Selective Inhibition of the Collagenolytic Activity of Human Cathepsin K by Altering Its S2 Subsite Specificity†

Fabien Lecaille,‡ Youngchool Choe,§ Wolfgang Brandt,|| Zhenqiang Li,‡ Charles S. Craik,§ and Dieter Brömme*,‡

Department of Human Genetics, Mount Sinai School of Medicine, Box 1498, New York, New York 10029, Department of Pharmaceutical Chemistry, Program in Chemistry and Chemical Biology and Graduate Group in Biophysics, University of California San Francisco, San Francisco, California 94143, and Institute of Plant Biochemistry, Martin Luther University Halle-Wittenberg, Halle/Saale, Germany

Received February 5, 2002; Revised Manuscript Received April 29, 2002

ABSTRACT: The primary specificity of papain-like cysteine proteases (family C1, clan CA) is determined by S2−P2 interactions. Despite the high amino acid sequence identities and structural similarities between cathepsins K and L, only cathepsin K is capable of cleaving interstitial collagens in their triple helical domains. To investigate this specificity, we have engineered the S2 pocket of human cathepsin K into a cathepsin L-like subsite. Using combinatorial fluorogenic substrate libraries, the P1−P4 substrate specificity of the cathepsin K variant, Tyr67Leu/Leu205Ala, was determined and compared with those of cathepsins K and L. The introduction of the double mutation into the S2 subsite of cathepsin K rendered the unique S2 binding preference of the protease for proline and leucine residues into a cathepsin L-like preference for bulky aromatic residues. Homology modeling and docking calculations supported the experimental findings. The cathepsin L-like S2 specificity of the mutant protein and the integrity of its catalytic site were confirmed by kinetic analysis of synthetic di- and tripeptide substrates as well as pH stability and pH activity profile studies. The loss of the ability to accept proline in the S2 binding pocket by the mutant protease completely abolished the collagenolytic activity of cathepsin K whereas its overall gelatinolytic activity remained unaffected. These results indicate that Tyr67 and Leu205 play a key role in the binding of proline residues in the S2 pocket of cathepsin K and are required for its unique collagenase activity.

Type I collagen is the major component of the organic bone matrix which is continually degraded and resynthesized during the bone remodeling process (*1*). It consists of covalently cross-linked triple helices containing two α1(I) and one α2(I) chains. Large regions of the collagen chains are composed of repeating Gly-Pro-Xaa sequences, where Xaa is mainly proline and/or 4-*trans*-L-hydroxyproline (*2*). Degradation of the triple helical type I collagen is achieved by several members of the matrix metalloprotease family (MMPs) such as MMPs 1, 2, 8, and 13 (*3*) and the papain-like cysteine protease, cathepsin K (*4*). MMPs selectively cleave collagens through a single scission across all three chains and generate about three-fourth and one-fourth length collagen fragments (*5*) whereas cathepsin K, similar to bacterial collagenases, cleaves type I and II collagens at multiple sites within their triple helical domains (*6*, *7*). In contrast, other cysteine proteases such as cathepsins L and B cleave collagens only in their nonhelical telopeptide regions (*8*).

Cathepsin K is predominantly expressed in osteoclasts and has been implicated in bone resorption (*9*−*11*). The specific role of cathepsin K in bone resorption was demonstrated by the discovery that deficiency in cathepsin K activity causes an inherited autosomal recessive bone dysplasia, pycnodysostosis (*12*). Several cleavage sites of cathepsin K in the triple helical regions of type I and II collagens have been identified (*6*, *7*). The foremost feature of those sites was the frequent presence of proline residues in the P2 position. The accommodation of a P2 proline residue in the S2 subsite pocket [subsite nomenclature from Schechter and Berger (*13*)] is unique for cathepsin K and in contrast to the preferred acceptance of bulky aromatic P2 residues by cathepsin L. Comparison of three-dimensional structures of cathepsins K and L [PDB, 1mem (*14*); PDB, 1cjl (*15*)] showed amino acid sequence differences at residues 67 and 205 in the primary S2 binding pocket (papain numbering is used in this paper). In cathepsin L, residues 67 and 205 are leucine and alanine, respectively, whereas tyrosine and leucine residues are present at the respective positions in cathepsin K.

To further explore this unusual preference for proline and to test the hypothesis that residues Y67 and L205 are critical for the ability of cathepsin K to cleave triple helical collagen, a cathepsin L-like double mutant, Tyr67Leu/Leu205Ala, of cathepsin K was generated. The subsite specificities of mutant cathepsin K and wild-type cathepsins K and L using combinatorial peptide libraries were compared, and the ability of the mutant protease to cleave triple helical collagens and gelatin was studied.

† This research was supported by NIH Grant AR46182, by a Biomedical Science grant from the Arthritis Foundation, and a postdoctoral scholarship from the Region Centre (France).

* To whom correspondence should be addressed. Phone: 212-659-6753. Fax: 212-849-2508. E-mail: Dieter.Bromme@mssm.edu.

‡ Mount Sinai School of Medicine.

§ University of California San Francisco.

|| Martin Luther University.

10.1021/bi025638x CCC: $22.00
Published on Web 06/04/2002

EXPERIMENTAL PROCEDURES

Materials. Z-Phe-Arg-MCA,[1] Z-Arg-Arg-MCA, and Z-Gly-Pro-Arg-MCA were purchased from Bachem Feinchemikalien Inc. (Bubendorf, Switzerland). Z-Leu-Arg-MCA and Z-Val-Arg-MCA were synthesized as previously described using standard procedures in peptide chemistry (*16*). Dithiothreitol (DTT) and EDTA were obtained from Sigma (St. Louis, MO). Wild-type human cathepsins K and L were expressed in *Pichia pastoris* as described (*17*; D. Brömme, unpublished results).

Generation of the S2 Subsite Mutant of Cathepsin K and Its Expression in P. pastoris. Human procathepsin K cDNA cloned into the *Eco*RI and *Not*I sites of the pBluescript SK-(+) phagemid (Stratagene, La Jolla, CA) was used as the template for in vitro site-directed mutagenesis. Each point mutation was introduced individually into the cDNA using the PCR ligation method with *Pfu* polymerase (New England Biolabs Inc., Beverly, MA). Oligonucleotides used for mutagenesis were as follows (the underlined bases code for the mutated amino acids at positions 67 and 205): Tyr67 → Leu, 5′-TGT GGA GGG GGC CTC ATG ACC AAT GCC-3′; Leu205 → Ala, 5′-GGC ATT GCC AAC GCG GCC AGC TTC CCC-3′. Unique *Eco*RI and *Not*I sites, introduced at the 5′ and 3′ ends of the amplified double mutant proenzyme PCR product, were used to ligate the cDNA into the pPIC9K expression vector (Invitrogen Corp., Carlsbad, CA). The double mutant construct was subsequently sequenced using a series of primers derived from either the vector or internal cDNA sequences. Sequence analyses were performed using an Applied Biosystems model 3777 automated sequencer. The pPIC9K vector containing the Tyr67Leu/Leu205Ala mutant cDNA was linearized with *Bgl*II and then electroporated into *P. pastoris* GS115 cells (Invitrogen Corp.). Several cathepsin K mutant protein expressing clones were obtained after phenotype screening according to the manufacturer's instructions.

Clones were grown in shaker flasks for 5 days according to the manufacturer's protocol, and the liquid culture media were concentrated using a YM10 ultrafiltration membrane with a cutoff size of 10 kDa (Amicon Inc., Beverly, MA). The expression of the mutant was detected by western blot analysis using a rabbit polyclonal antibody raised against human cathepsin K (MS2) as previously described (*18*).

Activation of the Precursor Protein and Purification of the Active Cathepsin K Mutant. The conversion of the pro form into the active enzyme was accomplished by treatment with pepsin as previously described (*18*). Briefly, the concentrated yeast culture supernatant containing the precursor form of the cathepsin K mutant was diluted with 2.5 volumes, to a final solution containing 100 mM sodium acetate, pH 4.0, 2.5 mM DTT, and 2.5 mM EDTA. Porcine pepsin (Sigma Chemical Co., St. Louis, MO) at a final concentration of 0.5 mg/mL was added, and the mixture was incubated for 40 min at 37 °C. The activation was monitored using the fluorogenic substrate Z-Phe-Arg-MCA (5 μM). The activated form of the Tyr67Leu/Leu205Ala mutant was purified by ion-exchange chromatography using SP-Sepharose (Amersham Pharmacia Biotech, Piscataway, NJ) as described in ref *4*. The active site concentration of the mutant protease was determined by titration with E-64 (*19*).

Combinatorial Fluorogenic Substrate Libraries. The substrate specificities of the S1−S4 subsites of the Tyr67Leu/Leu205Ala cathepsin K variant and wild-type cathepsins K and L were determined using tetrapeptide positional scanning synthetic combinatorial libraries as described previously (*20*). To determine the P1 specificity of the cathepsins, a P1 diverse library consisting of 20 sublibraries was used. In each sublibrary, the P1 position is systematically held constant with 1 of the 20 proteinogenic amino acids, omitting cysteine and including norleucine. The P2, P3, and P4 positions were randomized with an equimolar mixture of the 19 amino acids (cysteine was omitted and norleucine added), for a total of 6859 tetrapeptide substrate sequences per each sublibrary. Aliquots of 5×10^{-9} mol from each sublibrary were added to 20 wells of a 96-well Microfluor-1 U-bottom plate (Thermo Labsystems, Inc, Helsinki, Finland) for a final concentration of 7.3 nM compound per well. To determine the P2, P3, and P4 specificities of the cathepsins, a P1-Arg fixed library was used. The library was comprised of P2, P3, and P4 libraries in which either the P2, P3, or P4 position was spatially addressed with 19 amino acids (cysteine was omitted and norleucine was substituted for methionine) while the remaining two positions in each library were randomized. Aliquots of 9×10^{-9} mol from each sublibrary of the P2, P3, and P4 libraries were added to 57 wells (361 compounds per well) of a 96-well Microfluor-1 U-bottom plate for a final concentration of 0.25 μM per well. Hydrolysis reactions were initiated by the addition of 100 μL of enzyme (5 nM for cathepsin K and cathepsin K Tyr67Leu/Leu205Ala, respectively, and 0.5 nM for cathepsin L) and monitored fluorometrically with a Molecular Devices SpectraMax Gemini spectrophotometer with excitation at 380 nm and emission at 460 nm. Assays were performed at 25 °C in a buffer containing 100 mM sodium acetate, pH 5.5, 100 mM NaCl, 10 mM DTT, 1 mM EDTA, and 0.01% Brij-35.

Substrate Assays Using Z-X-Arg-MCS Peptides. Steady-state kinetics were performed with fluorogenic di- and tripeptide substrates as previously described (*4*). The enzymatic activity was followed by monitoring the release of the fluorogenic leaving group, MCA, at an excitation wavelength of 380 nm and an emission wavelength of 460 nm using the Molecular Devices SpectraMax Gemini spectromicrofluorometer. k_{cat} and K_M values were determined using nonlinear regression analysis. Cathepsins were assayed at 25 °C at fixed enzyme concentrations (1−5 nM) and variable substrate concentrations (1−200 μM) in 100 mM sodium acetate buffer, pH 5.5, containing 2.5 mM DTT and 2.5 mM EDTA.

pH Activity Profiles and pH Stability. The pH profiles of the recombinant Tyr67Leu/Leu205Ala cathepsin K variant and cathepsin K were determined at 0.5 μM substrate (Z-Phe-Arg-MCA) concentration ($S < K_M$, where the initial rate, v_0, is directly proportional to the k_{cat}/K_M value). The following buffers were used for the pH activity profile: 100 mM sodium citrate (pH 2.8−5.8), 100 mM sodium phosphate (pH 5.8−8.0), and 100 mM sodium borate (pH 8.0−10.0). All buffers contained 2.5 mM DTT, 2.5 mM EDTA, and 400 mM NaCl. A three protonation model was used for least-squares regression analysis of the pH activity data. The data

[1] Abbreviations: C-4S, chondroitin 4-sulfate; Z, benzyloxycarbonyl; E-64, l-3-carboxy-*trans*-2,3-epoxypropionylleucylamido(4-guanidino)-butane; MCA, 7-amino-4-methylcoumarin; SDS, sodium dodecyl sulfate; PAGE, polyacrylamide electrophoresis.

were fitted to the equation:

$$(k_{cat}/K_M)_{obs} = (k_{cat}/K_M)/([H^+]/K_1 + 1 + K_2/[H^+])$$

The pH stabilities of wild-type cathepsins L and K and its S2 subsite Ty67Leu/Leu205Ala mutant were studied at 28 °C in either 100 mM sodium acetate buffer, pH 5.5, or 100 mM potassium phosphate buffer, pH 6.5, or 100 mM Tris-HCl buffer, pH 7.5. All buffers contained 2.5 mM DTT and 2.5 mM EDTA. At appropriate time intervals, aliquots of the incubation mixture were withdrawn, and the activity was measured using 1 μM Z-Phe-Arg-MCA as substrate.

Collagen and Gelatin Digests. Active cathepsins K and L (600 nM) and Tyr67Leu/Leu205Ala mutant protein (600 nM) were incubated with 0.4 mg/mL soluble calf skin type I collagen (U.S. Biochemical Corp., Cleaveland, OH) or 0.6 mg/mL type II collagen from calf articular joints (Amersham Pharmacia Biotech, Piscataway, NJ) in 100 mM sodium acetate buffer, pH 5.5, containing 2.5 mM DTT and 2.5 mM EDTA. Collagen digestion was performed at 28 °C for 4 and 8 h in the presence of 0.15% (w/v) chondroitin 4-sulfate. The digestion reaction was stopped by the addition of 10 μM E-64. To measure the gelatinase activity of cathepsins K and L and the Tyr67Leu/Leu205Ala mutant, type I collagen was heated for 30 min at 70 °C prior to incubation with the proteases. In the presence of 10 nM cathepsin K, 10 nM cathepsin L, and 20 nM Tyr67Leu/Leu205Ala mutant, respectively, the reaction mix was incubated for 30 min at 28 °C. Collagen and gelatin digests were subjected to SDS−polyacrylamide electrophoresis using 4−20% Tris−glycine gels (Novex, San Diego, CA), and degradation products were visualized by Coomassie Blue staining.

Homology Modeling and Docking Calculation of the S2−P2 Interaction. The structures of human procathepsin L [PDB, 1cjl (*15*)] and cathepsin K with the vinyl sulfone inhibitor APC3328 [PDB, 1mem (*14*)] were used as templates to build a homology model of the S2 subsite cathepsin K mutant. Replacements of residues Tyr67Leu and Leu205Ala were obtained after energy minimizations (5000 iterations, 0.01 derivative) using Builder and Discover modules (Insight 2000 software; Molecular Simulations Inc., San Diego, CA). The structure of papain bound to a chloromethyl ketone inhibitor [PDB, 6pad (*21*)] was superimposed onto the model of the cathepsin K mutant and the wild-type cathepsin K structure to position the P2 Phe residue into the S2 subsite of the active site cleft of both enzymes. The inhibitor was kept as a seed group to replace Leu by Phe and Pro. Wild-type and mutant cathepsin K structures (keeping the backbone atoms fixed) were minimized to a convergence of the energy gradient less than 0.01 kcal·mol^{-1}·Å^{-1} using the TRIPOS force field (*22*) included in the SYBYL/MAXIMIN2 module (TRIPOS Associates Inc.). The minimization included electrostatic interactions based on Gasteiger partial charge distributions using a dielectric constant with a distance-dependent function $\epsilon = 1r$ and a nonbonded interaction cutoff of 8 Å (*23*).

Furthermore, both enzyme structures and the free ligands were minimized. For the estimation of the affinities of the three ligands to the enzymes, two methods were used. First, the interaction energies have been calculated by subtracting the energies of the free enzyme and of the free ligand from the corresponding enzyme−ligand complex. Second, using SCORE (*24*) the pK_d (which corresponds directly to the high-affinity binding: pK_i) was calculated. Residue volume (Å^3) was obtained from Rose et al. (*25*).

Table 1: Putative S2 Subsite Constituting Residues in Papain and in Cathepsins L and K and Its Tyr67Leu/Leu205Ala Mutant

residue[a]	papain	cathepsin L	cathepsin K	Tyr67Leu/ Leu205Ala
67	Tyr	Leu	Tyr	Leu
68	Pro	Met	Met	Met
133	Val	Ala	Ala	Ala
157	Val	Leu	Leu	Leu
160	Ala	Gly	Ala	Ala
205	Ser	Ala	Leu	Ala

[a] Papain numbering.

RESULTS AND DISCUSSION

Cathepsin K is responsible for the bulk degradation of type I collagen during bone remodeling. Absence of cathepsin K activity leads to an intralysosomal accumulation of undigested collagen fibrils in cathepsin K-deficient osteoclasts (*12*, *26*). Since various cathepsin K-catalyzed collagen cleavage sites have proline residues in the P2 position (*6*, *7*) and Gly-Pro-X is a repetitive motif in triple helical collagens, we were interested in studying the P2 substrate specificity of cathepsin K in detail and determining whether the binding of proline in the S2 subsite pocket is required for the intrahelical degradation of collagens by cathepsin K.

Engineering the Cathepsin K S2 Subsite Pocket into a Cathepsin-like Binding Pocket. The enzymatic specificity of papain-like cysteine proteases primarily depends on P2/S2 interactions (*27*). On the basis of several three-dimensional structures of papain-like proteases, the S2 binding pocket is formed by residues 67, 68, 133, 157, 161, and 205 (*28*). Protein sequence alignments revealed that the binding pockets of cathepsins K and L differ mostly in residues 67 and 205 (Table 1). Residue 67 is a tyrosine in cathepsin K and a leucine in cathepsin L whereas the hydrophobic 205 leucine residue in cathepsin K is replaced by a small alanine residue in cathepsin L. Analyses of 3-D structures of the S2 pockets in cathepsins K and L revealed significant differences in size and shape of the respective subsites. Residue 205, located at the bottom of the pocket, and residues 67, 133, and 157, forming the side part of the wall, tend to determine whether the pocket is larger or smaller. Met68, Ala133, Met/Leu157, and Gly/Ala160 are strictly conserved in most cysteine proteases, whereas residues 67 and 205 display significant variation. Cathepsins K and L display striking differences in their binding preferences for P2 residues. Whereas cathepsin L prefers aromatic residues and excludes proline in this position, cathepsin K readily accommodates leucine (*4*) and proline residues (*29*, *30*). Hence, residues 67 and 205 were selected as target sites for the modification of cathepsin K's S2 subsite specificity.

The recombinant Tyr67Leu/Leu205Ala cathepsin K mutant was expressed in *P. pastoris* as described in Experimental Procedures. The secreted mutant precursor protein was activated in the presence of pepsin at pH 4.0, and the proteolytic activity of the mature enzyme was monitored by the hydrolysis of Z-Phe-Arg-MCA. Maximal E-64 inhibitable Z-Phe-Arg-MCA hydrolysis was achieved after 40 min of pepsin-catalyzed activation at 37 °C. SDS−PAGE (12%

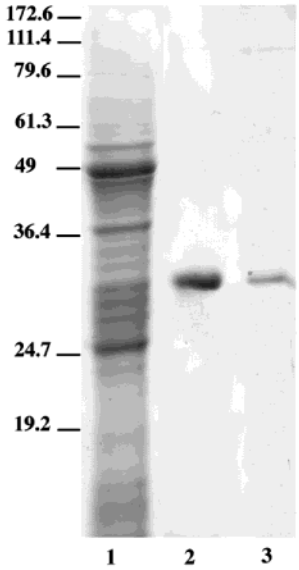

FIGURE 1: SDS−PAGE of the purified Tyr67Leu/Leu205Ala cathepsin K mutant (Coomassie Blue staining). Lanes: 1, crude supernatant fraction of the human Tyr67Leu/Leu205Ala mutant from *P. pastoris*; 2, purified Tyr67Leu/Leu205Ala mutant after passage through the SP-Sepharose column; 3, purified recombinant wild-type human cathepsin K. Molecular mass standards are indicated in the left lane.

Tris−glycine gel) of the active protein showed a single band after chromatography using a SP-Sepharose fast-flow column. The apparent molecular mass of 29 kDa for the mutant protein was identical to that of the mature wild-type cathepsin K as shown in Figure 1. Three liters of culture supernatant yielded 2 mL of a 1.25 μM (concentration determined by titration with E-64) enzyme solution after purification.

Profiling of the Tyr67Leu/Leu205Ala Mutant and Wild-Type Cathepsins K and L with Fluorogenic Substrate Libraries. The P1−P4 substrate specificities of the Tyr67Leu/Leu205Ala mutant and wild-type cathepsins K and L were determined. Assays of the cathepsins using P1 diverse libraries showed that arginine was the most favored amino acid at the P1 position (Figure 2, column P1) for all three activities (*28*). The overall P1 specificities of the cathepsins were similar to the P1 specificity of papain which had been determined using the same P1 library and reported previously (*20*). The order of preference for P1 residues was Arg > Lys > Nle, Gln > Met > Thr > Leu. Negatively charged (Glu, Asp), aromatic (Phe, Tyr, Trp, His), β-branched (Ile, Val, Pro), or very small amino acid residues (Ser, Ala, Gly) were only poorly or not accepted in the P1 position. Interestingly, P1 residues such as Glu, Ala, Ser, Gly, and Trp, which were not accommodated by the S1 subsite pocket of wild-type cathepsin K, allowed a low but significant hydrolysis of the substrates by the Tyr67Leu/Leu205Ala mutant. The relative hydrolysis rates for the cathepsin K mutant and wild-type cathepsin K were virtually identical (Figure 2, column P1), indicating that the S2 subsite mutation affected the P1 specificity only to a small degree.

Subsequently, P1−Arg fixed libraries were chosen to determine the P2−P4 substrate specificities of the cathepsins. This library revealed the specificity of the cathepsins at the P2 position to be the most strict among the P1−P4 positions. Cysteine proteases of the papain family have been shown to have primary substrate specificity at the P2 position and to have a preference for hydrophobic amino acids (*27*, *28*, *31*). Indeed, we observed that cathepsin K, its S2 subsite mutant, and cathepsin L exhibited preferences for hydrophobic amino acid residues. Cathepsin K, however, favored aliphatic amino acids and, most strikingly, proline (order of preference: Pro, Ile, Leu > Val, Nle) while cathepsin L favored both aromatic and aliphatic amino acid residues (Phe > Tyr > Trp, Val, Leu > Ile, Nle) at the P2 position (Figure 2, column P2). The assay of the cathepsin variant revealed that the mutations at the 67/205 positions caused a significant change of the P2 specificity of cathepsin K. The amino acid substitutions caused a shift of the specificity favoring aliphatic amino acids to favoring aromatic amino acids (Phe > Tyr ≫ Trp, Leu) as cathepsin L does. They also resulted in a total loss of the unique preference for proline at the P2 position. By changing two amino acids of cathepsin K to their counterparts in cathepsin L, cathepsin L-like substrate preferences are conferred to the cathepsin K variant.

The P3 specificities for all three enzymes were rather broad and showed a moderate preference for basic amino acid residues such as lysine and arginine (Figure 2, column P3). It should be noted that glycine in P3 is rather well accepted by cathepsin K when compared with the S3 specificity of cathepsin L. Glycine in P3 is part of the repetitive Gly-Pro-X motif in collagens. Unlike P2 and P1, however, almost all amino acids could be tolerated at the P3 position. The cathepsin K mutant specificity showed a slight increase in aromatic amino acids relative to both cathepsins K and L and revealed a loss in affinity to proline. The affinity toward glycine was not influenced by the mutations. The P4 specificity was also studied and shown to be much broader than the specificities of the P1, P2, and P3 positions and not to be affected by the S2 mutations (Figure 2, column P4).

Stability, pH Activity Profile, and Kinetic Analysis of the S2 Subsite Mutant of Cathepsin K. The overall catalytic activity of the mutant protease was not affected by the introduction of the double mutation as demonstrated by the comparable k_{cat}/K_M values for the most effective substrates for the mutant and wild-type enzymes (k_{cat}/K_M = 341756 $M^{-1}\cdot s^{-1}$ for wild-type cathepsin K toward Z-Leu-Arg-MCA; k_{cat}/K_M = 1089296 $M^{-1}\cdot s^{-1}$ for mutant cathepsin K toward Z-Phe-Arg-MCA). To determine a potential effect of the double mutation on the pH stability of the cathepsin K variant, the Tyr67Leu/Leu205Ala mutant and cathepsins K and L were incubated at pH 5.5, 6.5, and 7.5 at 28 °C, and the residual activities were monitored over time (Table 2). The pH stabilities of wild-type cathepsin K and the mutant protein were comparable for all three pH values tested, suggesting that the mutation did not impair the overall stability of the protease. For comparison, cathepsin L revealed increased instability at pH values close to neutrality as previously reported (*4*, *32*). In addition, pH activity profiles were analyzed to characterize the integrity of the ion pair of the active site. The pH profiles for the double mutant and wild-type cathepsin K showed no significant differences (Figure 3), indicating that the p*K* values of the catalytic groups, cysteine-25 and histidine-159, were largely unaffected. The pH activity profile of both enzymes was bell-shaped with flanking p*K* values of 3.91 ± 0.06 (pK_1) and 7.57 ± 0.07 (pK_2) for the Tyr67Leu/Leu205Ala variant. Values for wild-type cathepsin K were respectively 3.73 ± 0.06 and 7.55 ± 0.05. The very similar p*K* values between

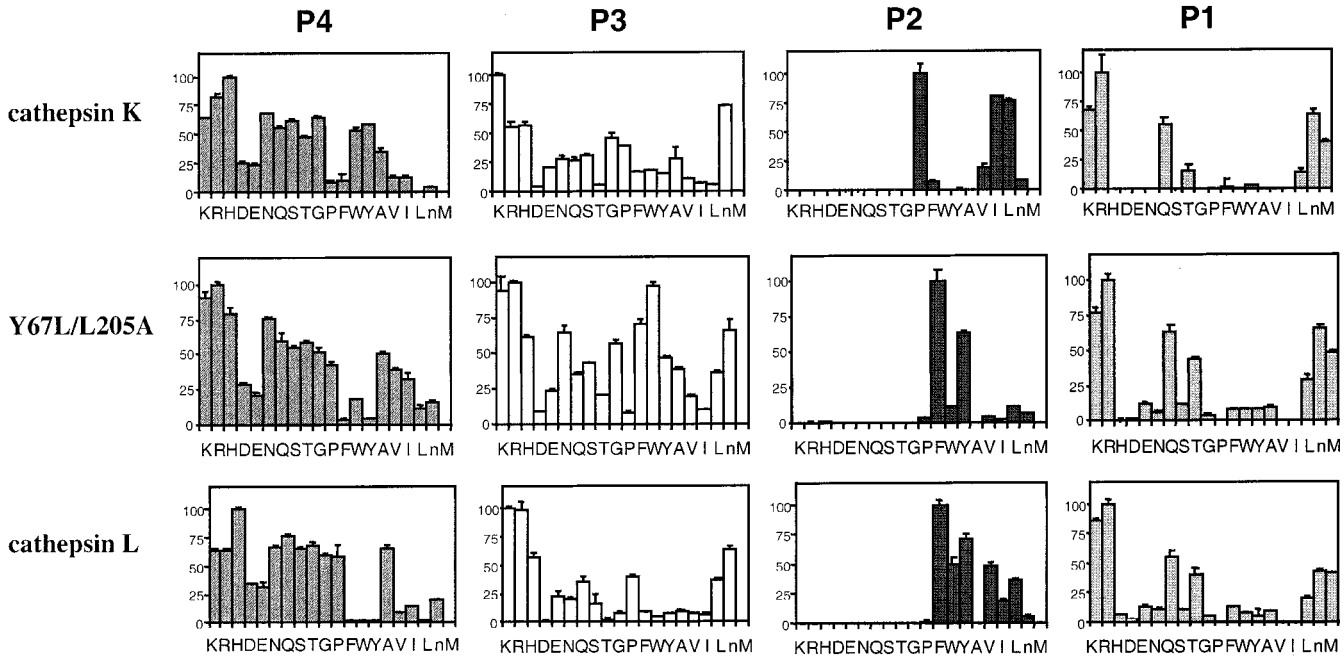

FIGURE 2: Substrate specificity profiles of cathepsin K, cathepsin K Tyr67Leu/Leu205Ala, and cathepsin L using positional scanning synthetic substrate libraries. The P1 diverse library and P2, P3, and P4 libraries were used for profiling P1−P4 specificities, respectively (shown in the P1, P2, P3, and P4 columns). The *y* axis is the relative activity compared to the highest activity of the library. The *x* axis provides the spatial address of the amino acids as represented by the one-letter code (with "n" representing norleucine). The error bars display the standard deviation from duplicate or triplicate experiments.

Table 2: pH Stability of the Tyr67Leu/Leu205Ala Mutant in Comparison with Recombinant Human Cathepsins K and L at 28 °C

protease	incubation time (h)	residual activity (%)		
		pH 5.5	pH 6.5	pH 7.5
Tyr67Leu/Leu205Ala	0.5	95	84	32
	1	59	39	22
	2	42	31	17
cathepsin K	0.5	92	60	59
	1	53	44	31
	2	51	21	8
cathepsin L	0.5	96	57	0
	1	87	15	0
	2	72	0	0

the mutant and wild-type enzyme are not surprising since the mutations did not involve charged residues and the S2 binding pocket is relatively distant from the catalytic site.

To confirm the change in substrate specificity observed with the tetrapeptide libraries for the cathepsin K variant, the hydrolysis of a series of di- and tripeptide substrates was evaluated. The kinetic parameters (K_M, k_{cat}, k_{cat}/K_M) of the cathepsin K mutant Tyr67Leu/Leu205Ala for the hydrolysis of synthetic substrates of the general structure Z-X-Arg-MCA (X is Phe, Leu, Val, Arg, or Gly-Pro) in comparison with cathepsins K and L are shown in Table 3. In agreement with the substrate library results, mutations introduced in cathepsin K dramatically changed the specificity of the S2 pocket. This was best reflected by an inversion of the K_M values for the P2 phenylalanine and leucine substrates. The mutation lowered the K_M value for the Phe substrate 5.3-fold and increased it 3.6-fold for the Leu substrate. Interestingly, the K_M values for the substrate series hydrolyzed by the Tyr67Leu/Leu205Ala mutant were similar to the K_M values observed for cathepsin L (Table 3). This trend was even more obvious for the tripeptide substrate Z-Gly-Pro-Arg-MCA, which displayed a K_M value >40 times higher than the best

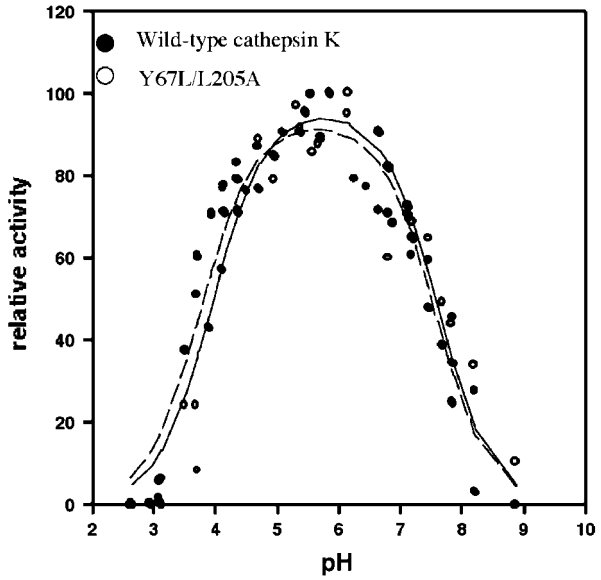

FIGURE 3: pH activity profile for the Tyr67Leu/Leu205Ala mutant of cathepsin K. Relative k_{cat}/K_M values for the hydrolysis of Z-Phe-Arg-MCA by cathepsin K and its S2 mutant were plotted against the pH values.

substrate, Z-Phe-Arg-MCA, for the Tyr67Leu/Leu205Ala mutant and cathepsin L. Despite a similar K_M value with cathepsin K for Z-GPR-MCA, the k_{cat} value was almost 3 orders of magnitude lower for the Tyr67Leu/Leu205Ala mutant and cathepsin L (Table 3). This is in agreement with previously reported findings (*29*, *30*) demonstrating that the proline residue in Z-Gly-Pro-Arg-MCA is well accepted in the S2 pocket of cathepsin K but not in that of cathepsin L. The k_{cat} values for the hydrolyses of the Z-Phe-Arg-MCA and Z-Leu-Arg-MCA substrates by the mutant enzyme doubled or halved, respectively, when compared with wild-type cathepsin K, indicating that the mutations did not

Table 3: Kinetic Parameters for Hydrolysis of Peptidyl-MCA Substrates by the Recombinant Human Cathepsin K Tyr67Leu/Leu205Ala Mutant and Wild-Type Cathepsins K and L[a]

	k_{cat} (s^{-1})	K_M (μM)	k_{cat}/K_M ($M^{-1}\cdot s^{-1}$)
Z-FR-MCA			
cathepsin K	0.73 ± 0.2	7.4 ± 0.8	98648
Tyr67Leu/Leu205Ala	1.52 ± 0.4	1.4 ± 0.8	1089296
cathepsin L	1.16 ± 0.3	0.8 ± 0.1	1446118
Z-LR-MCA			
cathepsin K	1.42 ± 0.01	4.2 ± 0.9	341756
Tyr67Leu/Leu205Ala	0.68 ± 0.3	15.0 ± 4.0	45333
cathepsin L	1.15 ± 0.2	12.7 ± 2.7	90551
Z-VR-MCA			
cathepsin K	0.58 ± 0.1	8.0 ± 1.6	72500
Tyr67Leu/Leu205Ala	0.20 ± 0.04	12.5 ± 2.5	18173
cathepsin L	1.05 ± 0.2	8.5 ± 1.0	123529
Z-RR-MCA			
cathepsin K	0.0025 ± 0.0005	25.0 ± 0.8	100
Tyr67Leu/Leu205Ala	0.04 ± 0.008	22.0 ± 4.5	1818
cathepsin L	0.07 ± 0.008	40.0 ± 10.0	1750
Z-GPR-MCA			
cathepsin K	3.8 ± 1.0	32.0 ± 14.0	118750
Tyr67Leu/Leu205Ala	0.007 ± 0.0005	57.0 ± 8.0	122
cathepsin L	0.011 ± 0.002	81.0 ± 26.0	135

[a] Experiments were carried out as described in Experimental Procedures. The data corresponded to triplicate experiments and were averaged ± SD.

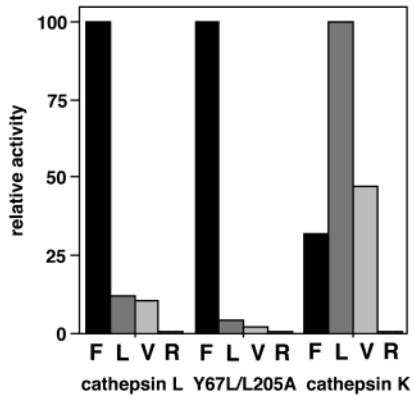

FIGURE 4: k_{cat}/K_M values for the hydrolysis of Z-XR-MCA by cathepsins L, K, and the Tyr67Leu/Leu205Ala cathepsin K mutant (normalized to the best substrate = 1): cathepsin K (Z-LR-MCA), 122632 $M^{-1}\cdot s^{-1}$; cathepsin L (Z-FR-MCA), 1446118 $M^{-1}\cdot s^{-1}$; Tyr67Leu/Leu205Ala mutant (Z-FR-MCA), 1089296 $M^{-1}\cdot s^{-1}$.

indiscriminately affect the catalytic site of the protease. Analysis of the second-order rate constant (k_{cat}/K_M) showed that the Tyr67Leu/Leu205Ala mutant specificity was comparable with that of cathepsin L (Figure 4).

The importance of residue Tyr67 for the accommodation of proline residues also was shown for cathepsin L-like proteases expressed by *Fasciola hepatica*. Cathepsin L2 from *F. hepatica* exhibited substantial activity against substrates with proline in the P2 position whereas other related cathepsin L-like proteases of the same organism lacked this specificity. Similar to cathepsin K, cathepsin L2 possesses a tyrosine residue in position 67. The replacement of Leu67 by tyrosine in cathepsin L5 significantly increased the ability of this variant to accommodate substrates with a proline residue in the P2 position (*33*).

Homology Model of the Tyr67Leu/Leu205Ala Mutant. A molecular model of the Tyr67Leu/Leu205Ala mutant was built on the basis of the crystal structures of human cathepsins K and L. Examination of the S2 pockets in both enzymes revealed differences in their sizes. The replacement of Tyr67 by a leucine residue resulted in an increase of the distance between the CA of Leu157 and the CD of Leu67 by 1.5 Å when compared with the distance between the OH group of Tyr67 (wild type) and CA of Leu157, thus increasing the volume of the S2 site in the cathepsin K variant. Thus, bulky aromatic residues could fit into the S2 subsite of the Tyr67Leu/Leu205Ala mutant while the size of the S2 subsite of wild-type cathepsin K was sterically restricted by the Tyr67 residue (Figure 5).

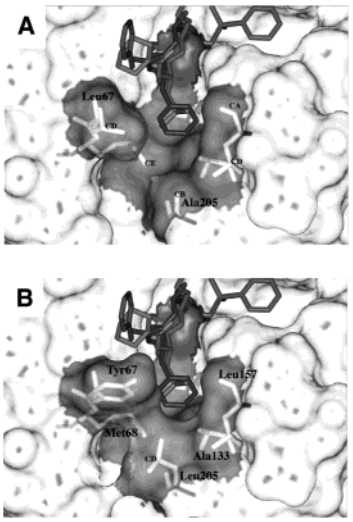

FIGURE 5: Connolly surface representation of the modeled S2 pocket of the Tyr67Leu/Leu205Ala mutant (A) compared to wild-type cathepsin K (B). The S2 pocket is displayed by transparency in dark gray. Residues constituting the S2 pocket are in white stick representation. The P2 Phe (dark gray) or Leu (light gray) residues are superimposed and derived from 6pad and 1mem structures, respectively.

Table 4: Computational Parameters of Pro, Leu, and Phe Residues into the S2 Pocket of Wild-Type Cathepsin K and the Tyr67Leu/Leu205Ala Mutant

			wild-type cathepsin K		Tyr67Leu/Leu205Ala	
P2 residues	log *P*	vol (Å^3)	interaction energy (kcal/mol)	pK_i	interaction energy (kcal/mol)	pK_i
Pro	1.128	112.7	−61.6	5.48	−68.2	5.04
Leu	2.167	166.7	−60.2	5.26	−70.6	5.75
Phe	2.423	189.9	−59.7	4.47	−71.6	6.69

Since the kinetic data discussed above revealed major differences between the binding affinities for Pro, Leu, and Phe residues inside the S2 pocket of both enzymes, docking calculations of these amino acids were performed. Theoretical volumes and hydrophobicities of the three amino acids and their interaction energies and pK_i values with the S2 subsite pockets of wild-type cathepsin K and its mutant protein are listed in Table 4. A correlation was obtained between docking score (interaction energies and pK_i values) of Pro, Leu, and Phe residues and their theoretical value of volume (Å^3) as well as with their hydrophobicity properties (expressed as a partition coefficient in octanol/water: log *P*). The appropriate pK_i values for the wild-type enzyme decreased with increasing size and hydrophobicity of the ligand whereas the trend was reversed in the mutant enzyme. The considerably lower

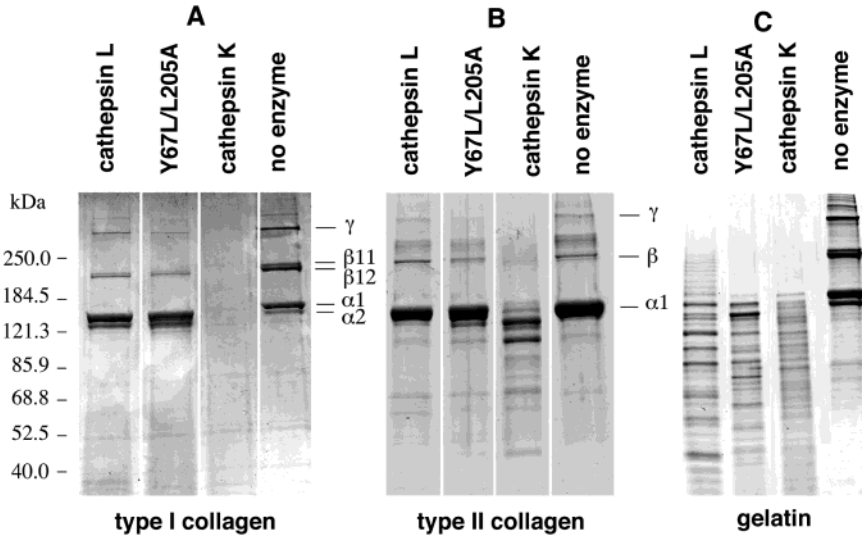

FIGURE 6: Comparison of collagenolytic and gelatinolytic activities of cathepsins L and K and its Tyr67Leu/Leu205Ala mutant. (A) Soluble type I or (B) type II bovine collagens were incubated in the presence of 0.15% chondroitin 4-sulfate at 28 °C with 600 nM respective cathepsin activities for 8 h. The gelatinolytic activities (C) were monitored by incubating heat-denatured type I collagen with the respective cathepsin activities for 30 min at 28 °C. Samples were analyzed by 4−20% SDS−PAGE. Undigested type I (A) or type II collagen (B) was used as a standard. Molecular mass standards are indicated in the left lane.

pK_i value for proline and the much higher value for phenylalanine in mutant cathepsin K reflects the decreased affinity for proline and the increased affinity for phenylalanine as observed in the kinetic analysis of the appropriate peptide substrates (Table 3).

Activity of the Tyr67Leu/Leu205Ala Mutant on Type I and II Collagens and Gelatin. Analyses of cathepsin K catalyzed cleavage sites revealed the repeated presence of proline residues in the P2 position with regard of the hydrolyzed bond. Our substrate specificity data derived from combinatorial peptide libraries clearly documented a unique preference for proline residues in the P2 position, a preference not observed in structurally related cathepsin L and a preference which could be deleted by the engineering of a cathepsin L-like S2 subsite pocket into cathepsin K. To investigate the ability of the Tyr67Leu/Leu205Ala cathepsin K variant to cleave interstitial collagens, type I and II triple helical collagens were incubated with the 600 nM mutant enzyme, and its activity was compared with those of cathepsins K and L at equal enzyme concentrations. To mimic the abundant presence of glycosaminoglycans in bone and cartilage, the degradation assays were performed in the presence of chondroitin 4-sulfate, a known activator of the collagenase activity of cathepsin K (*34*). Contrary to wild-type cathepsin K, which possessed a potent collagenolytic activity at acidic pH toward both collagens, the Tyr67Leu/Leu205Ala mutant and cathepsin L did not display any intrahelical collagenolytic activity (Figure 6A,B). All three enzymes displayed a residual activity of approximately 10% after 8 h incubation with the appropriate collagens, indicating comparable enzyme stabilities (determined with the synthetic substrate, Z-Phe-Arg-MCA; data not shown). Garnero et al. (*7*) have shown that, even at enzyme concentrations as high as 2.14 μM, cathepsin L does not cleave inside triple helical collagens. However, similar to cathepsin L (*4*, *35*), the Tyr67Leu/Leu205Ala mutant is able to cleave the γ and β forms of native type I and II collagens at pH 5.5 into the α forms by cleaving in the nonhelical telopeptide region. The disappearance of the γ and β forms after incubation with the Tyr67Leu/Leu205Ala mutant and a minor shift of the α chains in molecular weight suggest a cleavage activity for the mutant enzyme in the telopeptide region (Figure 6A). In contrast to its inability to cleave triple helical collagens, the Tyr67Leu/Leu205Ala variant of cathepsin K retained a strong gelatinolytic activity toward denatured type I collagen at pH 5.5 which was comparable to those of wild-type cathepsins K and L (Figure 6C). As expected, the pattern of gelatin cleavage products changed from wild-type cathepsin K to the mutant protein, reflecting the alteration of the specificity of the S2 binding pocket. Interestingly, some of the major cleavage products generated by the mutant protease migrated similarly to the products generated by cathepsin L, suggesting that the introduction of the cathepsin L-like subsite pocket into cathepsin K is also reflected in its specificity toward protein substrates. All three protease activities were capable of completely degrading gelatin into nondetectable small fragments at increased enzyme concentrations or extended incubation times. These findings clearly suggest that the S2 subsite specificity of cathepsin K is required for the cleavage of peptide bonds within the triple helical domain of collagens. The preference of the proline residue for the S2 subsite in cathepsin K indicates that the active site of cathepsin K specifically recognizes the Gly-Pro-X motif (with proline in P2 and glycine in P3) in the helical regions of collagens and that its triple helical collagenase activity requires the accommodation of proline (or hydroxyproline) residues in the S2 subsite. The exclusion of proline from the S2 subsite of cathepsin L may explain the enzyme's inability to cleave peptide bonds in the proline-rich helical area of collagens. However, it should be noted that the collagenolytic activity of cathepsin K requires additional structural features besides a proline residue accommodating the binding site. We have recently demonstrated that a pycnodysostosis-causing mutation in the cathepsin K gene (Y212C mutation) results in a catalytically active enzyme with gelatinase but no collagenase activity (*18*). This mutation, which is distant from its substrate binding cleft, is unlikely to affect the direct binding of the triple helical domain of collagen. Preliminary data

indicate that the collagenolytic activity of cathepsin K also depends on specific interactions between the protease and glycosaminoglycans which are interrupted by the Y212C mutation (*34*; D. Brömme, unpublished data). These interactions are presently being investigated in our laboratory.

In conclusion, we have demonstrated that cathepsin K exhibits a preference for proline residues in the P2 substrate position and that this unique affinity can be relinquished by altering two amino acids in the S2 subsite pocket of cathepsin K into residues present in the cathepsin L S2 subsite. This mutant completely lost the ability of wild-type cathepsin K to hydrolyze intrahelical peptide bonds in interstitial collagens whereas it retained its gelatinolytic activity. The identification of the P2 proline specificity as a critical requirement for the enzyme's collagenase activity may provide scaffolds for the design of highly selective antiresorptive cathepsin K inhibitors which may surpass the selectivity of drug candidates for cathepsin K with a leucine moiety (*36*).

REFERENCES

1. Delaisse, J. M., and Vaes, G. (1992) *Mechanism of mineral solubilization and matrix degradation in osteoclastic bone resorption*, CRC Press, Boca Raton, Ann Arbor, London, and Tokyo.
2. Beck, K., and Brodsky, B. (1998) *J. Struct. Biol. 122*, 17−29.
3. Overall, C. M. (2001) *Methods Mol. Biol. 151*, 79−120.
4. Brömme, D., Okamoto, K., Wang, B. B., and Biroc, S. (1996) *J. Biol. Chem. 271*, 2126−2132.
5. Welgus, H. G., Jeffrey, J. J., and Eisen, A. Z. (1981) *J. Biol. Chem. 256*, 9511−9515.
6. Kafienah, W., Bromme, D., Buttle, D. J., Croucher, L. J., and Hollander, A. P. (1998) *Biochem. J. 331*, 727−732.
7. Garnero, P., Borel, O., Byrjalsen, I., Ferreras, M., Drake, F. H., McQueney, M. S., Foged, N. T., Delmas, P. D., and Delaisse, J. M. (1998) *J. Biol. Chem. 273*, 32347−32352.
8. Kirschke, H., Kembhavi, A. A., Bohley, P., and Barrett, A. J. (1982) *Biochem. J. 201*, 367−372.
9. Bromme, D., and Okamoto, K. (1995) *Biol. Chem. Hoppe-Seyler 376*, 379−384.
10. Drake, F. H., Dodds, R. A., James, I. E., Connor, J. R., Debouck, C., Richardson, S., Lee-Rykaczewski, E., Coleman, L., Rieman, D., Barthlow, R., Hastings, G., and Gowen, M. (1996) *J. Biol. Chem. 271*, 12511−12516.
11. Tezuka, K., Tezuka, Y., Maejima, A., Sato, T., Nemoto, K., Kamioka, H., Hakeda, Y., and Kumegawa, M. (1994) *J. Biol. Chem. 269*, 1106−1109.
12. Gelb, B. D., Shi, G. P., Chapman, H. A., and Desnick, R. J. (1996) *Science 273*, 1236−1238.
13. Schechter, I., and Berger, A. (1967) *Biochem. Biophys. Res. Commun. 27*, 157−162.
14. McGrath, M. E., Klaus, J. L., Barnes, M. G., and Brömme, D. (1997) *Nat. Struct. Biol. 4*, 105−109.
15. Coulombe, R., Grochulski, P., Sivaraman, J., Menard, R., Mort, J. S., and Cygler, M. (1996) *EMBO J. 15*, 5492−5503.
16. Brömme, D., Steinert, A., Friebe, S., Fittkau, S., Wiederanders, B., and Kirschke, H. (1989) *Biochem. J. 264*, 475−481.
17. Linnevers, C. J., McGrath, M. E., Armstrong, R., Mistry, F. R., Barnes, M., Klaus, J. L., Palmer, J. T., Katz, B. A., and Brömme, D. (1997) *Protein Sci. 6*, 919−921.
18. Hou, W.-S., Brömme, D., Zhao, Y., Mehler, E., Dushey, C., Weinstein, H., Miranda, C. S., Fraga, C., Greig, F., Carey, J., Rimoin, D. L., Desnick, R. J., and Gelb, B. D. (1999) *J. Clin. Invest. 103*, 731−738.
19. Barrett, A. J., Kembhavi, A. A., Brown, M. A., Kirschke, H., Knight, C. G., Tamai, M., and Hanada, K. (1982) *Biochem. J. 201*, 189−198.
20. Harris, J. L., Backes, B. J., Leonetti, F., Mahrus, S., Ellman, J. A., and Craik, C. S. (2000) *Proc. Natl. Acad. Sci. U.S.A. 97*, 7754−7759.
21. Drenth, J., Kalk, K. H., and Swen, H. M. (1976) *Biochemistry 15*, 3731−3738.
22. Clark, M., Cramer III, R. D., and van Opdenbosch, N. (1989) *J. Comput. Chem. 10*, 982−1012.
23. Gasteiger, J., and Marsili, M. (1980) *Tetrahedron 36*, 3219−3228.
24. Wang, R., Liu, L., Lai, L., and Tang, Y. (1998) *J. Mol. Model. 4*, 379−394.
25. Rose, G. D., Geselowitz, A. R., Lesser, G. J., Lee, R. H., and Zehfus, M. H. (1985) *Science 229*, 834−838.
26. Everts, V., Aronson, D. C., and Beertsen, W. (1985) *Calcif. Tissue Int. 37*, 25−31.
27. Turk, D., Guncar, G., Podobnik, M., and Turk, B. (1998) *Biol. Chem. 379*, 137−47.
28. McGrath, M. E. (1999) *Annu. Rev. Biophys. Biomol. Struct. 28*, 181−204.
29. Aibe, K., Yazawa, H., Abe, K., Teramura, K., Kumegawa, M., Kawashima, H., and Honda, K. (1996) *Biol. Pharm. Bull. 19*, 1026−1031.
30. Xia, L., Kilb, J., Wex, H., Lipyansky, A., Breuil, V., Stein, L., Palmer, J. T., Dempster, D. W., and Brömme, D. (1999) *Biol. Chem. 380*, 679−687.
31. Rawlings, N. D., and Barrett, A. J. (1994) in *Methods in Enzymology* (Barrett, A. J., Ed.) pp 461−486, Academic Press, San Diego.
32. Turk, B., Dolenc, I., Turk, V., and Bieth, J. G. (1993) *Biochemistry 32*, 375−380.
33. Smooker, P. M., Whisstock, J. C., Irving, J. A., Siyaguna, S., Spithill, T. W., and Pike, R. N. (2000) *Protein Sci. 9*, 2567−2572.
34. Li, Z., Hou, W. S., and Bromme, D. (2000) *Biochemistry 39*, 529−536.
35. Maciewicz, R. A., and Etherington, D. J. (1988) *Biochem. J. 256*, 433−440.
36. Thompson, S. K., Halbert, S. M., Bossard, M. J., Tomaszek, T. A., Levy, M. A., Zhao, B., Smith, W. W., Abdel-Meguid, S. S., Janson, C. A., D'Alessio, K. J., McQueney, M. S., Amegadzie, B. Y., Hanning, C. R., DesJarlais, R. L., Briand, J., Sarkar, S. K., Huddleston, M. J., Ijames, C. F., Carr, S. A., Garnes, K. T., Shu, A., Heys, J. R., Bradbeer, J., Zembryski, D., and Lee-Rykaczewski, L. (1997) *Proc. Natl. Acad. Sci. U.S.A. 94*, 14249−14254.

BI025638X