluted using a linear salt gradient (from 100 to 1000 mM NaCl in 20 mM sodium acetate buffer, pH 6.0). The eluted fractions containing purified cystatin F (at 400 mM NaCl) were identified by SDS-PAGE and verified by N-terminal amino acid sequencing (22). The protein concentration was checked on a Model Q10 10FPLC column (Amersham Pharmacia Biotech) equilibrated in 50 mM ethanolamine buffer, pH 9.0, and eluted with a linear gradient of 0–500 mM NaCl in the same buffer. Gel filtration was performed on a Superdex 75 FPLC column (Amersham Pharmacia Biotech) equilibrated in 50 mM ammonium bicarbonate buffer, pH 8.0, at a flow rate of 0.5 ml/min.

**Protein Analyses**—Glycosylation analysis of cystatin F was performed using 0.2 mg of isolated recombinant antigen (see above) in Freund’s complete adjuvant (Difco) subcutaneously into a rabbit. The injection was repeated after 3 weeks, and the rabbit was bled every third week. The specificity of the antiserum was tested by immunoelectrophoresis of the recombinant cystatin F used as antigen; of concentrated protein- containing urine containing cystatins A, B, C, S, SN, and E and kininogen (23); and of recombinant cystatins C, D, and E (13, 24, 25). The IgG fraction of 50 ml of antiserum was isolated by absorption to protein A-Sepharose (Amersham Pharmacia Biotech) and subsequent elution with 0.1 M glycine buffer, pH 2.2. The eluate was immediately neutralized by addition of 2 M Tris buffer, pH 7.4.

**Antiserum Production and Construction of an ELISA for Cystatin F**—Quantification—an antisem against cystatin F was raised by injecting the rabbit urine containing cystatins A, B, C, S, SN, and E and kininogen (23); and the specificity of the antiserum was tested by immunoelectrophoresis of the recombinant cystatin F used as antigen; of concentrated protein-containing urine containing cystatins A, B, C, S, SN, and E and kininogen (23); and of recombinant cystatins C, D, and E (13, 24, 25). The IgG fraction of 50 ml of antiserum was isolated by absorption to protein A-Sepharose (Amersham Pharmacia Biotech) and subsequent elution with 0.1 M glycine buffer, pH 2.2. The eluate was immediately neutralized by addition of 2 M Tris buffer, pH 7.4.

**Identification of cDNAs Encoding Cystatin F**—A database containing >1,000,000 ESTs obtained from >850 different cDNA libraries has been generated by the Human Genome Sciences, Inc. and the Institute for Genomic Research using high-throughput automated DNA sequencing. Gene analysis of randomly selected human cDNA clones (17, 18). Sequence homology comparisons of each EST were performed against the GenBank™ Data Bank using the BLAST and BLASTN algorithms (19). ESTs having homology to previously identified sequences (p = 0.01) were selected for further investigation. The complete cDNA sequence of both strands of this clone was determined by sequencing the cDNA to the clone end.

**Baculovirus Expression for Production of Cystatin F**—The entire coding sequence of the cystatin F cDNA was amplified using standard polymerase chain reaction techniques with primers corresponding to the 5′- and 3′-sequences of the gene (upstream primer (with tailing BamHI site underlined), 5′-CGC GGA TCC GCG ATG CCG GCT-GGA-3′; and downstream primer (with Asp718 site), 5′-CGG GCC AGG GCG GCG TCA-3′). The amplified fragment was purified, digested with BamHI and Asp718, and again purified. The baculovirus expression vector pA2, derived from pNR704 (20, 21), was digested with BamHI and Asp718, followed by agarose gel purification. The open and purified pA2 vector was ligated with the amplified cystatin cDNA sequence encoding the T4 DNA ligase (Life Technologies, Inc.). Details on screening and cloning of the recombinant baculoviruses have been reported earlier (13).

Recombinant cystatin F was purified from 10 liters of SF9 cell supernatant after infection of the insect cells, using an IC Cel401 medium (JRH Biosciences) with 1% (v/v) fetal bovine serum. The harvested supernatants were pooled and clarified by centrifugation at 18,000 × g in a continuous flow centrifuge. The supernatant was loaded onto a strong cation-exchange column (Poros HS50, PerSeptive Biosystems) at pH 6.2. The column had previously been equilibrated in 20 mM sodium acetate buffer, pH 5.5, containing 100 mM NaCl. After loading the supernatant onto the HS50 column, the column was washed with the equilibration buffer, followed by successive elutions with the same buffer containing 250, 400, 600, and 1000 mM NaCl. The recombinant cystatin F eluted in the 400 mM NaCl fraction. This fraction was then loaded onto a size-exclusion chromatography column (Superdex S-75, Amersham Pharmacia Biotech), which had previously been equilibrated in 20 mM sodium acetate buffer, pH 5.5, containing 100 mM NaCl. The fractions containing cystatin F were identified by SDS-PAGE analysis and pooled for the next step. The pooled fractions were dialyzed against 40 mM Tris-HCl buffer, pH 8.0, containing 100 mM NaCl. This pool was then loaded onto a pre-equilibrated strong anion-exchange chromatography column (Poros HSQ50, PerSeptive Biosystems), and the flow-through fraction was collected. The flow-through pool from the HSQ50 column was finally re-equilibrated in 20 mM sodium acetate buffer, pH 6.0, containing 100 mM NaCl and then loaded onto a weak cation-exchange column (Poros CM20, PerSeptive Biosystems). The fractions containing cystatin F were eluted using a linear salt gradient (from 100 to 1000 mM NaCl in 20 mM sodium acetate buffer, pH 6.0). The eluted fractions containing purified cystatin F (~400 mM NaCl) were identified by SDS-PAGE and verified by N-terminal amino acid sequencing (22). The protein concentration was checked on a Model Q10 10FPLC column (Amersham Pharmacia Biotech) equilibrated in 50 mM ethanolamine buffer, pH 9.0, and eluted with a linear gradient of 0–500 mM NaCl in the same buffer. Gel filtration was performed on a Superdex 75 FPLC column (Amersham Pharmacia Biotech) equilibrated in 50 mM ammonium bicarbonate buffer, pH 8.0, at a flow rate of 0.5 ml/min.
formed by determining the monosaccharide content in a purified preparation of the recombinant protein. Approximately 10 μg of the protein was hydrolyzed with 0.2 M trifluoroacetic acid at 100 °C for 4 h. The hydrolysate was dried in a SpeedVac and reconstituted in 50 μl of deionized water, after which monosaccharides were analyzed on a Dionex PA-1 column used to separate the monosaccharides by isocratic elution with 12 mM NaOH. The monosaccharides were detected by integrated amperometry. Glycosylation analysis was also done by incubation with peptide N-glycosidase F (EC 3.2.2.18) under conditions recommended by the enzyme supplier (Oxford GlycoSystems, Abingdon, United Kingdom), followed by 16.5% SDS-PAGE and silver staining. N-terminal sequencing using an ABI492 sequencer (Applied Biosystems, Inc.) was carried out after electrophoresis on 4–20% SDS-polyacrylamide gels (Novex) using the buffer system by Laemmli (37), blotting onto a ProBlott membrane (Applied Biosystems, Inc.), staining with Ponceau S (0.2% in 4% acetic acid), and excision of the band of interest. Amino acid sequencing using an ABI492 sequencer (Applied Biosystems, Inc.) was performed on the Superdex 75 column run as described above, with the elution volumes for apropin (M = 6500), cytochrome c (M = 12400), chymotrypsinogen (M = 23400), and β-lactoglobulin (M = 35000) used for construction of a calibration curve. To analyze the content of free thiol groups in recombinant cystatin F, the Ellman reaction with 5,5'-dithiobis-2-nitrobenzoic acid (Sigma) was performed under conditions described earlier (38). One sample of isolated recombinant cystatin F was analyzed directly after purification (see above). A second sample of cystatin F and a sample of papain, used as a positive reaction control, was both analyzed after mild reduction to reduce oxidized sulphydryl groups. Papain (17.5 mg; Sigma type III) and cystatin F (0.36 mg) was incubated in 100 mM sodium phosphate buffer, pH 6.5, containing 1 mM dithiothreitol for 10 min. Removal of excess dithiothreitol was accomplished by chromatography on a HighTrap desalting column (Amersham Pharmacia Biotech) in 40 mM sodium phosphate buffer, pH 6.5. Portions of the desalted samples were directly analyzed for free thiols by the Ellman assay (22), respectively. The proportion of active enzyme in the papain sample was measured by E-64 titration (see below).

**Enzyme Assays—** The methods used for active-site titration of papain (using benzoyl-Arg-p-nitroanilide (Bachem Feinchemikalien, Buben-dorf, Switzerland) as substrate), titration of the molar enzyme inhibitory capacity in cystatin F preparations, and determination of equilibrium constants for dissociation (Kd) of complexes between cystatin F and cysteine proteinases have been reviewed (1). The enzymes used were papain (EC 3.4.22.2; from Sigma), activatable to 70–75% after affinity purification on Sepharose-coupled Gly-Gly-Tyr-Arg as detailed previously (39, 40), human cathepsin B (EC 3.4.22.1; from Calbiochem), the fluorogenic substrate used for Kd determinations was benzylxycarbonyl-Phe-Arg-7-amido-4-methylcoumarin (10 μM; Bachem Feinchemikalien), and the assay buffer was 100 mM sodium phosphate buffer (adjusted to pH 6.5 for papain and to pH 6.0 for cathepsin B and cathepsin L) containing 1 mM dithiothreitol and 2 mM EDTA. Steady-state velocities were measured before and after addition of cystatin F in assays at 37 °C, and Kd values were calculated according to Henderson (41). Corrections for substrate competition were made using Ke values determined for the substrate batch used under the assay conditions employed (60, 55, and 3.2 μM for papain, cathepsin B, and cathepsin L, respectively).

**RESULTS AND DISCUSSION**

**Discovery of a Novel Human Cystatin—** On analysis of EST sequences obtained from human cDNA libraries, several clones encoding polypeptides with low but significant homology (30–35%) to the cystatin C sequence (42) were identified. Some of these encoded cystatin E/M (13), but none encoded another secreted cystatin-related protein. A full-length human cDNA clone in the latter group (designated HCUD60) was identified in a library from CD34-depleted cord blood cells. The 727-base pair cDNA included an open reading frame encoding a 145-residue preprotein (Fig. 1), of which the first 19 theoretically (43) should constitute the signal peptide. The cDNA contained a typical consensus sequence for initiation of translation (44) around the start ATG codon of the open reading frame and a normal polyadenylation signal (ATTAAA) at a distance of 18 nucleotides from a poly(A) stretch in the 3'-end. An alignment of the deduced 126-residue mature protein sequence with the six known secreted single-domain human cystatins demonstrated that the novel protein was most similar to cystatin C of the Family 2 cystatins (34% identical residues) and a bit more distantly related to cystatins D, S, SN, and SA (30–32% identity) (Fig. 2). The resemblance to the recently reported atypical secreted recombinant protein is shown by an arrow. Residues involved in cysteine binding for other cystatins are boxed. The Aaa residues of two theoretical N-glycosylation sites (Asn-Xaa-(Ser/Thr)) are marked with asterisks.

**Human Cystatin F**

**Fig. 1.** Nucleotide and deduced amino acid sequences of clone HCUD60 cDNA encoding human pre-cystatin F. Numbering of the nucleotide sequence begins at the first nucleotide of the HCUDE60 cDNA. The oligonucleotides used as polymerase chain reaction primers to generate a fragment for ligation into a baculovirus expression vector are shown at the positions they hybridize to the cDNA. The polyadenylation signal is double-underlined. Amino acids of the translation product are given in one-letter code below the first nucleotide of each codon; the numbering of these starts with the initiator methionine as residue 1. The cleavage generating the N terminus of the insect cell-secreted recombinant protein is shown by an arrow. Residue involved in cysteine binding for other cystatins are boxed. The Aaa residues of two theoretical N-glycosylation sites (Asn-Xaa-(Ser/Thr)) are marked with asterisks.
Family 2 cystatins (32% identical residues). The sequence contained a central Gln-Xaa-Val-Xaa-Gly motif, a typical Gly residue in the N-terminal segment 44 positions earlier, and also a Pro-Trp pair toward the C-terminal end of the translation product, like the sequences of the human Family 2 or 3 cystatins and cystatin E/M. The sequence also contained 4 Cys residues toward the C-terminal end, corresponding to those forming disulfide bridges in cystatin C (47) are marked with solid brackets. The 2 additional Cys residues in the cystatin F sequence likely forming a third disulfide bridge are shown with a dashed bracket. The N-glycosylation sites in the cystatin F and cystatin E/M sequences are marked with asterisks. The secondary structures forming segments in cystatin C (48) and chicken cystatin (49) are indicated above the sequences (α, α-helix; β1–5, strands of the β-sheet). B, schematic illustration of evolutionary relationships. The evolutionary relationships between all known inhibitory active human cystatin domains are indicated. The phylogenetic tree was constructed using GROWTREE included in the GCG software package (Version 8.1; Genetics Computer Group Inc., Madison, WI). The phylogenetic distances were obtained according to the method described by Kimura (50). The reconstruction of the tree was done by the unweighed pair group method using arithmetic averages.
extended N-terminal segment, being 6–10 residues longer than those of other single-domain cystatins. The presence of a Cys residue in this segment, and another not seen in other cystatins in position 37 (Fig. 2A), indicated the presence of a third disulfide bridge in addition to the two ones present in all known secretory cystatins of higher animals. Assuming a structure of cystatin F similar to those of human cystatin C and chicken cystatin (48, 49), Cys-37 would be located in a loop just after an α-helix-forming segment, on the side of the molecule opposite from the enzyme-binding site. According to molecular modeling, this residue could be in close contact with Cys-1, provided that the extended N-terminal segment of cystatin F is stretched from the proposed anchored Gly-11 residue included in the enzyme-binding site, along the α-helix on the surface of the molecule. The extra disulfide could thus function to stabilize the N-terminal segment, presumably in a conformation that would affect the specificity of target enzyme inhibition for cystatin F. A motif search, in addition, showed two target Asn-Xaa(Ser/Thr) sequences for glycosylation at positions 36–38 and 88–90. The first site would be located in the same surface loop as the likely disulfide-bonded Cys-37 residue. The second site lies in a segment that, assuming three-dimensional structure homology to chicken cystatin and human cystatin C, forms a surface loop close to the fourth strand of the β-sheet. As both of these loops are on the opposite side from the enzyme-binding region, attached carbohydrate chains would likely not affect the function of cystatin F.

Characterization of Recombinant Cystatin F—To allow studies of the properties of cystatin F, recombinant production of the protein was attempted. The cystatin F cDNA was subcloned in a baculovirus expression vector as described under “Experimental Procedures.” Recombinant cystatin F was secreted into the cell media of Sf9 cell cultures at a level of 10 mg/liter of such cultures (see “Experimental Procedures”), yielding approximately equal quantities by the Sf9 cells. Prolonged incubation (Fig. 1) also demonstrated that the recombinant protein was 3.8 nM, whereas that determined by quantitative amino acid analysis was 12 μM. Because the affinity-purified papain used for the titration experiments was active at 70% (according to a parallel titration using a reference solution of E-64; see “Experimental Procedures”) and cystatins also bind the fraction of papain that is inactive against peptidyl substrates (1), these results demonstrated that the recombinant protein

---

**Fig. 3. Cystatin F is a glycoprotein.** A time course experiment for deglycosylation of recombinant cystatin F by incubation with peptide N-glycosidase F is shown. Samples taken from the incubation mixture were analyzed by SDS-PAGE using a 16.5% gel. The gel was silver-stained. Lane 1, isolated recombinant cystatin F, with no peptide N-glycosidase F added; lanes 2–6, cystatin F incubated with peptide N-glycosidase F at 37°C for 0.5 min, 30 min, 1 h, 3 h, and 6 h, respectively. The positions of relevant molecular weight marker bands are indicated to the left.
Human Cystatin F

**FIG. 4.** Gel filtration of recombinant cystatin F and native human cystatin F produced by U-937 cells. Gel filtration was performed on a Superdex 75 column as described under “Experimental Procedures.” Spent medium from serum-free cultures of U-937 cells (100 ml) was concentrated 200 times by ultrafiltration and applied to the column. Cystatin F content in the fractions was measured by ELISA (dashed line). Approximately 0.3 mg of isolated recombinant cystatin F was gel-filtered the same way and monitored by absorbance at 280 nm (solid line). The elution volumes of isolated recombinant cystatin C and cystatin E/M chromatographed on the same column under identical conditions are indicated by arrows.

was 45% active. Equilibrium constants for cystatin F complexes with papain, human cathepsin B, and human cathepsin L were determined by steady-state measurements in fluorogenic enzyme assays to allow dilution of the enzymes to fluorometric molar concentrations; under such assay conditions, the cystatin dissociated significantly from formed enzyme complexes. The results (Table I) show that cystatin F is a relatively tight-binding inhibitor of papain, with a $K_i$ value for the complex similar to that for the cystatin D complex, but 5 orders of magnitude higher than that for the cystatin C complex. The cystatin F affinity for cathepsin L was somewhat higher ($K_i = 0.31 \text{ nM}$) than for papain and moreover significantly higher than the affinity of cystatin D for cathepsin L. Like cystatin D, however, cystatin F did not appreciably inhibit cathepsin B. This contrasts with the inhibitory specificities of cystatins C and E/M, which both have the capacity to inhibit cathepsin B (Table I). A structural element that has been shown to be at least partly responsible for the target enzyme specificity of cystatin C and of crucial importance for efficient cystatin C binding of cathepsin B is the N-terminal segment with Arg-Leu-Val-Gly at positions 8–11 (40, 53). Leu-9 and Val-10, interacting with target enzyme substrate subpockets S3 and S2, respectively, are especially important for a fast interaction between the cystatin and cathepsin B (54, 55). The corresponding N-terminal segment in cystatin F is Val-Lys-Pro-Gly. The presence of Pro-10 in the proposed P2 position would likely not be favorable for binding of the N-terminal segment to cathepsin B, nor would a charged Lys-9 residue in the P3 position be beneficial, thus likely explaining the lack of cathepsin B inhibitory activity seen for cystatin F.

**Identification and Distribution of Cystatin F in Human Tissue**—To identify cystatin F in human samples, an ELISA for quantification of the cystatin was constructed using a polyclonal antiserum against recombinant cystatin F and cystatin F purified by amino acid analysis for construction of the standard curve (see “Experimental Procedures”). The ELISA showed a sensitivity of 0.5 ng/ml (defined as 2 S.D. values of blank readings), and the linear part of the standard curve allowed us to measure concentrations up to 30 ng/ml. The assay should thus be sufficiently sensitive to measure physiologically relevant concentrations of cystatin F in human samples. ELISA measurements demonstrated that blood (plasma and serum) contained cystatin F in low amounts. The cystatin F concentration in pooled normal serum was 0.89 ng/ml (mean value from three measurements). The concentration of cystatin F in urine from patients with tubular proteinuria, which has earlier been shown to contain significant amounts of the single-domain cystatins A, B, C, E, S, and SN and has proven a good starting material for purification of human cystatins (13), was below the detection limit, however. Measurement of bovine serum samples demonstrated that the ELISA cross-reacted with neither a putative bovine cystatin F-like inhibitor nor other bovine cystatins. Immunoblotting using the antiserum against cystatin F after separation of blood plasma proteins according to size (SDS-PAGE) or charge (agarose gel electrophoresis) gave a negative result. This was not unexpected, given the low blood concentration of the protein according to ELISA, which should mean that $<1 \text{ ng}$ of cystatin F would be present in the 30 µl that was the maximum volume that could be electrophoresed per lane. Fractionation of blood plasma by ammonium sulfate precipitation followed by concentration of the fractions and immunoblotting after SDS-PAGE separation indicated that cystatin F precipitated already in 25% saturated ammonium sulfate solution. The dissolved precipitate obtained in 25% saturated ammonium sulfate solution also contained high amounts of immunoglobulins. The immunoglobulin light chains co-eluted with cystatin F in attempts to further purify the cystatin F by ion-exchange chromatography or gel filtration according to immunoblotting of such fractions (data not shown).

Because the cystatin F concentration of $<1 \mu g$ in a background of some 70 g of other proteins/liter made blood plasma a far from ideal starting material for purification of natural cystatin F, we instead investigated human blood cell lines in culture for cystatin F secretion. Quantification of cystatin F in spent culture media from 10 cell lines (Table II) revealed that B cell lines generally did not produce appreciable amounts of cystatin F, but some T cell lines and one out of two monocyte-related cell lines secreted the inhibitor in significant amounts. The highest cystatin F production was seen for U-937 cells, which secreted 2.0 ng/10^6 cells/72 h, an amount equaling 10% of the cystatin C secretion from the same cells. Serum-free culturing of MOLT-4, Jurkat, and U-937 cells was attempted. The ELISA showed that conditioned medium from such cultures also contained cystatin F (Table II). The cystatin F immunoreactivity co-eluted with recombinant cystatin F in ion-exchange chromatography on Mono-Q and in gel filtration as demonstrated by ELISA measurements of fractions from such chromatographies (Fig. 4), but the amounts present in the fractions after attempts to start purification from 1-liter scale serum-free cultures of U-937 cells did not allow purification of the natural cystatin F to homogeneity and were not sufficient for N-terminal sequencing from electrophoresis blots. What the results clearly demonstrated, however, was that the cystatin F

| Cystatin | Papain | Cathepsin L | Cathepsin B |
|----------|--------|------------|------------|
| $K_i$    | $n_m$  | $n_m$      | $n_m$      |
| F        | 1.1    | ±0.26      | 0.31       | ±0.17      | 11| >1000 |
| C        | 0.000014 | <0.005     | 0.27       |
| D        | 0.9    | 18         | >1000      |
| E/M      | 0.39   | ND*        | 32         |

ND, not determined.

**TABLE I**

Inhibition of cysteine proteases by cystatin F

Equilibrium constants for dissociation of cystatin F complexes with papain and human cathepsins B and L were determined by steady-state kinetics. The corresponding values for cystatins C, D, and E/M, determined by similar methods (24, 51, 52), are shown for comparison. The 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.”
measured by ELISA was not a result of cross-reactions with other cystatins, as they elute in significantly lower
C or cystatin E/M mRNA (13, 56). The strongest mRNA signals
were seen for spleen and peripheral blood leukocytes; moderate
signals were observed for thymus and small intestine; and
apoptotic T cells, osteosarcoma cells, pancreatic islet tumor
cells, thymus, Jurkat cells, helper T cells (Th2), and peripheral
blood mononuclear cells stimulated with poly(I-C) (interferon
inducer). To put the distribution pattern of cystatin F clones in
the cDNA data bases in perspective, the fact that 51 of the 54
cystatin F clones were found in libraries directly related to the
immune system could be compared with the distribution pat-
tern of cystatin C clones in the same libraries. A total number
of 539 cystatin C clones were identified in 163 of the >650
libraries, without any obvious bias toward a certain tissue or
cell type. The combined results from the analyses of cystatin F
mRNA by Northern blotting and EST data base screening thus
suggest that cystatin F is a cysteine proteinase inhibitor pri-
marily produced by cells of the immune system, possibly with a
function crucial for specialized cells of the myeloid or T cell
lineages. In this context, it is noteworthy that a novel papain-
like cysteine proteinase, cathepsin W, with specific expression
in cytotoxic T lymphocytes has recently been reported (57).

Acknowledgments—The skilled technical assistance of Anne-Cath-
lene Löstfröm and Inger Nilsson is gratefully acknowledged. We thank
the Human Genome Sciences, Inc. Sequencing Group and Department
of Protein Development for assistance.

REFERENCES
1. Abrahamson, M. (1994) Methods Enzymol. 244, 685–700
2. Asfalg-Machleidt, I., Jochum, M., Klaubert, W., Inthorn, D., and Machleidt,
W. (1988) Biol. Chem. Hoppe-Seyler 369, suppl. 263–269
3. Köppel, P., Baici, A., Keist, R., Matzku, S., and Keller, R. (1994) Exp. Cell Biol.
32, 283–298
5. Battile, D. J., Burnett, D., and Abrahamson, M. (1996) Scand. J. Clin. Lab.
Invest. 56, 509–516
6. Cox, S. W., and Eley, B. M. (1989) J. Periodontal Res. 24, 353–361
7. Pennecchio, I. A., Leheshjoki, A. E., Stone, N. E., Willour, V. L., Virtanen, K.
Miao, J., D’Amato, R., Ramirez, L., Faham, S., Koskinen, M., Kukkonen, M.,
Warrington, J. A., Norio, R., de la Chapelle, A., Cox, D. R., and Myers, R. M.
(1990) Science 271, 1731–1734
8. Korant, B. D., Brizin, J., and Turk, V. (1985) Biochem. Biophys. Res. Commun.
127, 1072–1076
9. Björk, L., Grubb, A., and Kjellén, L. (1990) J. Virol. 64, 941–943
10. Collins, A. R., and Grubb, A. (1998) Oral Microbiol. Immunol. 13, in press
11. Björklund, H. V., Johansson, T. R., and Rinne, A. (1997) J. Virol. 71, 5658–5662
12. Rawlings, N. D., and Barrett, A. J. (1990) J. Mol. Evol. 30, 60–71
13. Ni, J., Abrahamson, M., Zhang, M., Fernandez, M. A., Grubb, A., Su, J., Yu,
G.-L., Li, Y., Parmelee, D., Xing, L., Coleman, T. A., Gentz, S., Thotakura,
R., Nguyen, N., Hesselberg, M., and Gentz, R. (1997) J. Biol. Chem. 272,
10853–10858
