Leukocystatin, A New Class II Cystatin Expressed Selectively by Hematopoietic Cells*

(Received for publication, November 5, 1997, and in revised form, April 6, 1998)

Sherin Halfon, John Ford, Jessica Foster, Lynette Dowling, Linda Luciani, Marissa Sterling, Yuming Xu, Mary Weiss, Mami Ikeda, Debra Liggett, Allison Helms, Christopher Caux, Serge Lebecque, Chuck Hannum, Satish Menon, Terrill McClanahan, Daniel Gorman, and Gerard Zurawski†

From the Departments of Molecular Biology and Immunobiology, DNAX Research Institute, Palo Alto, California 94304-1104 and §Laboratory of Immunological Research, 27 Chemin des Peupliers, B.P. 11, 69572 Dardilly, France

We describe a new cystatin in both mice and humans, which we termed leukocystatin. This protein has all the features of a Class II secreted inhibitory cystatin but contains lysine residues in the normally hydrophobic binding regions. As determined by cDNA library Southern blots, this cystatin is expressed selectively in hematopoietic cells, although fine details of the distribution among these cell types differ between the human and mouse mRNAs. In addition, we have determined the genomic organization of mouse leukocystatin, and we found that in contrast to most cystatins, the leukocystatin gene contains three introns. The recombinant proteins corresponding to these cystatins were expressed in E. coli as N-terminal glutathione S-transferase or FLAGTM fusions, and studies showed that they inhibited papain and cathepsin L but with affinities lower than other cystatins. The unique features of leukocystatin suggest that this cystatin plays a role in immune regulation through inhibition of a unique target in the hematopoietic system.

Cysteine proteases play many very important roles in the immune system. For instance, the de-ubiquinating enzymes are cysteine proteases, whereas lysosomal proteases are involved in antigen presentation both through the degradation of proteins to antigenic peptides and by processing the invariant chain of class II major histocompatibility complexes (1). However, the overexpression of these proteases can be detrimental to cells, as can their release into the extracellular space. Therefore, their activities in these cells are controlled by a variety of mechanisms, including the presence of macromolecular protease inhibitors.

The cystatins make up a class of very tight, reversible, competitive inhibitors of the papain family of cysteine proteases. Cystatins have been divided into four classes based upon their sequence and genomic structures within each family are highly conserved. Cystatins are expressed throughout the body in a tissue-specific manner. Mutations in some cystatins or alterations in the balance of these with their cognate cysteine proteases have been implicated in several diseases (2, 3). Many studies, involving changes of peptide sequence, have shown that three regions of the cystatin, which form a “wedge” that can associate with the active-site cleft, are all required for tight binding to the protease. These studies have been confirmed by the crystal structure of the cystatin B-papain complex (4) and supported by other structural studies showing that chicken egg white cystatin, a Class II cystatin, has the same fold as the Class I cystatin B (5, 6).

In this paper, we describe the characterization of a new Class II cystatin, leukocystatin, specifically expressed by hematopoietic cells. The unique features of the amino acid sequence suggest that the as yet unidentified target protease is not one of the commonly studied lysosomal cysteine proteases, although leukocystatin is an active inhibitor of these cathepsins.

In addition, the unusual genomic structure of the mouse protein and the amino acid sequences of both the human and mouse inhibitors suggest that they are quite divergent from other Class II cystatins.

MATERIALS AND METHODS

General—Antisera to the human protein (Josman Laboratories, Napa, CA), used for protein blotting, were produced in rabbits against the peptide GPFKTIKTNDPGVLQAR, which was synthesized on a lysine matrix (BioSynthesis, Inc., Lewisville, TX). Protein blots were detected with the Enhanced Chemiluminescent Detection System (Amersham Pharmacia Biotech). Chicken egg white cystatin was from PanVera (Madison, WI). Automated DNA sequencing was performed with a DyeTerminator Cycle Sequencing Ready Reaction kit on an Applied Biosystems Prism 377 DNA sequencer (both from Perkin-Elmer). PCR products were performed on a Perkin-Elmer 9600 with a GeneAmp PCR kit (Perkin-Elmer) followed by purification with QiAquick Gel Extraction kits (Qiagen, Santa Clarita, CA). Oligonucleotides were synthesized using an Applied Biosystems 394 synthesizer (Perkin-Elmer). cDNA Libraries—cDNA libraries are listed on the Southern blots (see Figs. 5 and 6) and were made with the SuperscriptII system (Life Technologies Inc.) as detailed elsewhere (Refs. 7 and 8 and references therein). Details are also available directly from the authors. In cases where the derivation is not obvious from the name, conditions used are listed below. Mouse libraries: 2) Braf:ER transfectant NIH3T3 cell line, ethanol-treated; 3) Mel14 bright CD4+ cells from spleen, polarized for 4 days with IFN-γ and anti-IL4, activated with anti-CD3 for 2, 6, or 10 days with IFN-γ, interferon-γ; IL, interleukin; EST, expression sequence tag; PCR, polymerase chain reaction; DTT, dithiothreitol.

* DNAX is supported by Schering-Plough Corporation. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked ‗advertisement‘ in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

† To whom correspondence should be addressed: Dept. of Molecular Biology, DNAX Research Institute, 901 California Ave., Palo Alto, CA 94304-1104. Tel.: 650-496-5255; Fax: 650-496-1214; E-mail: Gerard @dnax.org.
16 h and pooled; 4) as for 3, except polarized with IL4 and anti-IFN-γ; 5) naive CD4⁺ cells from an ovalbumin-specific TCR transgenic mouse, polarized for 3 weeks with IL12 and anti-IL4, activated with PMA and ionomycin for 2, 6, or 24 h, and pooled; 6) as for 5, but polarized with IL4 and anti-IL12; 8) D1.1 TH1 cell clone, 3 weeks after last antigen stimulation; 9) D1.1 TH1 cell clone, concanavalin A-stimulated for 15 h; 10) CDC35 TH2 cell clone, 3 weeks after last antigen stimulation; 11) CDC35 TH2 cell clone, concanavalin A-stimulated 15 h; 12) Mel14 dendritic cells derived from elutriated blood monocytes after 5 days in anti-CD28, IL4, and anti-IFN-γ; 13) α granulocytes derived from a different monocyte donor; 14) as for 12, but polarized for 7 days with IFN-γ, IL12, and anti-IL4; 14) as for 13, but polarized with IL4 and anti-IFN-γ; 18) B-cells from LPS-induced total spleen; 19) splenic dendritic cells derived by metrizamide enrichment; 21) RAW 264.7 cells activated with LPS 4 h; 22) bone marrow-derived macrophages grown in granulocyte-macrophage CSF and macrophage CSF; 24) macrophage cell line treated with LPS and anti-IL10 for 0.5, 1, 3, 5, 6, or 12 h and pooled; 25) as for 24, but treated with LPS and IL10. Rat libraries: 46) normal joint tissue, mock-infected; 47) pooled stromalcoel cell wall antigen-induced arthritic joints. Human libraries: 50) natural killer cell clone derived from peripheral blood of a large granular lymphocyte leukemia patient, treated with IL2; 51) 20 pooled natural killer cell clones activated with PMA and ionomycin for 6 h; 53) B-cell line JY activated with PMA and ionomycin for 3 h; 54) as for 53, but treated with anti-CD40 and IL4; 55) Mel14 dendritic cells derived from elutriated blood monocytes after 5 days in anti-CD28, IL4, and anti-IFN-γ; 56) as for 64, but treated with anti-CD28 and anti-CD3; 66) TH0 cell clone HY06 treated with specific antigenic peptide for 2, 6, or 12 h and pooled; 67) as for 66, but treated with anti-CD28 and anti-CD3 for 3, 6, or 12 h and pooled; 69) peripheral blood mononuclear cells (monocytes, T-cells, natural killer cells, granulocytes, and B-cells) treated with anti-CD3 and anti-CD28 for 6 h; 70) T-cell clone HY06 treated with specific antigenic peptide for 2, 6, or 12 h and pooled; 71) kidney epithelial carcinoma cell line CHO activated with PMA and ionomycin for 1 or 6 h and pooled; 72) lung fibroblast sarcoma line MRC5, treated as for 71; 73) hematopoietic precursor cell line TFl1, treated as for 71; 74) malignant tumor leiomyosarcoma; 75) normal myometrium; 77) dendritic cells derived from eluted blood monocytes after days in granulocyte-macrophage CSF and IL4, activated with tumor necrosis factor α, IL1α, and monocyte supernatant for 4 or 16 h and pooled; 78) as for 77, but activated with LPS; 79) as for 77, but not activated; 80) as for 79, but from a different monocyte donor; 81) 95% CD1a⁺ CD86⁺ dendritic cells derived from CD34⁺ stem cells after 12 days in granulocyte-macrophage CSF and tumor necrosis factor α, FACS-separated, activated with PMA and ionomycin for 1 or 6 h, and pooled; 82) as for 81, but with CD1a⁺ cells; 83) as for 81, but with CD1a⁺ cells; 84) 70% CD1a⁺ dendritic cells derived from CD34⁺ stem cells after 12 days in granulocyte-macrophage CSF and tumor necrosis factor α, activated with PMA and ionomycin for 6 h; 85) as for 84, but activated for 1 h; 86) as for 84, but not activated; 87) elutriated monocytes activated with LPS for 6 h; 88) elutriated monocytes activated with LPS for 1 h; 89) as for 87, but activated with LPS, IFN-γ, and IL10 for 4 or 16 h and pooled; 90) as for 89, but activated with LPS, IFN-γ, and anti-IL10; 91) as for 89, but from a different donor; 92) as for 90, but from a different donor; 93) U937 mononuclear cell line activated with PMA and ionomycin for 1 or 6 h and pooled; 94–105) from a 28-week-old fetus; 106) from a 28-week pregnancy; 107) from a 12-year-old patient. 4. Identification and Characterization of Human Leukocystatin mRNA—An average of 375 base pairs of highly unambiguous sequence (EST) from individual clones from cDNA libraries 84 and 86 were determined. The sequences were analyzed for possible encoded function by BLAST searches versus the public data bases, followed by BLASTP searches of the open reading frames (9). By this method, two leukocystatin ESTs (from a total of 1190) were identified. Both cDNAs were completely sequenced on each DNA strand and were found to be full-length. 5. Isolation and Characterization of Mouse Leukocystatin mRNA—160 pools of approximately 500 clones each from a mouse TH2 cDNA library (Ref. 10; Library 4) were amplified overnight to form sublibraries. Southern blots were performed on the sublibraries using a 32p-labeled 321-base pair probe to the human sequence (see below), washing with cross-species conditions (2x SSC, 0.1% SDS at 65 °C). One of the 160 sublibraries showed a positive signal. A bacterial stock from this pool was plated out, and colony hybridization was conducted under the same conditions to yield several possible positive clones, one of which was selected and found to encode a full-length copy of mouse leukocystatin. cDNA Library Southern Blots—The method presented by Bolin et al. (7) was followed. Briefly, NotI/Sall digests of 5 μg of cDNA library released the cDNA inserts from the vector. Digestion reactions were run on 1% agarose gels, transferred to Nytran + filters (Schleicher and Schuell), and cross-linked with a UV Stratalinker 1800 cross-linker (Stratagene, La Jolla, CA). For the human blot, a 321-base pair 32p-labeled probe was synthesized with [32pIDCTP (Amersham Pharmacia Biotech) using the Rediprime system (Amersham Pharmacia Biotech). Hybridization with 1.5 x 10⁶ cpm/ml was performed at 60 °C in ExpressHyb (CLONTECH, Palo Alto, CA) following washes in 0.5 x SSC, 0.1% SDS. The 343-base pair 32p-labeled mouse probe was made using a Prime-IT II kit (Strategene) followed by purification on a Centrisep column (Princeton Separations, Adelphi, NJ). Hybridization was performed in 0.5 x sodium phosphate, pH 7.2, 7% SDS, 0.5 μg EDTA at 65 °C followed by washes in 0.1 x SSC, 0.1% SDS. Intensities of the bands were quantitated with a Molecular Dynamics Personal Densitometer (Sunnyvale, CA) scan of the developed x-ray film (Kodak BioMax, Rochester, NY). 6. Genomic DNA Sequence—A 129SV mouse genomic library (Strategene) was screened with a 617-base pair 32p-labeled probe (complementary to the mouse cDNA sequence) in QuikHyb (Strategene) using washes recommended by the manufacturer. Of approximately 10 million clones, 4 were identified as being positive. Through a series of PCRs using primers that hybridized to various portions of the leukocystatin sequence, one clone was shown to contain the entire gene. In order to fully sequence this gene, a series of PCRs (26 reactions total) were performed. The resulting overlapping fragments were sequenced in both directions, generating 12 kilobases of sequence consistent with the cDNA sequence. Recombinant Protein Expression and Purification—The PCR primers listed in Table I were used to amplify the leukocystatin sequence with appropriate restriction sites. These amplimers were subcloned into the BamHI/NotI or BamHI/EcoRI sites of pGEX-4T-1 (Amersham Pharmacia Biotech) or HindIII/EcoRI sites of pFLAG (IBI, Eastman Kodak). The cloning regions of the constructs were completely sequenced, and DNA preparations using QIAfilter plasmid maxi kits (QIAGEN) were made for transformation into the Escherichia coli strains used for protein expression. The following E. coli strains were used to produce the GST fusion proteins:...
proteins: human short,\(^2\) W3110 (F\(^-\), thi-1); human long, X156F (lec-6, proC34, purE42, trpE38, thi-1, ara14, lacY1, galK2, xyl-5, txy7, azi-6, rpsL109, supF4)); mouse short, Ut4400 (azi-6, lacY1, leu6, mit-1, proC14, rpsL109, thi-1, trpE38, txy67, entoA303, lepA). The E. coli strain containing the pEXX-cystatin fusion was grown as a 15-liter fermentor for bacterial inclusion bodies, and 1 liter of cells was harvested in 13°C water (2% yeast extract, 0.5°C KH\(_2\)PO\(_4\), 20°C g/liter glycerol, 1 g/liter MgSO\(_4\), 50 mg/liter ampicillin). The culture was induced with 0.4°C isopropyl-1-thio-b-galactopyranoside at an A\(_{600}\) of 4 for 4 h at 37°C. Cell pellets were resuspended in 1 liter of TE with a Brinkmann (Westbury, NY) polytron PT3000 homogenizer and ruptured with three passes through a microfluidizer. The homogenate was centrifuged at 23,000 × g for 30 min, and pellets were washed with TE/1% Triton X-100, TE alone, and finally 2 mM guanidinium chloride/20 mM Tris-HCl, pH 8.25, 0.1% SDS, and 10% or 80% acetonitrile to remove insoluble material, the denatured protein was diluted 1000 × into 0.4°C guanidinium chloride, 50 mM Tris-HCl, pH 8.25, 2 mM EDTA, 1 mM EDTA, 1 mM Pefabloc (Enterchem, Stamford, CT), 10 mM DTT (10 m/lg of inclusion bodies). Following centrifugation at 23,000 × g for 30 min to remove insoluble material, the concentrated, and diafiltered into 50 mM Tris-HCl, pH 8. The protein was loaded on a glutathione-Sepharose 4B column at 2 ml/min. The column was washed with phosphate-buffered saline (PBS, Life Technologies, Inc.) and then PBS containing 0.5 M NaCl and subsequently eluted with 20 mM reduced glutathione/PBS. The fractions containing leukocystatin (as determined by Western blot) were pooled, diluted with 50 mM Tris-HCl, pH 7, and loaded on an S-Sepharose column. The bound protein was eluted with a linear 0 → 1 M NaCl gradient, the leukocystatin-containing fractions were pooled, and the fusion was cleaved with thrombin (4 units/µg of protein) for 1 h at 37°C. After thrombin inhibition with hirudin (0.2 units/µg of protein; Sigma) and 1 mM Pefabloc, the cleaved protein was loaded onto a Poros reverse phase column (PerSeptive Biosystems, Framingham, MA). The column was washed with 2% acetonitrile/0.1% trifluoroacetic acid and eluted with a 2% → 80% acetonitrile gradient containing 0.1% trifluoroacetic acid. Leukocystatin-containing fractions, as determined by Western blot, were pooled, dialyzed into 50 mM NaOAc, pH 4, and stored at 4°C. The expression of the FLAG-tagged cystatin was similar, with the following E. coli strains being used to produce the protein: human short, UT4400; human long, W3110. However, following cell harvesting, the periplasmic fraction was obtained by osmotic shock (1 h at 4°C in 50 mM Tris-HCl, pH 8, 2 mM EDTA, 20% sucrose, 0.1 mg/ml lysozyme). The volume was doubled using the above buffer, and benzonase (25,000 units/mg of extract; American International Chemical Inc., Natick, MA) was added. After incubation for 10 min, the suspension was centrifuged at 27,500 × g for 45 min. The inhibitor was purified from the supernatant by chromatography over a 5-m/2 M column (Kodak Scientific Imaging Systems) and eluted with 20 mM glycine hydrochloride, pH 3. Following dialysis into 20 mM sodium citrate, pH 4, those fractions containing cystatin were further chromatographed on an S-Sepharose column and eluted with a 0 → 1 M NaCl gradient in 20 mM sodium citrate, pH 4. N-terminal amino acid sequencing (ABI 476 Protein Sequencer) was performed for all forms and agreed with the predicted sequences. The mouse short and FLAG-human long materials were quantitated by amino acid analysis (Hewlett Packard AminoQuant using the manufacturer’s standards), and the concentration obtained for the FLAG-human long form agreed within 2-fold with that obtained by densitometry (Molecular Dynamics Personal Densitometer) scanning of a silver-stained (Daiichi, Integrated Separation Systems, Natick, MA) 10% Novex Bis-Tris gel using lysozyme as a concentration standard. Determination of protein concentrations for the other variants was obtained by densitometry of silver-stained gels using both lysozyme and the FLAG-human long cystatin as standards. Final yields of protein were as follows: mouse short, 0.57 mg/15 liters; human short, 0.012 mg/15 liters; human long, 1 mg/11 liters; FLAG-human short, 0.6 mg/15 liters; FLAG-human long, 1 mg/2 liters.

**Refolding of Chicken Egg White Cystatin**—2 mg of chicken egg white cystatin was concentrated to 100 µl and then incubated in 1 ml of 8% guanidinium chloride/50 mM Tris-HCl, pH 8.25, 2 mM EDTA, 1 mM Pefabloc, 10 mM DTT, 100 mM NaCl, and 1% glycerol at 37°C for 100 minutes. The protein was then incubated in 1 ml of 0.4 M guanidinium chloride, 50 mM Tris-HCl, pH 8.25, 2 mM EDTA, 1 mM Pefabloc, 2.5 mM reduced glutathione, 1 mM oxidized glutathione and renatured at 4°C overnight. The refolded material was filtered, concentrated, and diafiltered into 50 mM Tris-HCl, pH 8. The protein was loaded on a glutathione-Sepharose 4B column at 2 ml/min. The column was washed with phosphate-buffered saline (PBS, Life Technologies, Inc.) and then PBS containing 0.5 M NaCl and subsequently eluted with 20 mM reduced glutathione/PBS. The fractions containing leukocystatin (as determined by Western blot) were pooled, diluted with 50 mM Tris-HCl, pH 7, and loaded on an S-Sepharose column. The bound protein was eluted with a linear 0 → 1 M NaCl gradient, the leukocystatin-containing fractions were pooled, and the fusion was cleaved with thrombin (4 units/µg of protein) for 1 h at 37°C. After thrombin inhibition with hirudin (0.2 units/µg of protein; Sigma) and 1 mM Pefabloc, the cleaved protein was loaded onto a Poros reverse phase column (PerSeptive Biosystems, Framingham, MA). The column was washed with 2% acetonitrile/0.1% trifluoroacetic acid and eluted with a 2% → 80% acetonitrile gradient containing 0.1% trifluoroacetic acid. Leukocystatin-containing fractions, as determined by Western blot, were pooled, dialyzed into 50 mM NaOAc, pH 4, and stored at 4°C. The expression of the FLAG-tagged cystatin was similar, with the following E. coli strains being used to produce the protein: human short, UT4400; human long, W3110. However, following cell harvesting, the periplasmic fraction was obtained by osmotic shock (1 h at 4°C in 50 mM Tris-HCl, pH 8, 2 mM EDTA, 20% sucrose, 0.1 mg/ml lysozyme). The volume was doubled using the above buffer, and benzonase (25,000 units/mg of extract; American International Chemical Inc., Natick, MA) was added. After incubation for 10 min, the suspension was centrifuged at 27,500 × g for 45 min. The inhibitor was purified from the supernatant by chromatography over a 5-m/2 M column (Kodak Scientific Imaging Systems) and eluted with 20 mM glycine hydrochloride, pH 3. Following dialysis into 20 mM sodium citrate, pH 4, those fractions containing cystatin were further chromatographed on an S-Sepharose column and eluted with a 0 → 1 M NaCl gradient in 20 mM sodium citrate, pH 4. N-terminal amino acid sequencing (ABI 476 Protein Sequencer) was performed for all forms and agreed with the predicted sequences. The mouse short and FLAG-human long materials were quantitated by amino acid analysis (Hewlett Packard AminoQuant using the manufacturer’s standards), and the concentration obtained for the FLAG-human long form agreed within 2-fold with that obtained by densitometry (Molecular Dynamics Personal Densitometer) scanning of a silver-stained (Daiichi, Integrated Separation Systems, Natick, MA) 10% Novex Bis-Tris gel using lysozyme as a concentration standard. Determination of protein concentrations for the other variants was obtained by densitometry of silver-stained gels using both lysozyme and the FLAG-human long cystatin as standards. Final yields of protein were as follows: mouse short, 0.57 mg/15 liters; human short, 0.012 mg/15 liters; human long, 1 mg/11 liters; FLAG-human short, 0.6 mg/15 liters; FLAG-human long, 1 mg/2 liters.

2 Short form, amino acids 37–146; long form, amino acids 20–146. Amino acid numbering is defined in Fig. 3.
Gly19, would be approximately the same length as for other Class II cystatins but show no significant homology to those cystatins, and this is also reflected in the overall genomic structure (see below). In fact, little of the protein preceding the conserved glycine at position 37 is similar to other Class II cystatins both in length and in sequence. The predicted mature N-terminal region of leukocystatin is unusually long, being 8 amino acids longer than chicken egg white cystatin. Although the mouse and human N termini show 54% identity to each other—roughly the same amount as mouse and human cystatin C—they are only about 22% identical to the cystatin C N termini of the same species. As noted above, the human sequence, but not that of the mouse, has two possible start codons. However, PSORT would not predict a signal sequence for the amino acids coded by 266 to 1 but would instead predict that leukocystatin is a Type II transmembrane protein.

Other interesting features of the leukocystatin amino acid sequence include additional cysteine residues, one of which may be involved in stabilizing a homodimeric form of the protein (see below), and lysine residues at positions 35 and 84. Both lysine residues lie in the putative protease binding regions and replace nonpolar residues found in all other cystatins. The possible significance of these residues is examined under “Discussion.”

Genomic Structure—A mouse genomic sequence corresponding to leukocystatin was found, and approximately 12 kilobases was sequenced. The mouse leukocystatin gene (deposited in the GenBankTM data base under accession no. AF031825).

FIG. 1. cDNA sequence of human leukocystatin. The coding sequence is in boldface, with the corresponding protein sequence underneath. The upstream in-frame ATG discussed in the text is underlined. This sequence has been deposited in the GenBankTM under accession no. AF031824.

FIG. 2. cDNA sequence of mouse leukocystatin. The coding sequence is in boldface, with the corresponding protein sequence underneath. The last base of each exon is starred. The position of the polyadenylation signal is underlined. This has been deposited in the GenBankTM data base under accession no. AF031825.
inclusion bodies which had to be refolded. Although the GST fusion, however, was primarily recovered as soluble material from the periplasm; the FLAG-GST or FLAG fusion. Similar methods have been used previously to express other family members (16–20). The FLAG-GST moiety or enterokinase cleavage of the FLAG tag. SDS-polyacrylamide gel electrophoresis showed the expected molecular weights and that the proteins were reasonably pure (Fig. 3). The absence of readily-detectable mouse leukocystatin mRNA in naive and pre-T-cells. A moderate amount is also found in monocytes, whereas B-cells, dendritic cells, and some macrophage libraries show small amounts of cDNA corresponding to this protein. The small amounts seen in lymph nodes, thymus, and spleen probably result from resident lymphocytes. The absence of readily-detectable mouse leukocystatin mRNA in splenic and bone marrow mouse dendritic cells may reflect their lineage.  

Protein Production—Proteins of two types were made in E. coli as either an N-terminal GST or FLAG fusion. Similar methods have been used previously to express other family members (16–20). The FLAG-tagged material was isolated as soluble material from the periplasm; the GST fusion, however, was primarily recovered as inclusion bodies which had to be refolded. Although the Class I cystatin A requires refolding after expression in E. coli (21, 22), no other Class II cystatin is insoluble when expressed in E. coli, even when expressed as the GST fusion (20). The mature protein was isolated following thrombin cleavage of the GST moiety or enterokinase cleavage of the FLAG tag. SDS-polyacrylamide gel electrophoresis showed the expected molecular weights and that the proteins were reasonably pure (Fig. 7). Nonreduced gels, however, suggest that the long form may exist primarily as a dimer (Fig. 8): as DTT concentrations are varied from 0 to 8 mM, this protein product exists as a dimer and as a monomer, respectively, as determined by apparent molecular weights. Amino acid sequencing and detection with a leukocystatin-specific antibody confirmed that the correct proteins were isolated. Because freeze-drying the purified protein resulted in material that could not be resolubilized, the protein was stored at pH 4.0 at 4°C.

Inhibition of Cysteine Proteases—We studied the inhibition of three cysteine proteases (papain, cathepsin B, and cathepsin L) by normal methods (11). All studies were carried out near the optimal pHs of the proteases, and also in the presence of 5 mM DTT, which may partially denature the cystatins but was necessary to maintain maximal activity of the cysteine proteases. Using this method, we confirmed that chicken egg white cystatin binds tightly to cysteine proteases: an apparent inhibition constant of 90 pm was obtained versus cathepsin B, whereas binding was too tight to papain to quantitate. Identical results were obtained following denaturing and renaturing of the chicken egg white cystatin (data not shown). These results are consistent with published values (Table II). We found that binding of leukocystatin to the cysteine proteases studied is slow and is also weaker than other Class II cystatins (Table II). In fact, we could detect no inhibition of cathepsin B activity with leukocystatin (lower limit of detection, K<sub>i</sub> approx. 200 mM), although subnanomolar inhibition constants were found versus papain and cathepsin L. Whereas the different N-terminal forms of leukocystatin had little effect on the inhibition of papain, a 10-fold increase in affinity for cathepsin L was seen with the leukocystatin long form relative to the short form. Finally, although an affinity could not be determined quantitatively, we found that a cysteine-linked dimer of the leukocystatin long form was not as effective as a papain inhibitor as the reduced form. So, although leukocystatin is a functional Class II inhibitor, its unique amino acid sequence...
appears to interfere with binding to the commonly assayed cysteine proteases.

**DISCUSSION**

We have discovered a novel hematopoietic cell-specific Class II cystatin from an EST analysis of human dendritic cells. This protein, which we have called leukocystatin, has all the features of a Class II cystatin, but it has some notable characteristics. For example, leukocystatin contains lysine residues at two positions that are strictly hydrophobic (residue 35) and small, noncharged (residue 84) amino acids in all other characterized cystatins. Position 35 is thought to bind to the P3 site of the target protease (4, 5), so it is possible that this lysine substitution results in an especially high affinity for a cysteine protease with this preferred specificity. Because residues 81–85 in other cystatins usually form nonspecific hydrophobic interactions with the cognate protease (4), it is likely that the contacts formed by this region may also differ from those observed previously. Supporting this, computer modeling has shown that Lys84 would interfere with binding of leukocystatin to papain in this region (see below).

Leukocystatin contains a total of eight cysteines: the four that are conserved with other Class II cystatins, and four unique cysteines, two of which are in the leader region (Fig. 3). Conserved cysteine residues in the N-terminal portion are not seen in any other mature Class II cystatin molecule, although they do occur sporadically in other cystatin leader sequences. Two cysteine residues, in positions different from leukocystatin, also appear in the N-terminal region of each inhibitory kininogen domain, and a polymorphism in the cystatin D sequence introduces a cysteine in this area (18, 23). It is possible that the two additional leukocystatin cysteines in the putative mature protein form an intrachain disulfide and provide added stability. This, however, is not supported by the evidence. Participation of Cys63 in an intrachain disulfide could only occur if the leukocystatin structure is markedly different from chicken egg white cystatin or if the N terminus folds back because the structure of chicken egg white cystatin shows the amino acid corresponding to Cys63 at the end of an a-helix, 34 Å from the N terminus. Furthermore, nonreduced gels indicate that the long form, which contains Cys26, can dimerize (Fig. 8), whereas no evidence of dimerization exists for the short form, which contains Cys63 but not Cys26. Because only monomer is seen in reducing gels, this interaction is apparently mediated by an interchain disulfide, formed by Cys26 from two different molecules. This may be similar to the case of stefin B, a Class I cystatin, which has a cysteine at position 3 that is thought to mediate dimerization (24).

The mouse leukocystatin gene contains four exons (Fig. 4), unlike most other members of the Class II cystatins, which...
have three (25). Soyacystatin is the only known molecule containing one cystatin domain and having a gene encoding four exons. The additional exon in that case, however, lies in a unique C-terminal extension (26). Because the amino acids encoded by the first two exons of leukocystatin are very different from other Class II cystatins, it is clear that the evolution of this region is very different from other family members. The C-terminal genomic organization, however, is similar to the other Class II cystatins, with the intron/exon boundaries being conserved. Furthermore, the N-terminal portion of leukocystatin is not similar to Class I cystatins; the Class I genomic organization is different, with the first intron lying at a position between the first and second leukocystatin introns and with the second lying between the second and third leukocystatin introns (27, 28).

Several forms of leukocystatin were produced in *E. coli* and were active cysteine protease inhibitors. Although some of these products require refolding, the FLAG-tagged material was soluble, similar to other Class II members expressed in *E. coli*, including those used for comparison in Table II (16, 17). Although the Class I cystatin A requires refolding following overexpression in *E. coli*, this was shown to have no adverse effect on activity (21, 22). We further controlled for any effects that refolding may have on activity by examining denatured/renatured chicken egg white cystatin and found no difference in activity following this step.

We determined the apparent *K* <sub>i</sub> values of leukocystatin with papain and cathepsin L. These are compared in Table II with published values for other Class II cystatins. *K* <sub>i</sub> values in the literature vary for the same cystatin-protease pair, probably due to the differing lengths of the N termini in various cystatin preparations; these residues are easily proteolyzed during isolation of native cystatins. In general, the affinity of cystatins for cathepsin B is much weaker than the binding to cathepsin.
Leukocystatin

Inhibition constants (expressed in nM units) were determined by the procedure of Abrahamson, modified as detailed under “Materials and Methods.” Using the equation \[ \frac{[S]}{v_i} = \frac{1}{v_i (K_i/v_i)} (1 + [S]/K_i) \] with carbobenzyoxy-L-phenylalanyl-L-arginine-7-amino-4-methylcoumarin (papain, catL), or carbobenzyoxy-L-argininyl-L-arginine-7-amino-4-methylcoumarin (catB) as substrate at pH 6.7 (papain) or pH 5.5 (catB, catL). Published values corresponding to chicken egg white cystatin and cystatin C inhibition are indicated for reference. The long form corresponds to Gly\(^\text{37}\)-His/Gln\(^{146}\), and FLAG forms include the N-terminal DYKDDDDK sequence. The human long and short forms prepared by enterokinase cleavage of the FLAG-tagged material had activities identical to that prepared from the GST fusion. Chicken long cystatin denatured and renatured as described under “Materials and Methods” had activity identical to the native protein.

| Protein          | Papain | Cathepsin L | Cathepsin B |
|------------------|--------|-------------|-------------|
| Mouse short form | 0.062  | 0.27        | 20          |
| Human short form | 0.17   | 0.23        |             |
| Human long form  | 0.16   | 0.016       | >200        |
| FLAG-human short form | 0.10 |           |             |
| FLAG-human long form | 0.079 |           |             |
| Chicken egg white | <0.01 | 0.09        |             |
| Chicken egg white\(^a\) | 0.044 | 0.003       | 2.9         |
| Cystatin C\(^b\) | 0.000111 | <0.005     | 0.25        |

\(^a\) Ref. 16.  
\(^b\) Ref. 11.

the residue at position 35 may be a primary determinant of protease specificity because these N-terminal residues associate with the protease binding sites (4, 5). Lindahl et al. (32) have shown that an arginine substitution in cystatin C at the position equivalent to Pro\(^{30}\) can have a large impact on binding to cathepsin B or to papain and may even cause displacement of the N terminus from the protease (32). Although that position is probably more critical to tight binding to the cognate protease than the residue at 35, it demonstrates that changes in the amino acids at these positions can greatly affect the ability of cystatins to inhibit cysteine proteases. It is therefore likely that the native binding partner of leukocystatin is unlike that of the examined proteases. It is possible that the target is some as yet unidentified lysosomal protease, or even a protease from a different family. For instance, the ubiquitin-hydrolase UCH-L3 has recently been shown by x-ray crystallography to have a papain-like fold, being particularly similar to cathepsin B in the active-site cleft (33), and so may very well be inhibited by cystatins, although no evidence of this is yet in the literature. Particularly intriguing is the fact that this isozyme is primarily found in hematopoietic cells (34, 35), and is specific for the RGG sequence of ubiquitin. Furthermore, a domain of kininogen has shown inhibitory activity against calpains (36), and legumain is inhibited by chicken egg white cystatin (37), demonstrating that other families of cysteine proteases can be inhibited by these sorts of structures.

In support of a unique target for the leukocystatins, we found little difference in the binding abilities of long and short forms of cystatin with papain in the presence of 5 mM DTT. Based upon the results of experiments with chicken cystatin, in which N-terminally truncated forms were found to not be as efficient inhibitors as the full-length molecules (29–31), we would expect to see dramatic differences in the abilities of these variants to inhibit this enzyme. One possible explanation is that the S-amino acid N-terminal extension of the longer form impedes binding, although extensions have been shown to have little effect for other cystatins (38). Furthermore, the absence of an effect of the FLAG tag supports the idea that N-terminal extensions do not influence leukocystatin binding. The unique lysine at position 35 may, however, interfere with complex formation. There is also some evidence that the long forms

....

FIG. 7. Silver-stained SDS-polyacrylamide gel electrophoresis of recombinant leukocystatins. Each lane contains approximately 50 ng of protein. Lane A, molecular weight markers. Lane B, lysozyme. Lane C, chicken egg white cystatin. Lane D, human short leukocystatin. Lane E, human long leukocystatin. Lane F, FLAG-human short leukocystatin. Lane G, FLAG-human long leukocystatin. Lane H, mouse short leukocystatin.

FIG. 8. Nonreducing SDS-polyacrylamide gel electrophoresis of FLAG-human long leukocystatin treated with varying amounts of DTT. 50 ng of FLAG-human long leukocystatin in the presence of 8.0 (lane B), 4.0 (lane C), 1.6 (lane D), 0.8 (lane E), 0.4 (lane F), or 0 mM (lane G) DTT as indicated. Lane A contains molecular mass markers.
dimerize, and this may interfere with binding. Under the assay conditions, however, a large proportion is likely to be monomeric, as evidenced by the titration shown in Fig. 8. That the interchain cysteine primarily mediates this dimerization is evidenced by the fact that activity increases for the long form with increasing DTT concentrations. If we were to assume that the dimeric form did not bind at all to the studied cysteine proteases (and that all of the inhibition resulted from the monomeric form) this would only result in changing the apparent $K_i$ by a factor of less than 2 (in favor of tighter binding), because the effective concentration would be changed by this amount.

Although there was no effect on papain inhibition, there was a 10-fold increase in binding affinity to cathepsin L with the long form, showing that at least for this particular case, the N terminus contributes to binding. This supports the idea that the various portions of cystatins are differentially involved in association to individual proteases, even though the three-dimensional structures of these proteases are very similar. We would expect the native binding partner of leukocystatin to fully take advantage of the unique features in these sites.

Leukocystatin was shown by cDNA library Southern blots to be expressed selectively in hematopoietic cells. Examination of a wide variety of immune cell types suggests that the highest levels are expressed in T-cells, monocytes, and dendritic cells. Clearly, a search for a specific target protease should focus on the effector functions of these immune cell types. Currently, we are developing other tagged versions of leukocystatin, additional antibody reagents, and a mouse gene knockout to probe the effector functions of these immune cell types. Currently, we are developing other tagged versions of leukocystatin, additional antibody reagents, and a mouse gene knockout to probe in depth the biological role of this novel Class II cystatin.

In conclusion, we have characterized a new Class II cystatin, termed leukocystatin, which has a novel sequence, including unique lysine residues at two important protease binding sites, and a distinct distribution in hematopoietic cells.

Acknowledgments—We thank Felix Vega for peptide sequencing, Connie Huffine and Anh Quan for assisting with DNA sequencing, David Cambell for preliminary protease assays, and Jackie Timans for making the FLAG-human short cystatin construct.

REFERENCES
1. Riese, R. J., Wolf, P. R., Bromde, D., Natkin, L. R., Villadangos, J. A., Ploegh, H. L., and Chapman, H. A. (1996) *Immunity* 4, 357–366
2. Calkins, C. C., and Sloane, B. F. (1995) *Bioch. Chem. Hoppe-Seyer* 376, 71–80
3. Henschke, Y. M., Veerman, E. C., and Nieuw Amerongen, A. V. (1996) *Bioch. Chem. Hoppe-Seyer* 377, 71–86
4. Stubbs, M. T., Taber, B., Bode, W., Huber, R., Jeraldo, R., Lenarcic, B., and Turk, V. (1986) *EMBO J.* 5, 1939–1947
5. Bode, W., Engh, K., Musil, D., Thiele, U., Huber, R., Karshikov, A., Brzin, J., Kos, J., and Turk, V. (1988) *EMBO J.* 7, 2593–2599
6. Dieckmann, T., Mitschang, L., Hofmann, M., Kos, J., Turk, V., Auerswald, E. A., Jencsicke, R., and Oschkinat, H. (1993) *J. Mol. Biol.* 234, 1045–1059
7. Bolin, L. M., McNeil, T., Larios, L. A., DeVaux, B., Franz-Bacon, K., Gorman, D. M., Zurewski, S., Murray, R., and McClanahan, T. K. (1997) *J. Neurosci.* 17, 5493–5502
8. Murphy, E., Heny, S., Sher, A., and O’Garra, A. (1993) *J. Immunol. Methods* 162, 211–223
9. Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J. (1990) *J. Mol. Biol.* 215, 403–410
10. Gibb, K. J., and Coffman, R. L. (1994) *J. Immunol.* 152, 5180–5188
11. Abrahamson, M. (1994) *Methods Enzymol.* 244, 685–700
12. Bieth, J. (1974) in *Buerger Symposium V: Proteinase Inhibitors*, pp. 463–469, Springer-Verlag, New York
13. Williams, J. W., and Morrison, J. F. (1979) *Methods Enzymol.* 63, 437–467
14. Nakai, K., and Kanemura, M. (1992) *Genomics* 14, 897–911
15. Shortman, K., and Caux, C. (1997) *Stem Cells* 15, 409–419
16. Auerswald, E. A., Genenger, G., Assafal-Machleidt, I., Machleidt, W., Ungh, R. A., and Fritz, H. (1992) *Eur. J. Biochem.* 209, 837–845
17. Hall, A., Hakansson, K., Mason, R. W., Grubb, A., and Abrahamson, M. (1995) *J. Biol. Chem.* 270, 5115–5121
18. Freije, J. P., Balbin, M., Abrahamson, M., Velasco, G., Dalboge, H., Grubb, A., and Lopez-Otin, C. (1993) *J. Biol. Chem.* 268, 15737–15744
19. Ni, J., Abrahamson, M., Zhang, M., Fernandez, M. A., Grubb, A., Sa, 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
20. Setiopoulou, G., Anisowicz, A., and Sager, R. (1997) *J. Biol. Chem.* 272, 903–910
21. Shibuya, K., Kaji, H., Itoh, T., Ohyama, Y., Usui, K., Tate, T., Sakeda, A., Kumagai, I., Hiroi, I., Miura, K., Inagaki, F., and Samejima, T. (1995) *Biochemistry* 34, 12185–12192
22. Shibuya, K., Kaji, H., Ohyama, Y., Tate, S., Kainosho, M., Inagaki, F., and Samejima, T. (1995) *J. Biochem.* 118, 635–642
23. Balbin, M., Hall, A., Grubb, A., Mason, R. W., Lopez-Otin, C., and Abrahamson, M. (1994) *J. Biol. Chem.* 269, 23156–23162
24. Turk, V., and Bode, W. (1991) *FEBS Lett.* 285, 213–219
25. Saitoh, E., and Isemura, S. (1995) *Crit. Rev. Oral Biol. Med.* 4, 487–493
26. Misaka, T., Kuroda, M., Iwabuchi, K., Abe, K., and Arui, S. (1996) *Eur. J. Biochem.* 240, 609–614
27. Pennacchio, L. A., and Myers, R. M. (1996) *Genome Res.* 6, 1103–1109
28. Sato, N., Ishidoh, K., Uchiyama, Y., and Kominami, E. (1992) *Gene* 114, 257–260
29. Machleidt, W., Thiele, U., Laber, B., Assafal-Machleidt, I., Koster, A., Wiegand, G., Kos, J., Turk, V., and Bode, W. (1989) *FEBS Lett.* 243, 234–238
30. Lindahl, P., Abrahamson, M., and Bjork, I. (1992) *Biochem.* 281, 49–55
31. Bjork, I., Pol, E., Raub-Segall, E., Abrahamson, M., Rowan, A. D., and Mort, J. S. (1994) *Biochem.* 239, 219–225
32. Lindahl, P., Rippol, D., Abrahamson, M., Mort, J. S., and Storer, A. C. (1994) *Biochemistry* 33, 4384–4392
33. Johnston, S. C., Larsen, C. N., Cook, W. J., Wilkinson, K. D., and Hill, C. P. (1997) *EMBO J.* 16, 3787–3796
34. Wilkinson, K. D., Lee, K. M., Deshpande, S., Duersken-Hughes, P., Boss, J. M., and Pohl, J. (1989) *Science* 246, 670–673
35. Wilkinson, K. D., Deshpande, S., and Larsen, C. N. (1992) *Biochem. Soc. Trans.* 20, 631–637
36. Salvesen, G., Parkes, C., Abrahamson, M., Grubb, A., and Barrett, A. J. (1986) *Biochem.* 234, 429–434
37. Chen, J.-M., Dando, P. M., Rawlings, N. D., Brown, M. A., Young, N. E., Stevens, R. A., Hewitt, E., Watts, C., and Barrett, A. J. (1997) *J. Biol. Chem.* 272, 8090–8098
38. Hakansson, K., Huh, C., Grubb, A., Karlsson, S., and Abrahamson, M. (1996) *Comp. Biochem. Physiol. B. Biochem. Mol. Biol.* 114, 303–311