

Characterization of novel cathepsin K mutations in the pro and mature polypeptide regions causing pycnodysostosis

Wu-Shiun Hou,¹ Dieter Brömme,¹ Yingming Zhao,¹ Ernest Mehler,³ Craig Dushey,¹ Harel Weinstein,³ Clara Sa Miranda,⁴ Claudia Fraga,⁵ Fenella Greig,² John Carey,⁶ David L. Rimoin,⁷ Robert J. Desnick,^{1,2} and Bruce D. Gelb^{1,2}

¹Department of Human Genetics,

²Department of Pediatrics, and

³Department of Physiology, Mount Sinai School of Medicine, New York, New York 10029, USA

⁴Instituto de Genética Médica, Jacinto de Magalhães, Porto 4150, Portugal

⁵Hospital do Funchal, Madeira 9050, Portugal

⁶Department of Pediatrics, University of Utah, Salt Lake City, Utah 84112, USA

⁷Department of Pediatrics, Cedars-Sinai Medical Center, Los Angeles, California 90048-1804, USA

Address correspondence to: Bruce D. Gelb, Mount Sinai School of Medicine, One Gustave L. Levy Place, New York, New York 10029, USA. Phone: (212) 241-8662; Fax: (212) 360-1809; E-mail: gelb@msvax.mssm.edu

Received for publication May 19, 1997, and accepted in revised form January 15, 1999.

Cathepsin K, a lysosomal cysteine protease critical for bone remodeling by osteoclasts, was recently identified as the deficient enzyme causing pycnodysostosis, an autosomal recessive osteosclerotic skeletal dysplasia. To investigate the nature of molecular lesions causing this disease, mutations in the cathepsin K gene from eight families were determined, identifying seven novel mutations (K52X, G79E, Q190X, Y212C, A277E, A277V, and R312G). Expression of the first pro region missense mutation in a cysteine protease, G79E, in *Pichia pastoris* resulted in an unstable precursor protein, consistent with misfolding of the proenzyme. Expression of five mature region missense defects revealed that G146R, A277E, A277V, and R312G precursors were unstable, and no mature proteins or protease activity were detected. The Y212C precursor was activated to its mature form in a manner similar to that of the wild-type cathepsin K. The mature Y212C enzyme retained its dipeptide substrate specificity and gelatinolytic activity, but it had markedly decreased activity toward type I collagen and a cathepsin K-specific tripeptide substrate, indicating that it was unable to bind collagen triple helix. These studies demonstrated the molecular heterogeneity of mutations causing pycnodysostosis, indicated that pro region conformation directs proper folding of the proenzyme, and suggested that the cathepsin K active site contains a critical collagen-binding domain.

J. Clin. Invest. 103:731–738 (1999)

Introduction

Pycnodysostosis (Pycno), an autosomal recessive sclerosing skeletal dysplasia, recently was shown by a positional candidacy approach to result from the deficient activity of the lysosomal cysteine protease cathepsin K (EC 3.4.22.38; ref. 1). The disease is characterized by reduced stature, osteosclerosis, acro-osteolysis of the distal phalanges, frequent fractures, clavicular dysplasia, and skull deformities with delayed suture closure (2, 3). The fact that Pycno resulted from deficient cathepsin K activity proved the important physiological role of this lysosomal cysteine protease in bone matrix degradation and implied that inhibition of cathepsin K activity might be therapeutic for bone diseases characterized by excessive bone degradation, such as osteoporosis and certain forms of arthritis.

The cathepsin K gene, which was cloned originally from rabbit osteoclasts (4), and subsequently from several human tissues (5–8), was highly expressed in osteoclasts, the site of the Pycno pathology (9). The predicted 329-residue polypeptide sequence was highly homologous with those of cathepsins S and L (6, 7) and had the typical prepropeptide organization of cysteine proteases of

the papain family, including a 15-amino acid signal sequence and a 99-residue pro piece. A putative *N*-glycosylation site in the pro region, which is conserved among cathepsin K genes from all species sequenced to date (10), presumably facilitates lysosomal targeting of the proenzyme via the mannose 6-phosphate receptor pathway. The catalytic triad within the active site contains active cysteine, histidine, and asparagine residues that are conserved completely among all papain family members (6).

The human cathepsin K gene has been overexpressed in the baculovirus system, and the purified recombinant enzyme's physical and kinetic properties have been determined (11, 12). Cathepsin K was synthesized as a prepropeptide of 37 kDa, and the mature enzyme was a monomeric protein with an apparent molecular mass of about 29 kDa. The enzyme had a broad bell-shaped pH-activity profile with an optimum at 6.1 and had strong collagenolytic, elastase, and gelatinase activities, exceeding those of cathepsins S or L (11). These properties are consistent with the enzyme's ability to degrade organic matrix proteins in the subosteoclastic space.

Previously, three mutations in the mature region of the cathepsin K prepropeptide were reported in four unre-

lated Pycno families, including a missense (G146R), a nonsense (R241X), and a stop codon (X330W) mutation (1, 13). Transient expression of the X330W allele resulted in no immunologically detectable protein, despite normal message levels (1).

In this report, seven novel cathepsin K mutations are described, comprising four mature region missense mutations, a mature nonsense mutation, and two pro region point mutations. A nonsense mutation in the mature enzyme was found in several unrelated Hispanic Pycno families and suggested a common Iberian descent. Six missense mutants, expressed in *Pichia pastoris*, lacked detectable residual activity, with the exception of Y212C. This mutant had residual activity *in vitro* against the fluorescent substrate and gelatin, but the mature Y212C enzyme had little, if any, activity toward type I collagen.

Methods

Specimen collection and DNA sequencing. Blood samples were obtained with informed consent from eight unrelated families with one or more individuals affected with Pycno. The ethnic background of these families was as follows: northern European (two), Czech, Spanish, Portuguese (two), Indian, and Honduran. Genomic DNA was extracted from blood leukocytes using the Puregene Genomic DNA Isolation kit (Gentra Systems, Minneapolis, Minnesota). Exons 2–8 of the cathepsin K gene (14) were amplified from genomic DNA of Pycno patients by PCR, isolated, and sequenced by cycle sequencing with an ABI 377 Sequencer (Perkin-Elmer Corp., Norwalk, Connecticut, USA).

Confirmation of detected mutations. Putative cathepsin K mutations were confirmed by PCR-based analyses of normal, obligate heterozygous, and/or affected members from each family. In addition, each putative mutation was analyzed in genomic DNAs isolated from 50–55 unrelated normal Caucasian individuals to rule out possible polymorphisms. PCR products containing the mutation were amplified from genomic DNA, digested overnight with the appropriate restriction endonuclease (New England Biolabs Inc., Beverly, Massachusetts, USA), separated by horizontal electrophoresis in 2% agarose gels, and directly viewed with ethidium bromide. The five mutations that created or destroyed restriction sites were as follows: R241X (*Ava*I), Y212C (*Rsa*I), A277E (*Acc*I), A277V (*Acc*I), and R312G (*Taq*I). For the three mutations that neither created nor destroyed convenient restriction sites, either mismatch PCR (*Nla*IV for G79E; *Pst*I for Q190X) or the amplification refractory mutation system (ARMS) (K25X) was used for mutation detection (15, 16).

Expression and characterization of missense mutations. The human procathepsin K cDNA that had been cloned in pBlue-script SK(+) phagemid and mutated to eliminate the N-glycosylation site in the mature enzyme was used to generate constructs for the six Pycno missense mutations. Each point mutation was introduced individually into the cDNA using the PCR ligation method with Vent *Taq* polymerase (New England Biolabs Inc.), and each mutant construct was confirmed by sequencing. Unique *Eco*RI and *Not*I sites, introduced at the 5' and 3' ends of each amplified mutant proenzyme PCR product, were used to ligate each construct into the pPIC9K expression vector (Invitrogen Corp., Carlsbad, California, USA). The expression constructs for the wild-type and the six mutant cathepsin K (gly-) proenzymes were linearized with *Bgl*II and then electroporated into *P. pastoris* GS115 cells (Invitrogen Corp.). After genotype selection and phenotype screening according to the manufacturer's instructions, several clones were obtained for each of the constructs.

Clones were grown up in shaker flasks, and the liquid culture media were concentrated with a spin column (Amicon Inc., Beverly, Massachusetts, USA). The presence of the cathepsin K (gly-) proenzyme was detected by Western blot, using a rabbit polyclonal antibody raised against human cathepsin K (MS2) (provided by D. Brömme). Standard Northern analysis was performed after extraction of total RNA from lyticase-treated cells with RNazol (Tel-Test Inc., Friendswood, Texas, USA). Blots were hybridized with human cathepsin K cDNA and *Saccharomyces cerevisiae* actin (a gift from A. Caplan, Mount Sinai School of Medicine) that had been radiolabeled with [³²P]-dCTP (Amersham Pharmacia Biotech, Piscataway, New Jersey, USA) by the random hexamer method.

To assess cathepsin K enzyme activity, the concentrated medium with the highest yield of proenzyme for each construct was selected. The mature cathepsin K (gly-) proteins were activated from their precursor forms by diluting the medium to 2.5 volumes, to a final solution containing 50 mM sodium acetate, 2.5 mM dithiothreitol, and 2.5 mM EDTA (pH 4.0); adding porcine pepsin (Sigma Chemical Co., St. Louis, Missouri, USA) to a final concentration of 20 µg/ml; and then incubating the mixture at 37°C for 30 min.

The enzymatic activity of the wild-type and mutant cathepsin K (gly-) proteins was assayed using the fluorescent substrate Z-Leu-Arg-MCA at 5 µM in an assay buffer of 100 mM sodium acetate, 2.5 mM EDTA, 2.5 mM dithiothreitol (pH 5.5). Fluorescence was detected by a luminescence spectrometer (Perkin-Elmer Corp.), and enzyme kinetics were obtained by nonlinear regression analysis using the Grafit software package (Erihtacus Software Ltd., Staines, United Kingdom). To establish the pH profile for active cathepsin K enzymes, initial rates of substrate hydrolysis with 5 µM Z-Leu-Arg-MCA were determined in 100 mM sodium citrate for pH 2.8–5.6 and in 100 mM sodium phosphate for pH 5.8–9.7. The pH stability of wild-type and active mutant proteins was studied by incubating at 37°C in 100 mM sodium acetate buffer (pH 4.5 and 5.5) or in 100 mM potassium phosphate buffer (pH 6.5), containing 2.5 mM dithiothreitol and 2.5 mM EDTA. At several time points, the residual activity was determined using 5 µM Z-Leu-Arg-MCA as substrate. Substrate specificity of mutant enzymes with residual activity was determined by measuring enzyme kinetics with a panel of synthetic di- and tripeptide substrates. In addition, the mutant enzymes with residual activity against the synthetic substrates were also assayed with biologically relevant substrates. Activated wild-type and mutant enzymes were incubated at 28°C for 12 h with 0.4 mg/ml of soluble calf skin type I collagen (U.S. Biochemical Corp., Cleveland, Ohio) in 100 mM sodium acetate buffer (pH 5.5) containing 2.5 mM dithiothreitol and 2.5 mM EDTA. To measure the gelatinase activity, type I collagen was heated at 68°C for 15 min and then incubated with varying concentrations of the activated wild-type and mutant enzymes for 1 h at 28°C. Samples from the collagenase and gelatinase assays were subjected to SDS-PAGE, fixed with trichloroacetate, and viewed by Coomassie blue staining.

Determination of the processing sites of Y212C and wild-type cathepsin K enzymes. To determine the site(s) of processing of the cathepsin K propeptides, mature Y212C and wild-type cathepsin K proteins were separated on an SDS-PAGE gel and viewed by the colloidal blue staining method (Novex, San Diego, California, USA). The bands of interest were excised and subjected to in-gel digestion with trypsin or endopeptidase Lys-C, and the resulting proteolytic mixtures were extracted as described previously (17, 18). The molecular masses of the peptide mixtures were accurately determined by matrix-assisted laser desorption/ionization time-of-flight (MALDI-TOF) mass spectrometry using delayed ion extraction and ion reflection (Voyager-DE STR; PerSeptive Biosystems Inc., Framingham, Massachusetts, USA). NH₂- and COOH-terminal peptides were identified by

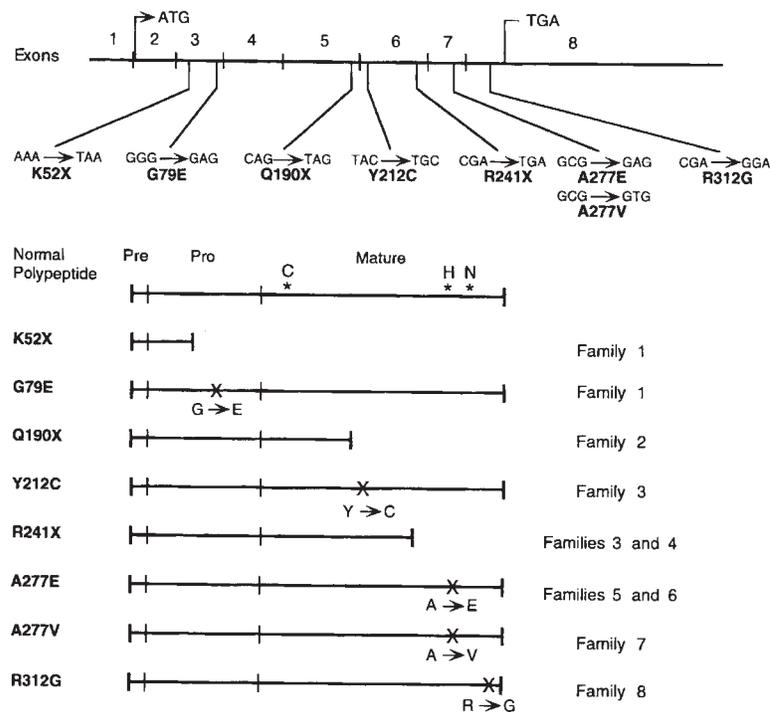


Figure 1

Novel pycnodysostosis mutations in the cathepsin K gene. DNA changes to codons and their predicted effects on the cathepsin K protein are indicated above and positioned on the cathepsin K gene, which is divided into eight exons. The initiation and stop codons are indicated in exons 2 and 8, respectively. The effects of these mutations on the cathepsin K prepropeptide are shown below. The positions of the cysteine, histidine, and asparagine residues that form the active triad in the active cleft are indicated on the normal polypeptide. The families in which individual mutations were found are indicated on the right.

comparing the measured masses with those calculated from the known protein sequence and the cleavage rules of the protease, using the software tool PepFound (19, 20). The mass accuracy, determined using internal calibration, was within 0.0001 kDa.

Results

Mutation detection and confirmation. Analysis of the cathepsin K coding region and adjacent intron/exon boundaries from genomic DNAs of patients from eight unrelated Pycno families revealed eight mutations, including seven novel lesions (Fig. 1). These mutations included two pro region changes consisting of a nonsense and a missense lesion, and six mature region changes consisting of two nonsense and four missense mutations. Of note, no readily apparent genotype-phenotype correlation was identified, because all patients appeared to be similarly affected.

To rule out the possibility that these cathepsin K base substitutions could be polymorphisms and not disease-causing mutations, alleles from ≥ 50 normal Caucasian individuals were analyzed for the respective lesions. Using the appropriate PCR-based assays, no K52X, G79E, Q190X, Y212C, R241X, or R312G alleles were detected among greater than 100 control alleles. In one control individual, the restriction assay for the A277E and A277V mutations produced a pattern identical to the A277E and A277V heterozygotes, but sequencing of the PCR product revealed only heteroallelism for a G→A change at nucleotide 936, which did not alter the amino acid

sequence, rather than the A277E and A277V mutations, which altered nucleotide 935. Thus, the possibility that these eight coding sequence changes were polymorphisms rather than mutations was considered highly unlikely.

Expression analysis of cathepsin K missense mutations. The wild-type and mutant cathepsin K proenzymes were expressed in the *P. pastoris* system, which has been shown previously (21) to result in high-level production and secretion into the medium (38.4 mg/l of purified active enzyme). In addition, as revealed by the completed *S. cerevisiae* genome project, yeast strains contain no papain-like cysteine proteases, except bleomycin hydrolase, which is an aminopeptidase with no overlap in substrate specificity. Coomassie blue staining of sodium dodecyl sulfate-polyacrylamide electrophoretic gels containing aliquots of media from the G79E, G146R, Y212C, A277E, A277V, and R312G constructs revealed the presence of the expected 43-kDa bands, which were quantitatively less than the band for the wild-type protein (data not shown). Immunologically detectable proteins of the expected mass from each mutant precursor were observed (Fig. 2), but, excepting Y212C, with reduced quantities relative to the wild-type control. In addition, degradative products of smaller mass were noted in the wild-type and Y212C media. Western blot analysis of whole cell lysates revealed a similar discrepancy in cathepsin K production, excluding a failure to secrete the mutant precursors (data not shown). Only supernatant from wild-type had a band of the expected mass (30 kDa)

Table 1
pH stability at 37°C of recombinant Y212C mutant enzyme compared with wild-type cathepsin K

Protease	Incubation time (h)	Residual activity (%)		
		pH 4.5	pH 5.5	pH 6.5
Y212C	0.5	67	46	8
	1	46	29	4
	2	32	6	0
	4	21	0	0
Wild-type	0.5	70	77	46
	1	40	63	32
	2	15	42	12
	4	5	28	8

of the mature enzyme, presumably generated by autoactivation. Northern analysis of total RNA extracted from the yeast revealed approximately equal levels of the cathepsin K transcripts in wild-type and mutant clones (data not shown), strongly suggesting instability of these mutant precursor proteins. These results were consistent with instability and/or improper folding of the mutant cathepsin K precursor proteins that were hydrolyzed by pepsin and other proteases.

Incubation of the precursor proteins with pepsin resulted in reduction of the wild-type and Y212C peptides to the expected mass of the mature enzyme (Fig. 2). The amount of Y212C enzyme generated was substantially less than the quantity of initial precursor, suggesting degradation by pepsin and other proteases. For the other mutant precursors, however, incubation with pepsin resulted in loss of the precursor peptide but no emergence of a 30-kDa protein.

As anticipated from the Western analyses, the fluorescent enzyme assays revealed no detectable protease activity of the G79E, G146R, A277E, A277V, or R312G mutant proteins toward the synthetic substrate. In contrast, low levels of activity were detected for Y212C mutant protein compared with the wild-type cathepsin K, which showed activity levels and kinetics comparable to those published previously (11). To further character-

ize the residual activity of the Y212C mutant protein, conditioned medium from a large-scale preparation was concentrated. The percentage of residual cathepsin K activity in the Y212C mutant protein was quantitated using a stoichiometric assay of the enzyme with the irreversible cysteine cathepsin inhibitor E-64, which binds to these proteases in a 1:1 molar ratio. After normalizing the activities of the Y212C and wild-type enzymes to molarity, the Y212C protein had a hydrolysis rate with the dipeptide substrate Z-Leu-Arg-MCA that was 41% of that with the wild-type cathepsin K (k_{cat} $0.64 \pm 0.07/s$ vs. $1.56 \pm 0.53/s$). The Michaelis-Menten constants for the two enzymes were similar (K_m $6.81 \pm 1.66 \mu M$ vs. $6.57 \pm 1.46 \mu M$, respectively), so the catalytic efficiency (k_{cat}/K_m) of the Y212C protein was only 40% of that for the wild-type enzyme (94,000 vs. 237,400 $l/M \cdot s$, respectively).

The pH-activity profiles of the Y212C and wild-type enzymes were comparable with pH optima at 6.1 and similarly broad profiles (Y212C $pK_1 = 4.2$ and $pK_2 = 8.4$; wild-type CTSK $pK_1 = 3.9$ and $pK_2 = 8.5$). Compared with wild-type enzyme, however, the Y212C activity was unstable at pH 5.5 and 6.5, although it was more stable than wild-type activity at the pH 4.5 (Table 1).

The substrate specificity of Y212C was determined using synthetic di- and tripeptide substrates (Table 2). In a pattern very similar to wild-type (11), the order of catalytic efficiency (k_{cat}/K_m) for the Y212C enzyme against four dipeptide substrates was Z-LR-MCA > Z-VR-MCA \approx Z-FR-MCA >> Z-RR-MCA. The Y212C enzyme had minimal catalytic activity toward the tripeptide substrate Z-GPR-MCA, which has been shown to be efficiently hydrolyzed by recombinant rabbit cathepsin K (22).

The ability of the Y212C and wild-type enzymes to degrade type I collagen was determined *in vitro* at pH 5.5 and 28°C. As shown in Fig. 3, there was pronounced collagen degradation by the wild-type enzyme after 12 hours, but only minimal collagenolysis by the Y212C enzyme. In contrast, the Y212C enzyme retained significant proteolytic activity against gelatin, although it did not hydrolyze the substrate as efficiently as did the wild-type activity (Fig. 3).

To ensure that the altered substrate specificity of the Y212C enzyme was not an artifact of abnormal processing of the propeptide, the NH₂-terminal residues of this

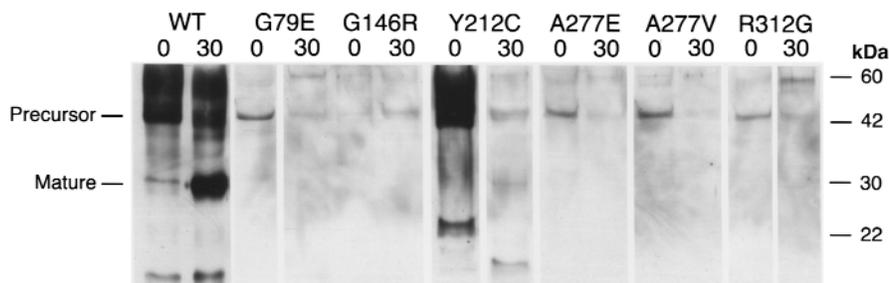


Figure 2
Immunoblot of mutant and wild-type cathepsin K proteins. Aliquots of *Pichia pastoris* culture supernatants, before and after 30 min of pepsin treatment (indicated as 0 and 30, respectively), were electrophoresed in SDS-PAGE and transferred to blots that were hybridized with polyclonal rabbit anti-human cathepsin K antibodies. The wild-type cathepsin K (WT) proenzyme (lane 1) was detected at a higher than expected molecular mass, which resolved after pepsin digestion (lane 2). Molecular mass standards (measured in kDa) are indicated in the right margin. There was no immunologically detectable cathepsin K protein in supernatants from pPIC9K vector-only control (data not shown). The G146R precursor protein, which was less abundant before pepsin treatment in this Western blot, was more abundant at 0 min than at 30 min in all subsequent repetitions

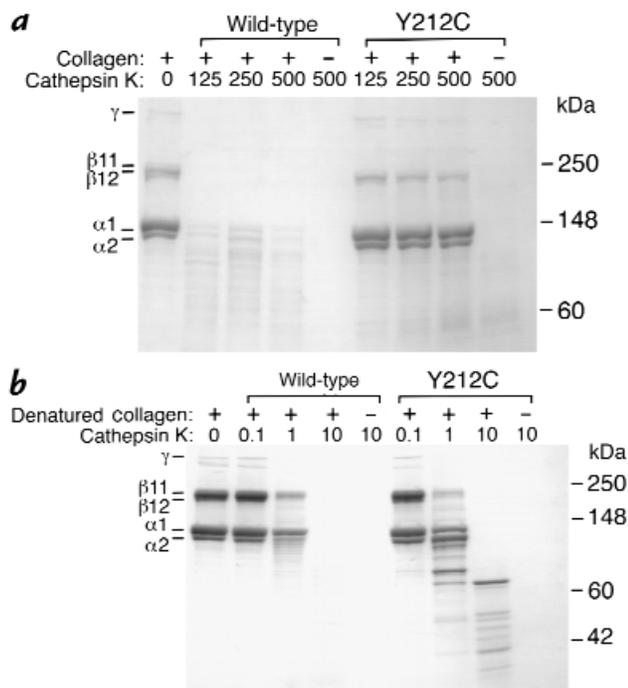


Figure 3
 SDS-PAGE of type I collagen (soluble calf skin collagen) after digestion with recombinant wild-type and Y212C cathepsin K enzymes. (a) Collagenase activity. Digestion of soluble calf skin collagen at 28°C and pH 5.5 for 12 h by wild-type and Y212C cathepsin K enzymes is shown. The presence (+) or absence (-) of collagen and the concentration of the cathepsin K in nM are indicated. (b) Gelatinase activity. Digestion of denatured soluble calf skin collagen by wild-type and Y212C cathepsin K enzymes is shown. The presence (+) or absence (-) of gelatin and the concentration of the cathepsin K in nM are indicated. Molecular mass standards (measured in kDa) are indicated in the right margin, and the collagen proteins are labeled in the left margin.

mutant, as well as the wild-type cathepsin K proteins, were compared. MALDI-TOF mass spectrometry of endopeptidase Lys-C digested mature proteins indicated that both the Y212C and wild-type enzymes were constituted by a mixture of enzymes with W¹¹¹EGRAPD, G¹¹³RAPD, and A¹¹⁵PD NH₂-termini. In addition, the Y212C enzymes contained peaks corresponding to NH₂-termini of Y¹⁰⁷IPEWEGRAPD and I¹⁰⁸PEWEGRAPD, suggesting less-complete maturation of some species. Finally, MALDI-TOF analysis of a tryptic digest of the mutant and wild-type enzymes demonstrated identical traces, except for peaks of 2.664 and 2.653 kDa in the former and latter, respectively, which corresponded to the predicted 24-residue tryptic fragments containing the C²¹² and Y²¹² residues. The ragged NH₂-termini were not detectable after the tryptic digest because trypsin cleaved at R¹¹², producing very short NH₂-terminal fragments that were lost during the elution and mass spectrometry.

Discussion

The seven novel cathepsin K mutations identified in eight unrelated Pycno families further document that defects in the gene cause this skeletal dysplasia and emphasize the degree of molecular genetic heterogeneity underlying this disease. Each point mutation was

confirmed by PCR-based allele-specific analysis of genomic DNA from the affected individuals and heterozygous parents, and none was found among more than 100 chromosomes from unrelated normal individuals. Notably, among the 12 families studied to date (1, 13), all of the disease-causing mutations were missense and nonsense mutations that altered exonic coding sequences. In addition, two of the lesions identified here involved a missense and a nonsense mutation in the pro region of the polypeptide. A single, rare polymorphism was found at nucleotide 936 (G→A change) that did not alter the amino acid sequence.

Among the detected mutations, R241X was found in family 3 from Spain, family 4 from Portugal, a previously described Mexican-American family (1), and an extended Mexican kindred (13). The finding of the R241X mutation in four unrelated families of Iberian ancestry, and its absence in other Pycno families, suggests that this lesion may be relatively common in that geographic region. Support for this concept would require population studies to determine the frequency of the R241X allele in the Iberian Peninsula and in other European and North African populations. If the mutation is of Iberian origin, then this mutation may be relatively ancient, because its transfer to one or both Mexican families probably occurred more than 300 years ago. Alternatively, but less likely, the R241X mutations could have arisen independently on several occasions, because it occurred at a CpG dinucleotide, a known hot spot for mutations (23). Of note, the A277E mutation was detected in two families that share no apparent ethnic background (Indian and Portuguese), although Vasco de Gama first traveled around the western coast of India in 1498 (24). This mutation also occurred at a CpG dinucleotide and could have arisen independently by unrelated mutational events. Future development of intragenic cathepsin K polymorphic markers would permit further analysis of the ancestral background of the R241X and A277E mutations.

The K52X and G79E mutations are the first identified in the 99-residue pro piece region of the cathepsin K prepolypeptide. The K52X nonsense mutation is predicted to terminate polypeptide synthesis in the pro region, thereby eliminating the entire mature enzyme. In contrast, the potential effect of the G79E missense mutation may be subtler. In general, lysosomal cysteine proteases are synthesized as prepolypeptides. The prepeptide is the signal sequence, and the pro piece subserves several functions, including promotion of protein fold-

Table 2

Substrate specificity of the recombinant Y212C mutant enzyme

Substrate	k _{cat} (1/s)	K _m (μM)	k _{cat} /K _m (1/M·s)	Relative activity (%)
Z-L-R-MCA	0.64 ± 0.07	6.81 ± 1.66	94,000	1.0
Z-F-R-MCA	0.32 ± 0.04	12.02 ± 1.72	26,600	0.28
Z-V-R-MCA	0.45 ± 0.07	14.28 ± 2.18	31,500	0.35
Z-R-R-MCA	NA	NA	NA	0
Z-G-P-R-MCA	0.13 ± 0.02	54.61 ± 11.3	2,380	0.025

NA, not available.

ing, protection of the nascent enzyme from neutral pH, trafficking of the polypeptide to the lysosome, and inhibition of protease function until the proenzyme reaches the lysosome, where intramolecular or intermolecular proteolysis will remove the propeptide and activate the enzyme. The cathepsin K pro region contains the only completely conserved potential *N*-glycosylation site (Asn¹⁰³) for lysosomal trafficking via the mannose 6-phosphate receptor pathway. Based on homology with the closely related procathepsin L polypeptide, for which the three-dimensional structure was recently solved (25), a portion of the cathepsin K pro piece sits in the active cleft, so that the enzyme remains nonfunctional until the pro piece is clipped in the lysosome. The G79E missense mutation altered the sixth residue of a seven-amino acid conserved propeptide motif, G-x-N-x-F-x-D (26), that is found among the members of the papain family of cysteine proteases (Fig. 4). This motif does exhibit some variability, and the amino acid residues found in position 6 among papain family members are alanine, serine, glycine, and threonine. Site-directed mutagenesis of residues 1, 3, 5, and 7 in the G-L-N-V-F-A-D motif in papain demonstrated that some substitutions resulted in complete loss of enzymatic function secondary to improper protein folding (note that residues 2, 4, and 6 were not mutagenized; ref. 26).

To model the changes produced by the G79E mutation, the method of conformational memories (27) was used to explore the conformational accessibility of the backbone and side-chain torsional angles in procathepsin K, using peptides consisting of nine amino acid residues with the wild-type or mutated residue occupying the central position; the electrostatic potentials (ESPs) of these constructs were calculated and displayed on the solvent-accessible surface with the program GRASP (28). The conformational search with the nine-residue peptide containing the E⁷⁹ mutation demonstrated a shift in the population of accessible torsion angles of Ψ_{79} compared with that observed in the crystal structure of cathepsin L (25) and readily available to the wild-type cathepsin K nonapeptide. This result implied a decreased flexibility for this conserved motif in the mutant pro piece, because decreased stability might result if Ψ_{79} of the E⁷⁹ mutant were to assume a value near that observed in cathepsin L. Comparison of the ESPs showed that the mutation had far-reaching consequences for the electrostatic properties of this portion of the propeptide. Whereas the negative ESP was confined to a relatively small region in the wild-type sequence near D⁸⁰, it was extended to a surface covering about four adjacent amino acid residues in the mutant. The results also showed the shift of the ESP over the entire nonapeptide to significantly more negative values that could impact the reactivity and recognition properties of this region of the protein. These predicted conformational and electrostatic effects on the pro piece suggest that the G79E mutation might cause improper folding of the cathepsin K proenzyme, resulting in a loss of protease activity, and were consistent with the results of the expression studies with the recombinant G79E mutant. Despite the production of an immunologically detectable proenzyme protein that included a mature

region with the normal cathepsin K primary structure, no mature protein or enzymatic activity was detected. This finding strongly implicated a perturbation in the secondary structure of the precursor protein caused by the G79E change in the propeptide that resulted in complete degradation, rather than activation, by pepsin.

Expression studies with the five missense mutants affecting the mature region of cathepsin K showed that four (G146R, A277V, A277E, and R312G) had no detectable mature protein or enzyme activity. The three-dimensional structure of cathepsin K shows that G¹⁴⁶ lies deep within the active-site cleft, so the substitution of the large, charged Arg residue was expected to have significant adverse effects upon this protein. The A277E and A277V mutations altered another residue in the active site, A²⁷⁷, which is immediately next to the invariably conserved H²⁷⁶ residue that forms the ion pair with the active cysteine (C¹³⁹) to effect protein catalysis. Moreover, the A²⁷⁷ residue is highly conserved among the papain family members, existing as either an alanine or a glycine residue in nearly all instances (29). It was anticipated that the replacement of A²⁷⁷ by a charged residue in the A277E mutation would obliterate the protease activity, but these expression studies revealed that even the A277V change, which appeared to introduce a less drastic alteration, resulted in a precursor protein that was unable to autoactivate and was unstable in the presence of pepsin. Thus, the oblitative effects of A277V suggest that there is little tolerance for larger amino acid side groups at this critical position in the active cleft.

The effects of the fourth missense mutation with no residual activity, R312G, are less obvious. The R³¹² residue does not lie within the active site, but rather resides on the surface. Based on the three-dimensional structure of human procathepsin L (25), R³¹² would not be expected to interact with the pro piece, so the R312G lesion presumably affects the precursor protein structure directly. The R³¹² residue is conserved among approximately two

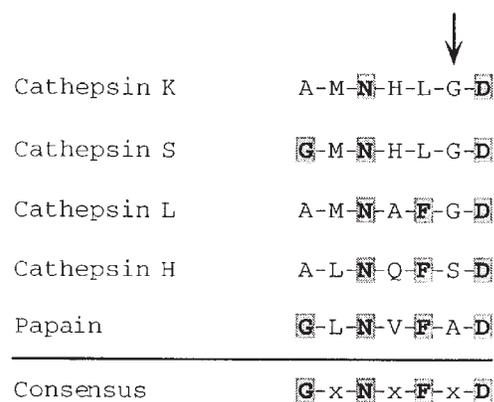


Figure 4

Alignment of conserved motif from the pro region of several members of the papain family of cysteine proteases. The seven-residue motif from four human cysteine cathepsins (GenBank/EMBL accession numbers: cathepsin K, U13665; cathepsin L, X12451; cathepsin S, M90696; cathepsin H, X16832) and papain (M15203) are aligned, and the consensus sequence described by Vernet *et al.* (26) is shown below. The site of the cathepsin K G79E mutation is indicated by the arrow.

thirds of the papain family members, never existing as a glycine residue (7), and the R312G mutation eliminates a positive charge, all consistent with the conclusion that this mutation significantly disrupts the surface structure of the cathepsin K proenzyme in this region.

The only mutant cathepsin K with residual protease activity was Y212C. Biochemical characterization performed revealed that the Y212C mutant had only modestly diminished enzymatic activity with dipeptide substrate specificity and pH profiles that were quite similar to those of the recombinant wild-type cathepsin K. The Y212C activity, however, was unstable at pH conditions around its pH optimum. Most critically, the mutant enzyme retained almost no collagenolytic activity, despite retaining good gelatinolytic activity, and showed significantly reduced activity toward Z-GPR-MCA, a tripeptide found previously to be an excellent and relatively specific synthetic substrate for cathepsin K (22). Mass spectrometric analysis demonstrated that the activation of the Y212 propeptide resulted in enzymes with NH₂-termini that were identical to those of the wild-type cathepsin K, although additional species with longer NH₂-terminal extensions were also noted. The presence of cathepsin K enzymes with variable NH₂-termini was also documented in a prior study of the activation of cathepsin K (30), as well as by similar work with cathepsins B and S (31, 32). Inspection of the three-dimensional structure of the mature enzyme (33) revealed that the NH₂-terminus is relatively far from the active site, residing on the opposite side of the protein, so that the extension of the NH₂-terminus by a small number of residues from the predicted cleavage site at A¹¹⁵ would not be expected to interfere with enzymatic function or substrate specificity. Because the Y212C enzymes are still capable of degrading denatured proteins as well as the synthetic dipeptides favored by wild-type cathepsin K, it suggests that the structural deformation caused by the Y212C mutation prevents engagement of the type I collagen substrate in the active cleft. It remains to be determined whether, like matrix metalloproteinases, cathepsin K will have a second, noncatalytic domain that is also required for collagen binding (34).

The Y²¹² residue does not lie in the active site, however, but rather on the surface of the mature protein. This residue is completely conserved among the six known human cysteine cathepsins, except for cathepsin H, in which it exists as a phenylalanine residue (7), and is never a cysteine residue among any papain family member (29). Because the Y212C mutation introduces a thiol group that might form a novel disulfide bridge or disrupt existing ones, it suggests that a structural deformation on the surface perturbs the conformation of the active site, preventing proper engagement of type I collagen. Future three-dimensional structural studies with Y212C protein may provide more precise information about the nature of these conformational changes.

In summary, seven new mutations in the cathepsin K gene have been identified, providing further evidence that the deficient activity of this enzyme causes Pycno and demonstrating the molecular genetic heterogeneity underlying this skeletal dysplasia. Among these lesions, two were in the region encoding the propeptide that is

presumably responsible for protein folding, lysosomal trafficking, and inactivation of the enzyme until it is cleaved in the lysosome. The lack of activity by the pro region missense mutant G79E suggests that the previously identified seven-residue pro region motif that it alters is critical for protein folding. Several missense and nonsense mutations in the mature polypeptide were detected, R241X being frequent in the Hispanic patients studied. Expression of five mature enzyme missense mutations revealed that four mutant enzymes had no detectable activity and that one, Y212C, had good protease activity and preserved dipeptide substrate specificity but retained minimal activity toward type I collagen and a synthetic tripeptide. This structure/function information suggests that the cathepsin K active site contains a critical collagen-binding domain.

Acknowledgments

This work was supported in part by National Institutes of Health research grants (1 R29 AR-44231 to B.D. Gelb, 5 R37 DK-34045 to R.J. Desnick, 5 PO1 HD-22657 to D.L. Rimoin, and DA00060-18 to H. Weinstein), a grant from the National Center for Research Resources for the Mount Sinai General Clinical Research Center (5 M01 RR00071), a grant for the Mount Sinai Child Health Research Center from the National Institutes of Health (5 P30 HD-28822), and a Basic Research Award from the March of Dimes Birth Defects Foundation (to B.D. Gelb).

1. Gelb, B.D., Shi, G.-P., Chapman, H.A., and Desnick, R.J. 1996. Pycnodysostosis, a lysosomal disease due to cathepsin K deficiency. *Science*. **273**:1137-1139.
2. Maroteaux, P., and Lamy, M. 1962. La pycnodysostose. *Presse Med*. **70**:999-1002.
3. Andrén, L., Dymling, J.F., Hogeman, K.E., and Wendeberg, B. 1962. Osteopetrosis acro-osteolytica: a syndrome of osteopetrosis, acro-osteolysis and open sutures of the skull. *Acta. Chir. Scand*. **124**:496-507.
4. Tezuka, K., et al. 1994. Molecular cloning of a possible cysteine proteinase predominantly expressed in osteoclasts. *J. Biol. Chem*. **269**:1106-1109.
5. Inaoka, T., et al. 1995. Molecular cloning of human cDNA for cathepsin K: novel cysteine proteinase predominantly expressed in bone. *Biochem. Biophys. Res. Commun*. **206**:89-96.
6. Shi, G.-P., et al. 1995. Molecular cloning of human cathepsin O, a novel endoproteinase and homologue of rabbit OC2. *FEBS Lett*. **357**:129-134.
7. Brömme, D., and Okamoto, K. 1995. Human cathepsin O2, a novel cysteine protease highly expressed in osteoclastomas and ovary: molecular cloning, sequencing and tissue distribution. *Biol. Chem. Hoppe Seyler*. **376**:379-384.
8. Li, Y.-P., et al. 1995. Cloning and complete coding sequence of a novel human cathepsin expressed in giant cells of osteoclastomas. *J. Bone Miner. Res*. **10**:1197-1202.
9. Everts, V., Aronson, D.C., and Beertsen, W. 1985. Phagocytosis of bone collagen by osteoclasts in two cases of pycnodysostosis. *Calcif. Tissue Int*. **37**:25-31.
10. Gelb, B.D., et al. 1996. Cathepsin K: isolation and characterization of the murine cDNA and genomic sequence, the homolog of the human pycnodysostosis gene. *Biochem. Mol. Med*. **59**:200-206.
11. Brömme, D., Okamoto, K., Wang, B.B., and Biroc, S. 1996. Human cathepsin O2, a matrix protein-degrading cysteine protease expressed in osteoclasts. *J. Biol. Chem*. **271**:2126-2132.
12. Bossard, M.J., et al. 1996. Proteolytic activity of human osteoclast cathepsin K. *J. Biol. Chem*. **271**:12517-12524.
13. Johnson, M.R., et al. 1996. A nonsense mutation in the cathepsin K gene observed in a family with pycnodysostosis. *Genome Res*. **6**:1050-1055.
14. Gelb, B.D., et al. 1997. Structure and chromosomal assignment of the human cathepsin K gene. *Genomics*. **41**:258-262.
15. Haliassos, A., et al. 1989. Modification of enzymatically amplified DNA for the detection of point mutations. *Nucleic Acids Res*. **17**:3606.
16. Newton, C.R., et al. 1989. Analysis of any point mutation in DNA. The amplification refractory mutation system (ARMS). *Nucleic Acids Res*. **17**:2503-2516.
17. Qin, J., et al. 1997. A strategy for rapid, high-confidence protein identification. *Anal. Chem*. **69**:3995-4001.
18. Ogryzko, V.V., et al. 1998. Histone-like TAFs within the PCAF histone

- acetylase complex. *Cell*. **94**:35–44.
19. Fenyó, D., Zhang, W., Chait, B.T., and Beavis, R.C. 1996. Internet-based analytical chemistry resource: a model project. *Anal. Chem.* **68**:721A–726A.
 20. Chait, B.T., and Kent, S.B. 1992. Weighing naked proteins: practical, high-accuracy mass measurement of peptides and proteins. *Science*. **257**:1885–1894.
 21. Linnevers, C.J., et al. 1997. Expression of human cathepsin K in *Pichia pastoris* and preliminary crystallographic studies of an inhibitor complex. *Protein Sci.* **6**:919–921.
 22. Aibe, K., et al. 1996. Substrate specificity of recombinant osteoclast-specific cathepsin K from rabbits. *Biol. Pharm. Bull.* **19**:1026–1031.
 23. Cooper, D.N., and Krawczak, M. 1990. The mutational spectrum of single base-pair substitutions causing human genetic disease: patterns and predictions. *Hum. Genet.* **85**:55–74.
 24. Ryan, A.N. 1992. Vasco da Gama. In *Encyclopedia Americana*. Grolier Inc. Danbury, CT. 259–260.
 25. Colombe, R., et al. 1996. Structure of human procathepsin L reveals the molecular basis of inhibition by the prosegment. *EMBO J.* **15**:5492–5503.
 26. Vernet, T., et al. 1995. Processing of the papain precursor. *J. Biol. Chem.* **270**:10838–10846.
 27. Guarnieri, F., and Weinstein, H. 1996. Conformational memories and the exploration of biologically relevant peptide conformations: an illustration for the gonadotropin-releasing hormones. *J. Am. Chem. Soc.* **118**:5580–5589.
 28. Nicholls, A., Sharp, K.A., and Honig, B. 1991. Protein folding and association: insights from the interfacial and thermodynamic properties of hydrocarbons. *Proteins*. **11**:281–296.
 29. Berti, P.J., and Storer, A.C. 1995. Alignment/phylogeny of the papain superfamily of cysteine proteases. *J. Mol. Biol.* **246**:273–283.
 30. McQueney, M.S., et al. 1997. Autocatalytic activation of human cathepsin K. *J. Biol. Chem.* **272**:13955–13960.
 31. Rowan, A.D., Mason, P., Mach, L., and Mort, J.S. 1992. Rat procathepsin B. Proteolytic processing to the mature form *in vitro*. *J. Biol. Chem.* **267**:15993–15999.
 32. Bromme, D., et al. 1993. Functional expression of human cathepsin S in *Saccharomyces cerevisiae*. Purification and characterization of the recombinant enzyme. *J. Biol. Chem.* **268**:4832–4838.
 33. McGrath, M.E., Klaus, J.L., Barnes, M.G., and Bromme, D. 1997. Crystal structure of human cathepsin K complexed with a potent inhibitor. *Nat. Struct. Biol.* **4**:105–109.
 34. Shingleton, W.D., Hodges, D.J., Brick, P., Cawston, T.E. 1996. Collagenase: a key enzyme in collagen turnover. *Biochem. Cell Biol.* **74**:759–775.