Proteolytic Activity of Human Osteoclast Cathepsin K

EXPRESSION, PURIFICATION, ACTIVATION, AND SUBSTRATE IDENTIFICATION*

(Received for publication, October 27, 1995, and in revised form, March 13, 1996)

Mary J. Bossardt, Thaddeus A. Tomaszek, Scott K. Thompson, Bernard Y. Amegadzie, Charles R. Hanning, Christopher Jones, Jeff T. Krydlat, Dean E. McNulty, Fred H. Drake, Maxine Gowen, and Mark A. Levy**

From the Departments of **Medicinal Chemistry, ¶Gene Expression Sciences, †Protein Biochemistry, and ‡Cellular Biochemistry SmithKline Beecham Pharmaceuticals, King of Prussia, Pennsylvania 19406

Human cathepsin K is a recently identified protein with high primary sequence homology to members of the papain cysteine protease superfamily including cathepsins S, L, and B and is selectively expressed in osteoclasts (Drake, F. H., Dodds, R., J ames, I., Connor, J., Debouck, C., Richardson, S., Lee, E., Rieman, D., Barthlow, R., Hastings, G., and Gowen, M. (1996) J. Biol. Chem. 271, 12511-12516). To characterize its catalytic properties, cathepsin K has been expressed in baculovirus-infected SF21 cells and the soluble recombinant protein isolated from growth media was purified. Purified protein includes an inhibitory pro-leader sequence common to this family of protease. Conditions for enzyme activation upon removal of the pro-sequence have been identified. Fluorogenic peptides have been identified as substrates for mature cathepsin K. In addition, two protein components of bone matrix, collagen and osteonectin, have been shown to be substrates of the activated protease. Cathepsin K is inhibited by E-64 and leupeptin, but not by pepstatin, EDTA, phenylmethylsulfonyl fluoride, or phenanthroline, consistent with its classification within the cysteine protease class. Leupeptin has been characterized as a slow binding inhibitor of cathepsin K (k_{obs}[I] = 273,000 m^{-1} s^{-1}). Cathepsin K may represent the elusive protease implicated in degradation of protein matrix during bone resorption and represents a novel molecular target in treatment of disease states associated with excessive bone loss such as osteoporosis.

Remodeling of the human skeleton is an ongoing cyclical process that involves phases of bone resorption and replacement. Resorption of bone is carried out by multinucleated cells of hematopoietic lineage known as osteoclasts, while osteoblasts are responsible for deposition of new bone matrix. Osteoclasts resorb bone by creating an extracellular compartment, which is maintained at low pH, on the bone surface. The acidic environment removes the mineral phase of the underlying bone, exposing the organic, proteinaceous matrix to proteolytic degradation. Following this cycle, the recruitment of osteoclasts to the site would begin the process of laying down a new protein matrix that is subsequently mineralized.

Several studies have suggested the involvement of cysteine class proteases in matrix remodeling, including demonstrations that prototypic class inhibitors such as leupeptin, E-64, and cystatin (1-4) are effective in models of osteoclast-mediated bone resorption (5). These inhibition results along with other circumstantial observations, such as the low pH activity of the cysteine protease located within osteoclasts, have most often been interpreted as evidence for the involvement of cathepsins B, S, or L in degradation of protein components within the bone matrix. Recently, a novel protein with high sequence homology to cysteine proteases of the papain/cathepsin superfamily (6-10) has been shown to be highly expressed within the osteoclast, but not in cells from spleen, liver, kidney, muscle, or lung (11). In contrast, relatively low levels of cathepsins S, L, and B were found within the osteoclast (11). This unique and selective cellular distribution has prompted reference to this protein as cathepsin K (11). Its selective presence within the osteoclast suggests that cathepsin K is the previously elusive cysteine protease involved in bone resorption, and consequently may represent a potential target for therapeutic intervention of disease states involving excessive bone loss. Biochemical, functional, and structural studies of human cathepsin K have been initiated to further develop this concept. In this report are included our approaches to protein heterologous expression, purification, and processing leading to demonstrations of catalysis by mature cathepsin K with peptide and protein substrates as well as its interactions with prototypic protease class inhibitors.

MATERIALS AND METHODS

Prestained molecular weight markers were purchased from Amer sham Corp. and Novex. Precast SDS-PAGE gels (12% and 15%) were obtained from Bio-Rad. Rapid Coomassie Blue protein stain was acquired from Diversified Biotech. Fluorogenic peptides and prototypic protease inhibitors were obtained from Bachem, Sigma, or Novo Biochem. [3H]Propionylated rat type I collagen was purchased from Du Pont NEN. Protein concentrations were estimated by the BCA dye reagent (Pierce) or the Bradford method using the commercially available Bio-Rad protein assay. Two of the potential substrates (Cbz-Leu-Leu-AMC and Cbz-Leu-Leu-Leu-AMC) were prepared by standard methods of peptide synthesis; details of their syntheses will be presented separately.

Cells and Virus

SF21 insect cells and wild-type baculovirus were purchased from Invitrogen Corp. The cells were cultured in Grace's medium (Invitrogen)

1 The nomenclature of this protein sequence has taken several forms, including carboxypeptidase K (6), cathepsin O (7), cathepsin X (8), and the rabbit homologue, OC2 (9). A second, unrelated sequence also has been referred to as cathepsin O (31). Upon our inquiry, the Nomenclature Committee of the International Union of Biochemistry and Molecular Biology has assigned the name cathepsin K (EC 3.4.22.38) to the protease described in this paper.

2 The abbreviations used are: PAGE, polyacrylamide gel electrophoresis; Cbz, benzoylcarbonyl; AMC, aminomethylcoumarin.

* 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: Mail Code UW-2103, Dept. of Medicinal Chemistry, SmithKline Beecham Pharmaceuticals, P. O. Box 1539, King of Prussia, PA 19406. Tel.: 610-270-6602; Fax: 610-270-4162.
Proteolytic Activity of Human Osteoclast Cathepsin K

Construction of Expression Vectors and Recombinant Viruses

The human cathepsin K cDNA was engineered for expression as follows. A Bluescript vector containing the cDNA fragment encoding the human cathepsin K open reading frame was cleaved with BamHI and SpeI. The resulting 1.5-kilobase pair fragment containing the human cathepsin K coding region was gel-purified and ligated into the baculovirus transfer vector PVL1393 that had been digested with BamHI and XbaI, generating the plasmid pBacCatK. This construct was designed to express a full-length molecule (including the first 15 NH2-terminal amino acids, which constitute the signal peptide).

For expression of recombinant viruses, SF21 cells were cotransfected with purified AcNPV linear DNA purchased from Pharmingen and pBacCatK vector using the liposome-mediated transfection technique as described (12) and then incubated at room temperature for 4 days. The supernatants from the transfection were collected and used to select for the appropriate recombinant viruses as described (12), amplified, and stored as virus stocks for subsequent experiments.

Analysis of Proteins from Cathepsin K Recombinant Virus-Infected Cells

SF21 cells were separately infected with about 10 plaque-forming units of either purified recombinant cathepsin K virus (vBacCatK) or a nonrecombinant control HPDE IVA virus (13) per cell at 27 °C in a serum-containing medium. Twenty-four hours after infection, the cells were pelleted, resuspended in a serum-free medium, and incubated at 27 °C for an additional 72 h. Aliquots of the supernatant collected from the infected cells were adjusted to contain 2% sodium dodecyl sulfate (SDS) and 10% mercaptoethanol and boiled for 5 min prior to electrophoresis in a 0.1% SDS, 12% polyacrylamide gel. Western blots were carried out as described previously (13) using a 1:1000 dilution of anti-human cathepsin K antisera. For time course expression of the target cathepsin K protein, aliquots were collected 48, 72, or 96 h after infection with the recombinant viruses. For large scale production of the protein, cells at a density between 1 and 2 × 10^9/ml were infected with the vBacCatK recombinant virus.

Purification of Procathepsin K

Protein Capture—Ten liters of baculovirus medium (sterile filtered) containing secreted recombinant procathepsin K were diluted with an equal volume of 10 mM Hepes, pH 8.0. The pH was adjusted to 8.0 with NaOH and the conductivity determined to be below 7 millisiemens/cm. The sample was loaded onto a 100-ml S-Sepharose Fast Flow (Pharmacia Labs) column at 30 cm/h previously equilibrated with 20 mM Hepes containing 100 mM NaCl and 5 mM EDTA, pH 8.0. The column was washed with 20 mM Hepes, pH 8.0, at a flow rate of 10 cm/h. The protein was eluted with a 0–1 M NaCl linear gradient over 30 column volumes.

Desalting—Based on SDS-PAGE and immunoblots of fractions from the S-Sepharose Fast Flow column, fractions containing the desired protein were pooled. The pool was desalted by dialfiltration using a 7-milliliter Minisette tangential flow system using 20 mM Hepes, pH 8.0, for dialfiltration. The conductivity was monitored during the dialfiltration, and when it reached ~7 millisiemens/cm dialfiltration was stopped. Any insoluble material in the dialyzed sample was removed by centrifugation prior to the next step.

Dye Mimetic Chromatography—The dialfiltered protein solution was loaded onto a blue no. 2 dye mimetic (American International Chemical, Natick, MA) column at 30 cm/h previously equilibrated with 20 mM Hepes, pH 8.0, 50 mM NaCl. Bound proteins were eluted with a 0.05–1.0 M linear NaCl gradient at the same flow rate.

Superdex 75 Chromatography—Pooled procathepsin K was loaded onto a 24-ml FPLC Superdex 75 column at a flow rate of 20 cm/h. The column was eluted under isocratic conditions with equilibration buffer (20 mM Hepes, 0.15 M NaCl, pH 8). Fractions were analyzed by SDS-PAGE and Western blot, and the fractions containing the purified procathepsin K protein were combined. NH2-terminal sequence analysis of the purified protein (LYPEEILD) demonstrated removal of the leading 15 amino acids, corresponding to the proposed (pre) signal sequence of procathepsin K (see “Discussion” below).

Processing and Activation of Procathepsin K

Heat Activation of Cathepsin K—Purified procathepsin K was incubated at 60 °C in standard activation buffer consisting of 50 mM sodium acetate and 20 mM cysteine adjusted to the final pH (variable) with NaOH. At predetermined time points, aliquots were removed from the individual incubation mixtures and quick frozen in liquid nitrogen. After completion of the time course, assay buffer consisting of 100 mM sodium acetate, 20 mM cysteine, and 5 mM EDTA, pH 5.5, was added directly to the frozen samples and the samples were transferred to a 96-well plate. Activity was monitored by the increase of fluorescence (excitation at 360 nM; emission at 460 nM) accompanying release of AMC. Fluorescence was monitored at ambient temperature with a Labsystems Fluoroskan plate reader. The assay was initiated by addition of substrate (Cbz-Phe-Arg-AMC) to a final concentration of 100 μM.

Studies on Alternative Methods of Cathepsin K Activation—Several studies were simultaneously carried out to compare different approaches to activation of cathepsin K. Two methods were used to initiate the activation process: (a) brief exposure of procathepsin K to elevated temperatures, and (b) addition of an aliquot of preactivated cathepsin K to a larger sample of procathepsin K at 4 °C. Specifically, procathepsin K (~1.2 mg/ml) that had been stored at -80 °C (in 20 mM Hepes, 500 mM NaCl, 0.1% CHAPS) was diluted 11-fold into activation buffer at defined pH. Samples initiated by exposure to elevated temperatures were incubated for 10 min at either 37 °C or 50 °C followed by an additional 30 min at room temperature. The samples were then cooled to 4 °C. In the studies of activation upon addition of catalytic cathepsin K, a sample of procathepsin K was heat-activated at 50 °C for 15 min in pH 4.0 buffer. After incubating at room temperature for an additional 30 min, a small aliquot of this heat treated protein was added to a significantly larger amount (~20-fold excess of procathepsin K to the heat-activated protein) of procathepsin K. These samples were incubated at 4 °C for 6 days. Aliquots were removed from all samples for determination of enzyme activity and proteolytic processing by SDS-PAGE. Proteolytic activity was monitored with 100 μM Cbz-Phe-Arg-AMC as substrate in a final buffer containing 4% Me2SO, 150 mM sodium acetate (pH 5.5), 20 mM cysteine and 5 mM EDTA. Processing of procathepsin K to mature protein was monitored by SDS-PAGE using 15% gels.

Characterization of Fluorogenic Peptide Substrates

Standard assay conditions for determining kinetic constants of potential fluorogenic peptide substrates used 100 mM Na acetate at pH 5.5 containing 20 mM cysteine and 5 mM EDTA. Stock substrate solutions were prepared at concentrations of 10 or 20 mM in Me2SO. Concentrations of substrate were varied for each test compound with final Me2SO concentration maintained at 10%. Independent experiments found that this level of Me2SO had no effect on enzyme activity or kinetic constants. A protein concentration of 3–5 nM was used for each assay as estimated by the Bio-Rad Bradford protein kit. All assays were conducted at ambient temperature. Product fluorescence was monitored with a Perceptive Biosystems Cytofluor II fluorescent plate reader. The linear portions of the initial velocity data from product progress curves were analyzed by the HYPER program as described by Cleland (14) to generate Kav and kobs values. A standard curve with AMC was used in the conversion of fluorescence to molar units.

Inhibition Studies

Studies with prototypic protease inhibitors were conducted with 20 μM Cbz-Phe-Arg-AMC as substrate at pH 5.5 under the conditions described above. All inhibitors were prepared as stock solutions in Me2SO; final Me2SO concentration was held constant at 10%. The time-dependent inhibition of cathepsin K by leupeptin was evaluated using progress curve analysis. Product progress curves were obtained in the absence and presence of inhibitor under conditions as described above. All reactions were initiated by addition of enzyme to solutions of substrate and inhibitor. Values for kobs at each concentration of inhibitor were computed for individual curves by directly fitting of the data to Equation 1, where [AMC] is the concentration of product formed over time, t, is the initial reaction velocity and vss is the final steady state rate.

\[
\frac{[AMC]}{vss} = v_t + (v_0 - v_{ss})(1 - \exp^{-kt})/k_{obs}
\]  

Equation 1

A complete discussion of this kinetic treatment has been presented elsewhere (15).

Evaluation of Proteinaceous Substrates

Degradation of Fibrinogen—Studies with fibrinogen as a potential substrate for recombinant cathepsin K were conducted as described (7).

Processing of Collagen—[3H]Protoprolated collagen (rat tail type I) was used as a potential cathepsin K substrate by a procedure based on the recommendation provided by the manufacturer (DuPont NEN). Assays were conducted at pH 5.5 and were neutralized to promote
precipitation of unreacted collagen prior to analysis.

Cleavage of Osteonectin by Mature Cathepsin K—A solution of human platelet osteonectin (Haematologic Technologies Inc., Essex Junction, VT) was incubated in the presence of 30 nM mature cathepsin K in 75 mM sodium acetate pH 5.5 buffer containing 15 mM cysteine and 4 mM EDTA. Samples were removed at predetermined intervals, immediately added to SDS-PAGE treatment buffer, and heated for 10 min at 95°C. Processing of osteonectin was monitored by SDS-PAGE using 12% gels. The protein bands from a separate gel were electroeluted onto a polyvinylidene difluoride membrane in preparation for NH₂-terminal sequencing.

Amino Acid Analysis

Aliquots of protein (5–10 μg each) were hydrolyzed in vacuo under 6 N HCl for 20 h at 110°C. A 1–2 μg sample of the resulting hydrolyzates were analyzed by ion-exchange amino acid analysis using post-column ninhydrin detection on a Beckman 6300 analyzer equipped with a System Gold data acquisition system.

Amino-terminal Sequence Analysis

Sequence analysis was performed on an Applied Biosystems model 470A gas-phase protein sequencer equipped with a Beckman 126/166 system for on-line phenylthiohydantoin analysis; data were acquired using System Gold chromatography software. Samples were either spotted directly onto Polybrene-coated GF/C filters (Applied Biosystems) or electroblotted onto polyvinylidene difluoride type supports (Problott), and standard Applied Biosystems sequencing cycles were used.

Sequence Alignments

Sequences were aligned and compared using the computer software package (version 8) provided by the Genetics Computer Group (GCG), Madison, WI.

Antibodies to Cathepsin K

Generation of antibodies to cathepsin K used for Western blot analyses has been described (11).

RESULTS AND DISCUSSION

Primary Sequence Analysis—Members of the papain superfamily are typically transcribed and translated as inactive precursors to their catalytically active mature forms (16, 23, 24). Each protein is expressed as a prepro-form, which includes a short NH₂-terminal signal sequence of approximately 12–16 amino acids followed by an intervening leader sequence of approximately 100 residues referred to as the pre- and pro-sequences, respectively. In the case of papain and cathepsins S, L, H, and B, removal of both the pre- and the pro-leader sequences is required for generation of native, mature, catalytically active proteases.

The primary sequence alignment of the pre-proforms of cathepsin K and human cathepsins S and L is presented in Fig. 1. Of the members of this family, the greatest primary sequence homology for the full-length prepro-protein forms occurs between cathepsin K and human cathepsin S (56% identity and 71% similarity using BESTFIT). Progressively lower homology is found with human cathepsin L (51% identity, 68% similarity), cathepsin H (42% identity, 59% similarity), and cathepsin B (27% identity, 53% similarity). The sequence similarity (55%) and identity (41%) between papain and cathepsin K is comparable to that for cathepsin H. Even higher homology is observed between the mature, catalytically active forms of these proteins and that predicted for mature cathepsin K. As examples, the homologies between the mature cathepsin K resulting from removal of the signal and leader sequences and that of human cathepsins S and L are 59% and 60% identity with 73% and 76% similarity, respectively.
acids, which is analogous to that found for the other cathepsins. The putative signal sequence contains a positively charged amino acid (Lys5) close to the initial methionine and a subsequent stretch of hydrophobic amino acids terminated by a consensus alanine (Ala15). The proposed mature, catalytically active form of cathepsin K was predicted to result from cleavage between Arg114 and Ala115 according to the alignment of a consensus sequence defined by P9 (Pro), P4 (Ser), P3 (Val), and P0 (Asp). The active site cysteine, histidine, and asparagine catalytic triad involved in proteolytic catalysis of the papain protease family members can be similarly identified from the sequence alignment to be Cys739, His776, and Asn796 of cathepsin K. All six of the Cys residues within the mature forms of papain and the cathepsins that are thought to be involved in three structural intramolecular disulfide bonds for stabilization of mature enzyme are conserved within the protein sequence of cathepsin K (see Fig. 1). The highest degree of overall homology among the family members can be found in the vicinity of the active site cysteines, which are located in the highly conserved amino-terminal region of the mature enzyme forms.

Cloning and Expression of Human Cathepsin K in Baculovirus-infected Cells—In order to generate sufficient quantities of biologically active human cathepsin K for biochemical, structural, and pharmacological characterization, the human cathepsin K cDNA encoding the pre-proenzyme was subcloned into a baculovirus expression vector; expression was put under the control of the Polh (polyhedrin) promoter. From this, recombinant viruses were generated and the production of recombinant cathepsin was measured in supernatant collected from SF-21 cells infected with cathepsin K recombinant virus (vBacCatK) by Western blot analysis with cathepsin K-specific antiserum (11). As shown in Fig. 2A (lanes 3 and 4), both of these recombinant viruses expressed cross-reacting protein bands of 37 kDa, the approximate molecular mass for the proenzyme from which the leader pre-sequence had been removed. This protein band was not detected in samples prepared from uninfected SF21 cells (Fig. 2A, lane 1).

Baculovirus genome encodes an endogenous cathepsin-like protein (17); however, lysates prepared from uninfected cells or cells infected with a nonrecombinant virus (18) produced no detectable protein band by immunoblotting (Fig. 2A, lane 2) using cathepsin K specific antiserum. This observation eliminates the possibility that viral infection of the cells enhances levels of an endogenous cathepsin-like protein, discounting the possibility of a false positive signal with the cathepsin K antisera. Thus, the immunoreactive protein bands detected either in media collected from infected cells or in the cell pellets were produced only upon infection with recombinant vBacCatK virus.

To determine the optimal time of expression and pattern of accumulation of recombinant human cathepsin K protein, SF21 cells were infected with the appropriate recombinant viruses and soluble protein samples from various times after infection were analyzed by Western blotting (data not shown). Very small levels of cathepsin K expression were detected 48 h after viral infection, while the recombinant 37-kDa protein was shown to accumulate through 96 h, consistent with the regulation of the strong late Polh promoter driving expression (19).

Secretion of Pro-cathepsin K—The level of expression of secretory proteins using baculovirus-infected cells lags behind that of cytosolic proteins (20); in addition, the amount of these proteins produced and the efficiency of secretion depends on the individual signal peptides (20). In order to determine the amount of expressed protein that was secreted relative to that retained within the cells, the supernatant and cell pellets from SF21 cells infected with vBacCatK were evaluated by Western blot analysis. As shown in Fig. 2B (lanes 3 and 4), a protein band with a molecular mass of approximately 37 kDa was detected in samples prepared from both supernatant and pellet, respectively. Interestingly, an additional band of 39 kDa was detected in the sample prepared from infected cell pellets Fig. 2B (lane 4). The size of this larger band is consistent with that predicted for the full-length, pre-proprotein, including the 15-amino acid amino-terminal signal sequence. The presence of cathepsin K retained within the cells suggests that the efficiency of secretion after synthesis was not quantitative. As noted below, amino-terminal sequencing of the 37-kDa band (following purification) isolated from the growth medium confirmed removal of the pre-signal sequence from the full-length recombinant construct.

Purification of Cathepsin K—All of the extracellular, secreted recombinant 37-kDa human cathepsin K was captured efficiently from the SF21 baculovirus expression media on a S-Sepharose Fast Flow column. After appropriate washing, the desired protein was eluted in the middle portion of the NaCl gradient with nearly quantitative recovery of the immunoreactive protein. After dialfiltration and centrifugation, cathepsin K was completely captured on an AIC-Blue-2 Dye mimetic column and subsequently was eluted near the middle of the 0.05–1 m NaCl gradient with approximately 70% recovery. In a final purification step, target protein was chromatographed over a Superdex 75 sizing column. The protein eluted as a sharp symmetrical peak at an apparent molecular mass of 35 kDa, which was near the 37-kDa size observed by SDS-PAGE and consistent with a monomeric form. The resulting single
Proteolytic Activity of Human Osteoclast Cathepsin K

A protein band was estimated to be greater than 95% homogeneous by SDS-PAGE (Fig. 5, lane 4) and was confirmed to be the desired protein by Western blot analysis. Amino-terminal sequence analysis (LYPEE . . .) confirmed that the isolated protein represented the pro-enzyme form in which the leader peptide (MWGLKLVLPPVVSFA) had been removed. No underlying secondary sequences were observed.

Demonstration of Cathepsin K Catalysis—Although cathepsin K was shown to have high sequence homology to known cysteine proteases, peptide or protein substrates were unknown at the time of our initial efforts, thereby posing a challenge in demonstrating proteolytic activity. Due to anticipated structural similarity between cathepsin K and the other proteins in the papain superfamily, it was considered likely that known peptide substrates of the established cathepsins also could be processed by cathepsin K. This concept was significantly strengthened by published data with cathepsins S, L, and B, where it has been shown that different peptide sequences can influence the affinity ($K_m$) and the catalytic efficiency ($k_{cat}/K_m$), yet many common peptide substrates can be cleaved by each protease (21). As a consequence, it was considered likely that substrates based on those known for other members of the papain superfamily would be recognized, although probably not optimized, as substrates for cathepsin K. This approach has recently been taken with cruzain, a cysteine protease from Trypanosoma cruzi that also is a member of the papain superfamily (22).

One of the more common methods used in evaluation of catalytic properties of the cathepsin cysteine proteases employs the cleavage of a fluorogenic (such as AMC) molecule from the carboxyl terminus of a small peptide with the general form,

\[
(\text{AA})_n - \text{AMC} \rightarrow (\text{AA})_n - \text{CO}_2\text{H} + \text{AMC}
\]

where activity is monitored upon release of the signal molecule (AMC) from a peptide of amino acid (AA) sequence of n residues. The primary uncertainty regarding the recognition of such a peptide as a substrate by cathepsin K would involve identity and number of the amino acids within the sequence.

In initial attempts to demonstrate catalytic activity with recombinant cathepsin K obtained by baculovirus expression, samples of media containing the soluble 37-kDa protein were analyzed with the fluorogenic peptides (see below) and rat tail samples of media containing the soluble 37-kDa protein were recombinant cathepsin K obtained by baculovirus expression, all of which were removed over a period of 30 min and assayed at pH 5.5 using Cbz-Phe-Arg-AMC as substrate.

It was not detected without the temperature treatment, suggesting that the conditions of preincubation resulted in the processing and activation of inactive cathepsin K precursor. A separate set of studies has shown that maximal enzyme activity is achieved at assay conditions near pH 5.5 (data not shown).

Processing of Procathepsin K Is Required for Enzymatic Activation—The treatment of procathepsin K at low pH and elevated temperature leading to its activation to catalyze proteolysis of a small peptide substrate has been shown to be coincident with processing of the 37-kDa procathepsin K to smaller protein components. For example, upon activation at 37°C, maximal activity toward Cbz-Phe-Arg-AMC was found to occur after exposure to activation conditions for 45–60 min. Western blot analysis of these same samples showed conversion of the initial 37-kDa procathepsin K into smaller proteins primarily of 33, 27, and 10–12 kDa in size. Enzyme activity seemed to correlate with formation of the 27-kDa protein, the approximate molecular mass of the predicted mature enzyme. The 10–12-kDa fragment is likely to derive from the pro-leader sequence that is removed upon activation.

Follow-up studies have demonstrated that the time course of procathepsin K activation and processing is temperature-dependent with lower temperature both slowing the rate of processing and generating an activated enzyme preparation of significantly higher specific activity. To allow efficient formation of the mature catalytically active form of cathepsin K, several studies were conducted to identify preferred experimental conditions for enzyme activation. In one experiment, samples of procathepsin K were subjected to variable conditions for the initiation of activation, including short exposure to elevated temperatures or addition of a catalytic aliquot of preactivated cathepsin K, followed by incubation at pH 3.5 to 6.0 at 4°C to encourage additional processing and accumulation of the mature enzyme. From these studies it was determined that maximal activity toward Cbz-Phe-Arg-AMC (assays conducted at pH 5.5) was obtained under incubation conditions of pH 4.0. As shown in Fig. 4, a maximum specific activity was achieved after 24 h of incubation at 4°C with the sample that had been initiated by addition of heat preactivated cathepsin K. The resulting

3 During preparation of this manuscript, use of Cbz-Phe-Arg-AMC as a substrate for cathepsin K, referred to as cathepsin O2, was reported (32). Alternative methods for purification and activation as well as additional kinetic characterization of this protease has been described subsequent to the review of this paper (33).
specific activity using these conditions was greater than 10-fold higher than that which had been achieved by the direct heat activation procedures, suggesting a significant improvement in the quality of the resulting activated mature enzyme population.

Analysis of samples from this study by SDS-PAGE with Coomassie staining for protein demonstrated that catalytic activity is associated with a significant accumulation of the 27-kDa protein. Fig. 5 shows a representation of protein processing corresponding to the best activation conditions from the set of studies outlined above (pH 4.0; Fig. 4, solid squares). Integration of the bands from this gel suggests an overall yield of >60% in conversion of the 37-kDa procathepsin K to the 27-kDa mature enzyme.

Amino-terminal analysis of the 27-kDa protein demonstrated the presence of two amino-terminal sequences, Rap-P3-P2-P1-AMC and Rap-P3-P2-P1-AMC, indicating that processing was offset by one and two residues toward the amino terminus from that predicted (APRSSYVKKY) by the primary sequence alignments with cathepsins S and L (Fig. 1). Further studies to characterize the temporal dependence of NH₂-terminal processing upon activation are in progress.

Specificity of Fluorogenic Peptide Substrates—Several fluorogenic peptides of varying amino acid sequence were investigated as potential substrates for the mature 27-kDa cathepsin K. Of approximately 40 peptide sequences with the general structure Cbz-P₂-P₁-AMC, it was demonstrated that an amino acid with hydrophilic side chain such as Arg or Lys in P₁ along with an amino acid having a small, hydrophobic side chain within P₂ were greatly favored as substrates of mature cathepsin K (Table I). In general, sequences having amino acids with hydrophilic side chains at P₂ were not favorable substrates. As example, while Cbz-Leu-Arg-AMC was one of the most efficient peptide substrates for activated cathepsin K, no release of AMC was observed with Cbz-Leu-Leu-AMC in the presence of relatively high enzyme concentrations. A small amount of AMC release was observed at 5 μM Cbz-Phe-Ala-AMC, but limited solubility at higher concentrations precluded further characterization of this molecule. In addition to these sequences, 20 single amino acid AMC molecules were evaluated and none were substrates of cathepsin K. Three of the better substrates, Cbz-Leu-Leu-Arg-AMC, Cbz-Leu-Arg-AMC, and Cbz-Phe-Arg-AMC, demonstrated inhibition at concentrations greater than their Kᵦᵦ values, and two of the non-substrate AMC peptides were shown to be moderate inhibitors of cathepsin K catalysis. Using Cbz-Phe-Arg-AMC as substrate, inhibition by Cbz-Leu-Leu-AMC and Cbz-Leu-Leu-Leu-AMC was estimated to be characterized with Kᵦᵦ values of 3 μM and 0.4 μM, respectively. The inhibition of cathepsin K demonstrated by substrates and close structural analogues can be explained by the possibility of multiple binding modes for these molecules. No activity was detected with any of the substrates in the presence of the purified 37-kDa procathepsin K, providing further support that cathepsin K proteolytic activity requires processing to the mature 27-kDa protein.

Inhibition Studies Supporting That Cathepsin K Is a Cysteine Protease — The activity resulting from heat activation of pro-cathepsin K was initiated by incubation at 37 °C for 10 min, pH 4.5 (●), or incubation at 50 °C for 10 min, pH 4.5 (□), or addition of preactivated cathepsin K, 4°C, pH 4.0 (▲). Following the activation step, the samples were incubated at 4°C to promote formation and accumulation of the mature, catalytically active enzyme. Procathepsin K was studied further with a series of prototypic protease class inhibitors to provide functional support that the enzyme could be classified as a member of the cysteine protease class. As shown in Fig. 6, significant inhibition was observed...
with the classical cysteine protease inhibitors E-64 ($IC_{50} \approx 5$ nM) and leupeptin ($IC_{50} \approx 70$ nM), while minimal effect was seen with pepstatin and phenylmethylsulfonyl fluoride, inhibitors of aspartyl and serine proteases, respectively. No inhibition was observed by addition of EDTA or phenanthroline, classical inhibitors of metalloproteases (data not shown). These results are consistent with classification of mature cathepsin K as a cysteine protease.

Additional experiments have shown that leupeptin is a time-dependent inhibitor of cathepsin K and is significantly more potent than originally estimated in the profiling studies with the prototypic class inhibitors. Representative progress curves showing the release of AMC from peptide substrate are depicted in Fig. 7. Product formation over the assay period in the absence of inhibitor was shown to be linear with incubation time, while curvature of the progress curves was observed in assays that included leupeptin. The shape of these product progress curves is consistent with an increasing loss of enzyme activity with time, which is characteristic of the slow binding of inhibitor to enzyme as described by Equation 1 (15). A plot of the observed rates of enzyme inhibition ($k_{obs}$) from a series of progress curves versus the concentration of leupeptin appears linear (Fig. 7, inset), yielding from the slope of this replot an approximate value of 273,000 M$^{-1}$s$^{-1}$ for $k_{diss}$/[inhibitor], the apparent second order rate constant of inactivation.

**Proteolytic Cleavage of Osteonectin**—The identification, isolation and localization of cathepsin K from cellular sources involved in matrix maintenance and remodeling, such as the osteoclast, contribute to the concept that the physiological role for the mature form of this enzyme might involve the processing of key structural proteins. Degradation of a proteinaceous substrate, fibrinogen, by COS cell extracts containing a recombinant form of cathepsin K has been reported recently (7). The experimental conditions that were used in these studies (pH 4.5 in a highly reducing environment) are similar to those discussed above for cathepsin K activation. In following up this report, experiments in our laboratory have shown that complete fibrinogen degradation catalyzed by activated cathepsin K requires an approximately stoichiometric concentration of mature enzyme to the protein substrate.

This relatively poor ability of cathepsin K to degrade fibrinogen led to the search for alternative matrix-associated substrates. The high levels of cathepsin K detected within osteoclasts (11) concentrated our focus on constituents within bone. Type 1 collagen represents the major structural protein in bone, comprising approximately 90% of the protein matrix. The remaining 10% of matrix in bone consists of several other elements, including osteocalcin, osteopontin, osteonectin, thrombospondin, fibronectin, and bone sialoprotein (29). While the exact roles for these non-collagenous proteins is not well understood, they appear to serve as cell adhesive proteins, and may play a role in matrix mineralization (30).

Two of these proteins, collagen and osteonectin, have been shown to be substrates for mature cathepsin K. As in the case with fibrinogen, incubation of activated enzyme at approximately equimolar concentrations to $[3H]$propionylated collagen resulted in partial release of radiolabel from this modified matrix protein (data not shown). In contrast, proteolytic processing of osteonectin was achieved using much lower (catalytic) concentrations of activated cathepsin K. As depicted in Fig. 8, limited proteolysis of parent osteonectin (42 kDa) resulted in the generation of several smaller protein fragments in a time-dependent manner, with accumulation of three main bands of 34, 14, and 10 kDa after 120 min of incubation. Lane 1 contains molecular size standards; lanes 2 and 4 were empty.
Arg·Val-Lys-Lys-Ile-His...). Most interesting is that the amino terminus from the 14-kDa fragment predicts the presence of the dipeptide Leu-Arg at positions P2-P1, which is consistent with the sequence from one of the better fluorogenic peptide substrates for cathepsin K (Table I). Initial attempts to obtain amino-terminal sequence data of the 10-kDa fragment (band C, Fig. 8) were unsuccessful due to insufficient sample. Efforts to identify the cleavage site defined by this product are continuing. It is anticipated that these observations will lead to the identification of improved peptide-based substrates and small molecule inhibitors for human osteoclast cathepsin K.

Conclusions—Cathepsin K is a newly identified protein with high sequence homology to the related cysteine proteases of papain and the mammalian cathepsins (7, 9–11). Its concentration in osteoclasts has been shown to be high relative to that for cathepsins S, L, and B (11), suggesting that cathepsin K significantly contributes to the enzymatic activity that has been attributed to cysteine class protease(s) implicated in degradation of the protein components within bone during the resorptive stage of organ remodeling.

To gain a better understanding of its functional role within the osteoclast in support of this proposal, we have undertaken the biochemical and functional characterization of human cathepsin K. Expression of pre-procathepsin K with the baculovirus system has been successful at generating soluble recombinant protein. The 37-kDa protein isolated from growth media of infected SF21 cells, which was purified to greater than 95% homogeneity, has been shown to have been processed with removal of the pre-proleader sequence, a common characteristic of papain and the cathepsins. Conditions of low pH in the presence of cysteine have been identified, which affect the conversion of procathepsin K to a mature catalytically active 27-kDa protein. Catalytic activity of the mature enzyme toward peptide and protein substrates occurs at low pH, consistent with the published observations that have led to speculation that the protease in osteoclasts involved in bone resorption may be cysteine class protease(s) implicated in degredation of the protein components within bone during the resorptive stage of organ remodeling.

Acknowledgments—We are grateful to Allan Shatzman, John Gleason, and Brian Metcalf for the support of this work; Ganesh Sathe and Joyce Mao for oligonucleotide synthesis; Laura Grayson for scale-up baculovirus infections; and Dan Veber for helpful discussions. We thank Professor Alan Barrett, Chair of the Advisory Panel of Protease Nomenclature to the Nomenclature Committee of the International Union of Biochemistry and Molecular Biology, for promptly responding to our request to resolve the question of nomenclature for this new protease.

REFERENCES

1. Delaissé, J. M., Eeckhout, Y., and Vaes, G. (1984) Biochim. Biophys. Res. Commun. 125, 441–447.
2. Delaissé, J. M., Eeckhout, Y., and Vaes, G. (1980) Biochem. J. 192, 365–368.
3. Lerner, U. H., and Grubb, A. (1992) J. Bone Miner. Res. 7, 433–439.
4. Delaissé, J. M., Boyde, A., Macnachan, E., Ali, N. N., Saw, C. H. J., Eeckhout, Y., Vaes, G., and Jones, S. J. (1987) Bone 8, 305–313.
5. Hill, P. A., Buttle, D. J., Jones, S. J., Boyde, A., Murata, M., Reynolds, J. J., and Meikle, M. C. (1994) J. Cell. Biochem. 56, 138–139.
6. Inakaka, T., Bibe, G., Ishibashi, O., Tetsuka, K., Kumeawa, M., and Kudukato, T. (1995) Biochem. Biophys. Res. Commun. 206, 89–96.
7. Shi, G.-P., Chapman, H. A., Bhari, S. M., Deli, C., Reddy, V. Y., and Weiss, S. J. (1995) FEBS Lett. 357, 129–134.
8. Li, Y. P., Alexander, M. B., Wucherpfennig, A. L., Chen, W., Yelick, P., and Stadenko, P. (1994) Mol. Biol. Cell. 5, 339a.
9. Tetsuka, K., Tetsuka, Y., Maekawa, K., Satoh, N., Kamioka, H., Hakeda, Y., and Kumeawa, M. (1994) J. Biol. Chem. 269, 1106–1109.
10. Bromme, D., and Okamoto, K. (1995) J. Biol. Chem. Hoppe-Seyler’s Physiol. 376, 379–384.
11. Drake, F. H., Didski, R., Jones, S. J., Connor, J., Debouck, C., Richardson, S., Lee, E., Rieman, D., Barthrow, R., Hastings, G., and Gowen, M. (1996). J. Biol. Chem. 271, 12511–12516.
12. Summers, M. D. and Smith, G. E. (1987) J. Biol. Chem. 262, 17308–17313.
13. Rawlings, N. D., Pearli, L. H., and Buttle, D. J. (1992) J. Biol. Chem. Hoppe-Seyler’s Physiol. 373, 1211–1215.
14. Amegadzie, B. Y., Ahn, B.-Y., and Moss, B. (1992) J. Biol. Chem. 267, 7411–7420.
15. Cleland, W. W. (1992) Adv. Enzymol. Relat. Areas Mol. Biol. 63, 201–231.
16. Wiederanders, B., Bromme, D., Kirschke, H., Figura, K., Schmidt, B., and Peters, C. (1992) J. Biol. Chem. 267, 13708–13713.
17. Delaisse, J. M., Eeckhout, Y., and Vaes, G. (1980) Biochim. Biophys. Res. Commun. 125, 441–447.
18. Eakin, E. A., Mills, A. A., Harth, G., McKerrow, J. H., and Craik, C. (1992) J. Biol. Chem. 267, 7411–7420.
19. Cohen, L. W., Fluharty, C., and Diel, L. C. (1990) Gene (Amst.) 88, 263–268.
20. Taylor, M. A., Pratt, K. A., Revell, P., Baker, K. C., Sumner, I. G., and Stachura, P. W. (1992) Protein Eng. 5, 455–459.
21. Vernet, T., Khouri, H. E., Laflamme, P., Tessier, D. C., Musil, R., Gour-Salin, P., Storer, A. C., and Thomas, D. Y. (1993) J. Biol. Chem. 268, 4832–4838.
22. Hasnain, S., Hira, T., Tam, A., and Mort, J. S. (1992) J. Biol. Chem. 267, 3713–3719.
23. Termini, J. D. (1993) in Primer on the Metabolic Bone Diseases and Disorders of Mineral Metabolism (Favus, M., ed) pp. 21–24, Raven Press, New York.
24. Baron, R. (1993) in Primer on the Metabolic Bone Diseases and Disorders of Mineral Metabolism (Favus, M., ed) pp. 3–9, Raven Press, New York.
25. Vélasco, G., Ferrando, A. A., Puente, X. S., Sánchez, L. M., and López-Otín, C. (1994) J. Biol. Chem. 269, 27136–27142.
26. Palmer, J. T., Rasnjak, D., Klaus, L. L., and Brönme, D. (1995) J. Med. Chem. 38, 3193–3196.
27. Brönme, D., Okamoto, K., Want, B. B., and Biocin, S. (1996) J. Biol. Chem. 271, 2126–2132.