Articles

The Crystal Structure of Human Procathepsin K‡

Judith M. LaLonde,§ Baoguang Zhao,§ Cheryl A. Janson,|| Karla J. D'Alessio,|| Michael S. McQueney,|| Michael J. Orsini,⊥ Christine M. Debouck,# and Ward W. Smith*,§

Departments of Structural Biology, Protein Biochemistry, and Molecular Genetics, SmithKline Beecham Pharmaceuticals, 709 Swedeland Road, King of Prussia, Pennsylvania 19406, USA and Kimmel Cancer Center, Thomas Jefferson University, Philadelphia, Pennsylvania 19107

Received September 16, 1998; Revised Manuscript Received November 2, 1998

ABSTRACT: Cathepsin K is a cysteine protease present in human osteoclasts that plays an important role in bone resorption. Cathepsin K is synthesized as an inactive proenzyme and activated under conditions of low pH. Autoproteolytic processing of the N-terminal 99 amino acid propeptide produces the active, mature form of cathepsin K. It is presumed that the activation of procathepsin K *in vivo* occurs in the bone resorption pit, which has a low-pH environment. We have determined the structure of human procathepsin K at 2.8 Å resolution. The structure of the mature enzyme domain within procathepsin K is virtually identical to that of mature cathepsin K. The fold of the propeptide of procathepsin K is similar to that observed in procathepsins B and L despite differences in length and sequence. A portion of the propeptide occupies the active site cleft of cathepsin K. Hydrophobic interactions, salt bridges, and hydrogen-bonding interactions are observed in the structure of the propeptide and between the propeptide and the mature enzyme of procathepsin K. These interactions suggest an explanation for the stability of the proenzyme. The structure of procathepsin K contributes to an understanding of the molecular basis of inhibition by the propeptide portion of the molecule and activation of this important member of the cysteine protease family.

Cathepsin K is a recently discovered proteinase belonging to the papain family of cysteine proteinases. It has been implicated in the resorption of the bone matrix (*1*−*6*). Bone remodeling is a dynamic process that involves bone resorption and formation. The resorption phase of this process is carried out by osteoclast cells, which adhere to the surface of bone, leading to the creation of an extracellular compartment, the resorption pit, which is maintained at low pH and into which the osteoclast secretes proteolytic enzymes. In this acidic environment, the mineral region of the underlying bone is removed and the protein matrix is exposed to degradation by proteolytic enzymes. The second phase of the remodeling involves the recruitment of osteoblast cells to the appropriate sites, where the rebuilding or layering of a new protein matrix becomes mineralized.

Like most of the proteinases, cathepsin K is synthesized and secreted from the cell as an inactive proenzyme that is converted to its mature active form by proteolytic cleavage of a 99 amino acid propeptide from the amino-terminus. The processing of procathepsin K to mature cathepsin K *in vitro* is autocatalytic at pH 4.0 (*7*). The bone resorption pit environment is one of low pH (*8*). This is consistent with the observation that cathepsin K plays an important role in bone remodeling and has led to the suggestion that activation of procathepsin K *in vivo* may occur in the resorption pit.

An understanding of the molecular basis of inhibition and activation of cysteine proteases is enhanced by knowledge of the three-dimensional structure of their proenzymes. Recently the structures of procathepsin B (pCatB)[1] and procathepsin L (pCatL) have been determined (*9*−*11*). Sequence comparison shows that mature cathepsin K shares 76% sequence similarity with cathepsin L and 58% with cathepsin B. There is no significant sequence homology between the propeptide portions of the proenzymes (*12*).

We have reported the crystal structures of mature human cathepsin K−inhibitor complexes (*13*, *14*). These structures are being used to design novel, potent inhibitors of human osteoclast cathepsin K for use in the treatment of diseases of bone remodeling. Here we report the crystal structure of human procathepsin K at 2.8 Å. Examination of the structure

‡ The coordinates have been deposited in the Brookhaven Protein Databank, Accession Number 1BY8.

* Correspondence should be addressed to this author. Tel: (610) 270-7017. Fax: (610) 270-4091. E-mail: ward_w_smith@sbphrd.com.

§ Department of Structural Biology, SmithKline Beecham Pharmaceuticals.

|| Department of Protein Biochemistry, SmithKline Beecham Pharmaceuticals.

⊥ Kimmel Cancer Center, Thomas Jefferson University.

Department of Molecular Genetics, SmithKline Beecham Pharmaceuticals.

[1] Abbreviations: pCatK, procathepsin K; mCatK, mature cathepsin K.; pCatL, procathepsin L; pCatB, procathepsin B; PBL, propeptide-binding loop; PEG, poly(ethylene glycol); Cbz, carbobenzyloxy.

10.1021/bi9822271 CCC: $18.00 © 1999 American Chemical Society
Published on Web 12/23/1998

Table 1: Data and Refinement Statistics

space group	$C222_1$
cell	56.78, 156.55, 96.26, 90. 90. 90
resolution limit	2.64 Å
no. of observations	54 032
no. of reflections	12 036
$R_{sym}(I)$	6.9%
% complete shell 2.85−2.80 (Å)	95%
% complete shell 2.74−2.64 (Å)	60%
I/sigma *I* for shell 2.85−2.80 (Å)	2.01
refinement	
resolution	10−2.6 Å
no. of reflections F>2 sigma (F)	8659
no. of atoms	2447
average *B*-factor (Å^2)	32
R-factor/*R*-free	0.214/0.344
RMS deviations	
bonds (Å)	0.006
angles (deg)	1.010
dihedrals (deg)	28.104
impropers (deg)	0.688

of procathepsin K and a comparison with procathepsin L (*11*) provide an understanding of components important in inhibition and autoactivation of cathepsin K.

MATERIALS AND METHODS

Crystallization and X-ray Diffraction Measurements. Procathepsin K was expressed in *Escherichia coli* and refolded and purified from inclusion bodies as described elsewhere (K. J. D'Alessio et al., to be published). The purified protein was concentrated in 25 mM Na_2HPO_4, 1 M NaCl, pH 7.0, using a Centricon-10 (Amicon) to 20 mg/mL then diluted 1:1 with 25 mM Na_2HPO_4 at pH 7.6 to yield a 10 mg/mL solution containing 0.5 M NaCl. The hanging drop vapor diffusion method was used for crystallization trials. Several conditions to promote crystallization were explored using a number of precipitants. Systematic variation of the concentrations of PEG 400, PEG 6000, and PEG 8000 was the most promising and led to the best crystallization conditions. Crystals of procathepsin K were obtained from a solution of 8−10% PEG6000, 0.1 M Sodium citrate at pH 5.0. Crystals of procathepsin K grew to a size of approximately 0.3 mm^3 in 10 days at 4 °C. X-ray diffraction data were measured at room temperature from a single crystal using a Siemens two-dimensional position-sensitive detector (Table 1). Crystals of procathepsin K are orthorhombic, space group $C222_1$, with cell constants of $a = 56.78$ Å, $b = 156.55$ Å, and $c = 96.26$ Å. The crystals contain one molecule in the asymmetric unit and contain 60% solvent.

Structure Determination. The structure of procathepsin K was solved in a manner similar to that employed for human procathepsin L (*11*) and rat and human procathepsin B (*9*, *10*). This consisted of a combination of molecular replacement to locate the position of the mature portion of the proenzyme and density modification to image the propeptide residues. The coordinates for 215 residues from the structure of human cathepsin K (*13*) were used as a starting model in molecular replacement. By using the program X-plor (*15*), we calculated the cross-rotation function using data from 10 to 3.3 Å with a radius of integration of 22 Å and a P1 cell for the search model with dimensions of $a = 88.0$ Å $b = 74.0$ Å, and $c = 68.0$ Å. The Patterson search was carried out with pseudo-orthogonal Euler angles $\theta 1$ from 0° to 180°, $\theta 2$ from 0° to 90°, and $\theta 3$ from 0° to 720° in space group P1. The highest peak was 15.7 sigma. The translation search was computed using data from 10 to 3.5 Å and gave a peak at 26 sigma. The resulting model yielded an *R*-factor of 0.43. The *R*-factor for the model dropped to 0.40 after rigid body refinement from 10 to 3.5 Å and to 0.31 after 200 cycles of positional refinement.

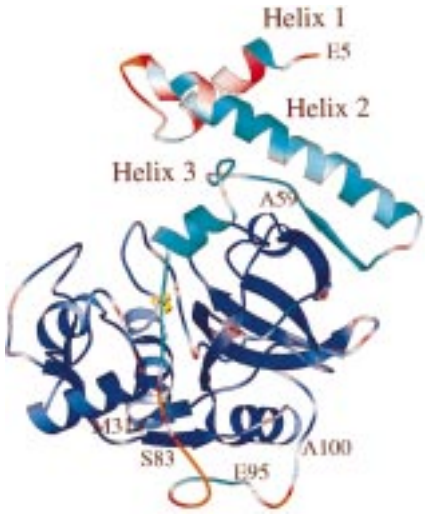

FIGURE 1: Ribbon diagram depicting the three-dimensional structure of procathepsin K. The color of the structure varies according to B-factor with the lowest B-factors in the mature region of the structure drawn in royal blue and the lowest B-factors in the propeptide drawn in cyan. The color changes from royal blue or cyan to red as the B-factors increase to 55. Regions with B-factors over 55 Å^2 are colored orange. The side-chain atoms of the catalytic cysteine (Cys124) are drawn in yellow at the center of the active site. Cleavage sites occur at E4 (not shown), A59, S83, and E95. The sequence of the mature structure begins at A100. Residues 1−4 are disordered in the structure and are not shown. Figure drawn with BobScript (*26*).

Inspection of the electron density map after positional refinement clearly showed the helical regions of the propeptide, which were added to the model. The electron density maps were improved using solvent flattening, histogram matching, and skeletonization as implemented in the program DM from the CCP4 suite (*16*). The modified electron density map was used to build the remaining portion of the propeptide segment of procathepsin K. Conventional positional refinement was used to refine the structure during model building. The structure was refined using torsion angle dynamics as implemented in X-plor version 3.8 (*17*, *18*) with the results shown in Table 1. The data are 95% complete in the shell from 2.85 to 2.8 Å, but all data to 2.6 Å were used in the refinement (Table 1). Electron density maps were calculated using Sigma-A weighting and cross-validation (*19*, *20*). Finally, one cycle of unrestrained group B-factor refinement (one B-factor for the main-chain atoms and one B-factor for the side-chain atoms of each residue) was performed as implemented in X-plor (*15*). Residues 1−4 from the N-terminus of the propeptide are disordered and not observed in our structure. The model shows only one residue, Val 26, in a disallowed region of the Ramachandran plot. The final model of procathepsin K consists of residues 5−314 with an R-factor of 0.214 and an *R*-free of 0.344.

The C-terminal portion of the propeptide (83−99) has high B-factors (Figure 1). The average B-factors are 51 and 58 Å^2 for main-chain and side-chain atoms, respectively. The

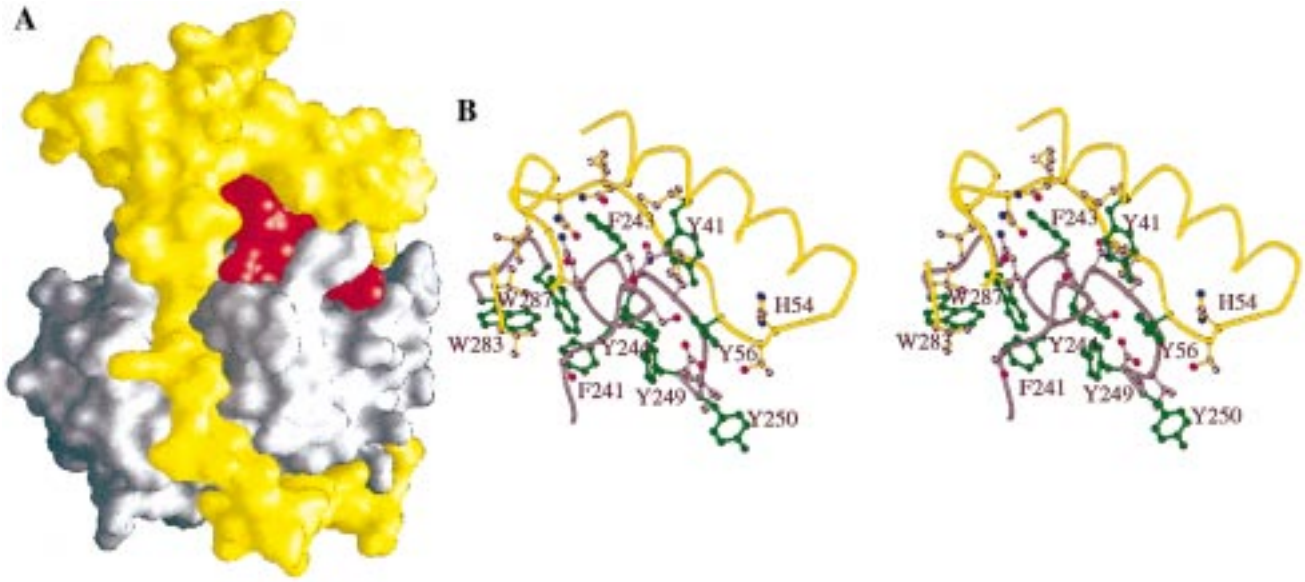

FIGURE 2: (A) Molecular surface formed by the propeptide (yellow) and mCatK (gray and red) in pCatK. The propeptide-binding loop (PBL), residues 236−251, is colored red. Part A generated by the program GRASP (*21*). (B) Stereoview of hydrophobic interactions and hydrogen bonding (defined as having a distance less than 3.5 Å between donor and acceptor) between the propeptide and mature protein. Aromatic residues are drawn in green. Residues N38, I42, T55, L58, A59, M60, M61, L63, G64, T67, V71, V72, S237, S239, S240, Q242, V248, and D251 are shown by atom color: carbon, gray; nitrogen, blue; and oxygen, red. Figure drawn with BobScript (*26*).

backbone atoms of residues 84−89 have B-factors greater than 50 Å^2. The side-chain atoms of residues 82−86, 89, 92, 95 to 96, and 98 to 99 have B-factors above 50 Å^2. These residues must be considered disordered. We interpret these high B-factors in terms of thermal disorder rather than an incorrect model because the model fits the density of the original unbiased density modified map (data not shown).

RESULTS

Overall Structure. The propeptide consists of residues 1−99; however residues 1−4 are disordered in our structure. The mature part of cathepsin K is referred to as mCatK and consists of residues 100−314 (Figure 1). The structure of the mCatK domain in procathepsin K (pCatK) is virtually identical to that of the mCatK structure alone with an rms difference for 215 CA atoms of 0.47 Å. The propeptide forms a globular domain comprising three α-helices and one β-strand located above the β-sheet domain of mCatK as shown in Figure 1. The first and second helices (residues 7−18 and residues 24−51, respectively) cross each other at an approximate 90° angle. A turn follows the second helix and changes the direction of the polypeptide. This positions the short β-strand (residues 55−58) adjacent to one β-strand in mCatK (Figure 1). These two β-strands hydrogen bond to form a two-stranded β-sheet. Helix 3 (residues 68−74) sits adjacent to the substrate-binding cleft (Figure 1). An extended loop stretches from the globular domain through the active site (residues 74−81). Two turns occur in the C-terminal segment of the propeptide prior to connecting to the amino-terminus of mCatK (Figure 1). The total surface area formed by the interface between the propeptide and mCatK is 3986 Å^2 as calculated with the program GRASP (*21*).

The propeptide can be divided into three segments based on conformation and interaction with mCatK: the globular domain, residues 5 to 73, the active cleft-binding segment, residues 74 to 81 and the C-terminal segment, residues 82 to 99. The overall B-factor for the structure is 32 Å^2. The average B-factor for the propeptide is 40 Å^2, higher than that of mCatK, 28 Å^2. The larger B-factors in the propeptide are mostly in the C-terminal segment in which the average B-factor is 55 Å^2 (Figure 1). This is in contrast with the average B-factors for the globular domain and the active site regions, 36 and 31 Å^2, respectively. The C-terminal segment appears to be a highly flexible region of the propeptide and the implications of this flexibility in understanding the mechanism of activation are discussed below.

Interactions between the Propeptide and mCatK: Globular Domain. The propeptide globular domain is anchored to the mature protein by hydrophobic interactions and hydrogen bonding. The hydrophobic faces of helices 2 and 3, and the β-strand pack against the apolar surface of mCatK (Figure 2A). Residues 236 to 251 form the apolar surface of mCatK. This stretch of polypeptide contains a single turn of α-helix, residues 239 to 243, and is referred to as the propeptide-binding loop (PBL) (*10*, *11*) (Figure 2B). The PBL is straddled by the propeptide, including residues 55 to 58 of the β-strand, which allows the β-strand to hydrogen bond with a β-strand from mCatK (residues 244 to 249) forming an antiparallel β-sheet. Aside from the β-sheet hydrogen bonding between the PBL and the propeptide, other hydrogen bonds are formed that anchor the propeptide to the PBL. These are main-chain to side-chain hydrogen bonds (defined as having a distance less than 3.5 Å between donor and acceptor) between Gln242 and Asn61, Gln242 and Ser68; Val248 and Thr55; Tyr250 and Tyr56; Ser68 and Gln242; Leu63 and Gln242; a main-chain hydrogen bond occurs between Phe243 and Ala59.

The hydrophobic surface is formed principally by Phe243 (Figure 2B). Phe243 forms key interactions between the PBL and the propeptide by forming van der Waals contacts with Asn38, Ile42, Leu58, Asn61, Leu63 and Gly64. A second important hydrophobic contact region exists between Tyr56 on the propeptide and Tyr244, Gly247, Val248, Tyr249,

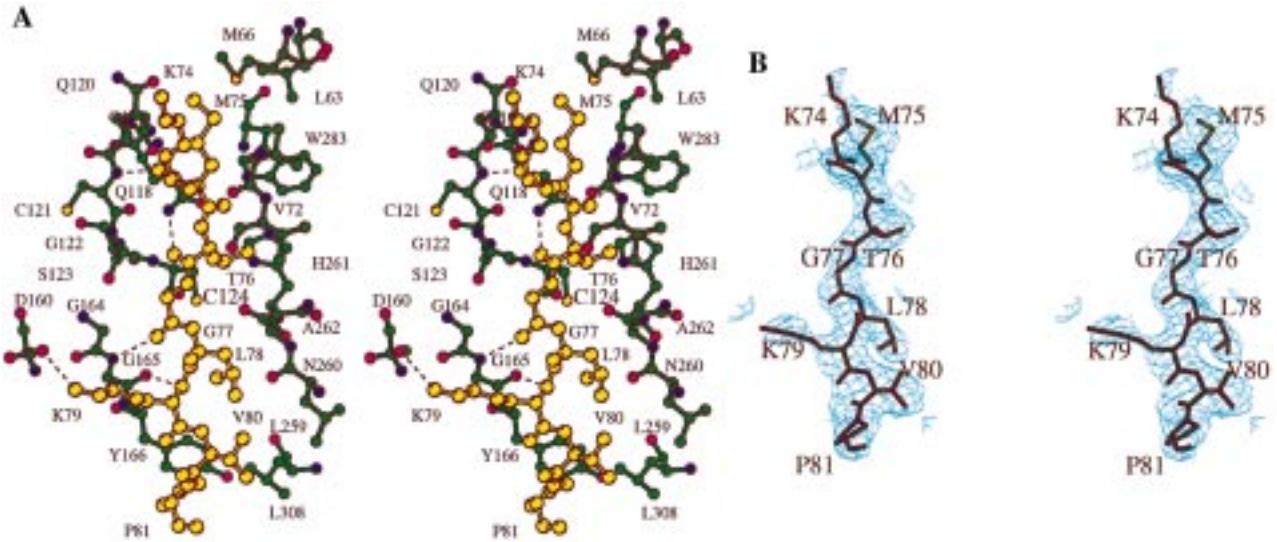

FIGURE 3: (A) Interactions of the propeptide of procathepsin K in the active site. Atoms of the propeptide lying in the active site are drawn in yellow, and interacting atoms from pCatK are in green, red, and blue for carbon, oxygen, and nitrogen atoms, respectively. (B) Electron density map of the propeptide of procathepsin K in the active site. The map was calculated using $(2F_o - F_c)$ coefficients and is contoured at 2 sigma. Figures drawn with BobScript (*26*).

Tyr250 and Asp251 on the PBL. Aromatic stacking occurs between Tyr56 and Tyr244, Tyr249 and Tyr250. This aromatic network extends into the mCatK to include Phe241, Phe243, Tyr244, Tyr249, Tyr250, Trp283, and Trp287 (Figure 2B).

Interactions between the Propeptide and mCatK: Active Cleft-Binding Segment. Residues 74−81 of the propeptide lie in the active site (Figure 3A). The electron density map of the active site is shown in Figure 3B. The orientation of the propeptide in the cleft is opposite to that postulated for a natural substrate (*22*). Helix 3 positions the propeptide to enter the active site at the S′-subsites and continues through the S-subsites. Numerous electrostatic and hydrophobic contacts are formed with residues in the subsites (Figure 3A). The S′-subsites (located above the catalytic Cys124 in Figure 3A) are occupied by Thr76 (S1′) and Met75 (S2′). The S-subsites (located below Cys124 in Figure 3A) are occupied by Gly77 (S1), Leu78 (S2), and Lys79 (S3). Lys74 sits at the end of helix 3, near the S′-subsites of mCatK, forming a main-chain hydrogen bond with Cys121. Although not positioned in an S′-subsite pocket, the amino group from Lys74 forms a hydrogen bond with the amido side chain of Gln120. Met75 sits in the S2′-subsite, with its side chain positioned in the hydrophobic region created by the side chains of Leu63 and Met66 from the propeptide and Trp283 from mCatK. Met75 forms a hydrogen bond from its main-chain carbonyl to the amide group of Gln118 and to the indole nitrogen of Trp 283. The S1′-subsite is occupied by Thr76 where it forms hydrogen bonds from its side-chain hydroxyl group to the main-chain carbonyl oxygen of Val72 of the propeptide. A main-chain hydrogen bond also occurs between the carbonyl oxygen of Thr76 and the amide nitrogen of Gln118. The carbonyl oxygen of Thr76 is also 3.4 Å away from the catalytic sulfur of Cys124. Gly77 lies in the S1-subsite, the center of the active site. Its main chain forms van der Waals contacts with the catalytic thiol of Cys124. Gly77 also forms main-chain hydrogen bonds between its nitrogen and carbonyl oxygen with the carbonyl oxygen of Lys79 and the nitrogen from Gly165, respectively. A residue with a larger side chain would not fit in this position. In the S2-subsite, the main-chain nitrogen of Leu78 is also within van der Waals radius of the catalytic thiol of Cys124 (3.1 Å). Leucine 78 forms hydrophobic contacts with Leu259, His261, and Asn260, which form the P2 specificity pocket. In the S3-subsite, Lys79 forms main-chain hydrogen bonds between its nitrogen and the carbonyl oxygen atoms of Gly77 and Gly165. Another hydrogen bond is formed between Lys79 carbonyl oxygen and the nitrogen of Pro81. The side-chain amino group of Lys79 faces away from the catalytic site and forms a salt bridge with Asp160. Val80 forms a hydrophobic contact with Leu78 and Leu259. Pro81 forms a hydrophobic contact with Leu308, but more importantly, it forms extensive contacts with Tyr166 by stacking against its aromatic ring.

Interactions between the Propeptide and mCatK: C-Terminal Segment of the Propeptide. The C-terminal portion of the propeptide (83−99) has high B-factors (Figure 1). Cygler et al. (*10*) and Coulombe et al. (11) each make similar observations of high B-factors in the C-terminal domain of the propeptide in the structures of rat procathepsin B and human procathepsin L. Because the C-terminal segment of the propeptide is disordered, specific interactions with mCatK cannot be inferred. However, the C-terminal segment of the propeptide contains two of the four cleavage sites, Ser83 and Glu95 (see below). McQueney et al. (*7*) observed that these two cleavage sites occur one or two residues from a proline, Pro81 or Pro94. Cathepsin K may prefer sites that have structural features unique to proline residues (*7*). This would be consistent with a preference by cathepsin K for the proline-rich substrate, type I collagen (*23*). Indeed the two proline residues, Pro81 and Pro94, have lower B-factors than the other residues in the C-terminal portion of the propeptide.

Interactions between the Propeptide and mCatK: Electrostatic Interactions. One salt bridge occurs between the propeptide and mCatK, Lys79 to Asp160. It is located in the active site and was described above. However, there are five well-defined salt bridges within the propeptide globular domain (salt bridges defined here as having an ionic interaction less than 3.5 Å): Asp8 to Arg32, Glu36 to Arg32,

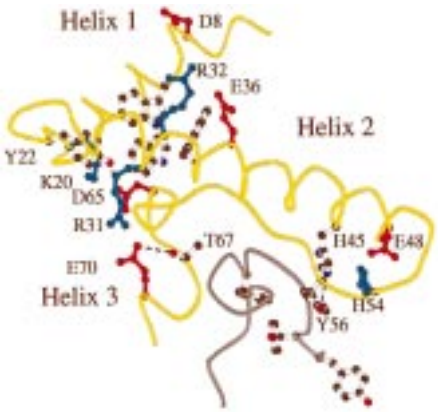

FIGURE 4: Electrostatic interactions at a distance of 3.5 Å or less in the propeptide. The backbones of propeptide and mCatK are shown in yellow and gray, respectively. Shown are acidic residues colored in red, basic residues in blue, and some of the hydrophobic residues (colored by atom) that form the electrostatic and hydrophobic network in the globular domain of the propeptide. Figure drawn with BobScript (*26*).

Glu48 to His54, Asp65 to Lys20, and Glu70 to Arg31 (Figure 4). Additional hydrogen bonds occur between side chains of Asp65 and Tyr22; Thr67 and Glu70; and His45 and Tyr56. Networks of electrostatic and hydrophobic interactions connect these residues. From the salt bridge Glu48 to His54, His54 forms a hydrophobic stacking interaction with His45, and His45 hydrogen bonds to Tyr56, which in turn stacks with Tyr244, Tyr249, and Tyr250 in mCatK. A second network occurs from the salt bridge Asp65 to Lys20, to the hydrogen bond Asp65 and Tyr22, and to the aromatic stacking interactions of Tyr22 with Trp11, Trp11 with Trp35, and Trp35 with Trp14. We postulate that disruption of the salt bridges would expose the hydrophobic core of the propeptide. The packing of the hydrophobic face of the propeptide against mCatK would then be disturbed, allowing cleavage of the propeptide to occur more readily.

Cleavage Sites in pCatK. Activation of the proenzyme *in vitro* is autocatalytic at pH 4 and is believed to be a bimolecular process in which one procathepsin K molecule is cleaved by another activated molecule (*7*). Cleavage sites during autoprocessing occur at Glu4 (the beginning of the N-terminal domain), Ala59 (the β-sheet region), Ser83, and Glu95 (the end of the C-terminal segment of the propeptide) (*7*). N-terminal sequence analysis indicates that the mCatK is a mixture of mature enzymes with the N-termini of Gly98, Arg99, and Ala100 (*7*). These forms of mCatK are presumably produced by nonspecific proteolysis after cleavage at Glu95. The residue Glu4 is not observed in our structure. The residues Ser83 and Glu95 are part of the disordered C-terminal segment of the propeptide. This region is solvent accessible and appears to protrude to one side of the proenzyme. The observed disorder may reflect the inherent flexibility in the polypeptide that facilitates cleavage.

Another cleavage site is observed at Ala59 in the mutant form (Cys124Ser) of pCatK (*7*). Peptide fragments resulting from cleavage at Ala59 were observed only after treatment of the proenzyme containing a Cys124 to Ser mutation with mature cathepsin K (*7*). This mutant underwent limited proteolysis when 1% wild-type mature cathepsin K was added, but did not process completely to the mature form at pH 4.0. Proteolysis stops because molar equivalents of active mature cathepsin K are not generated from the mutant pCatK, which lacks the ability to autoprocess. We believe that cleavage occurs at Ala59 in wild-type pCatK, but the peptide fragments resulting from cleavage at this site may degrade too rapidly to be observed (M. S. McQueney, unpublished results). Ala59 lies in the pCatK sequence Leu58-Ala59-Met60-Asn61-His62-Leu63-Gly64, which corresponds to the Gly-X-Asn-X-Phe-X-Asp consensus sequence for the initial cleavage in the autoactivation of papain (*24*). In mCatK, leucine is the preferred substrate residue in the S2-subsite. Cleavage at Ala59 places Leu58 in the S2 specificity pocket. Furthermore, docking of this propeptide fragment (58−64) in the active site shows that the hydrophobic side chain of Met60 may occupy either S1′ or S2′ much as the side chain of Met75 binds in the S2′-subsite of our structure. Therefore, the propeptide (74−81) sitting in the active site could block cleavage by mimicking one of the actual cleavage sites of the propeptide (Ala59 in the sequence Leu58-Ala59-Met60-Asn61-His62-Leu63-Gly64). The cleavage site Ala59 is located at the end of the short β-strand that occurs from residues 55−58. These residues form a β-sheet with residues 244−249 in mCatK. Ala59 forms a main-chain hydrogen bond between its nitrogen and the carbonyl oxygen of Phe243. The scissile bond between Ala59 and Met60, however, does not form main-chain hydrogen bonds to mCatK. Examination of a molecular surface of the interface between the propeptide and mCatK reveals that residues Ala59 and Met60 are partially buried (not shown). We believe that this hydrophobic interface would have to be disrupted for cleavage to occur at Ala59.

Comparison of the Propeptide with Cathepsin K-Inhibitor Complexes. We previously reported the structure of a symmetric inhibitor, 1,3-bis[[$N^{(a)}$-[(phenylmethoxy)carbonyl]-L-leucyl]amino]-2-propanone, bound to cathepsin K (*14*, PDB code 1AU0). This inhibitor spans both the S- and S′-subsites and shows strong inhibition (K_i = 22 nM) (*14*). Comparing the positions of the propeptide and this inhibitor in the active site provides insight into interactions that are important for the inhibition of the enzyme. There is little correspondence between the main-chain positions of the propeptide and the inhibitor (Figure 5B). The hydrogen-bonding pattern observed between the propeptide and mCatK is more extensive than that observed in the inhibitor−cathepsin K complex where only one main-chain hydrogen bond is formed. However, occupation of the subsites by hydrophobic groups is similar. Starting at the S2′-subsite, the methionyl side chain of Met75 and the Cbz group of the inhibitor occupy nearly the same position. This is also true for the side-chain atoms of Leu78 of the propeptide and the inhibitor in the S2 specificity pocket. Valine80 and the second Cbz group from the inhibitor occupy the same position of the S4-subsite. So while the pattern of main-chain hydrogen bonding is not conserved, the binding of hydrophobic side-chain atoms in the subsites of the active site is preserved in the inhibitor and the propeptide.

We have reported the crystal structure of cathepsin K in complex with the epoxide inhibitor E-64, (l-3-carboxy-2,3-trans-epoxypropionyl-leucyl-amino(4-guanidino)butane) (ref 13, PDB code 1ATK). Comparing the structure of cathepsin K/E-64 with procathepsin K shows that E-64 also mimics the binding of the propeptide (Figure 5A). Like the propeptide, the inhibitor binds in the active site in the reverse

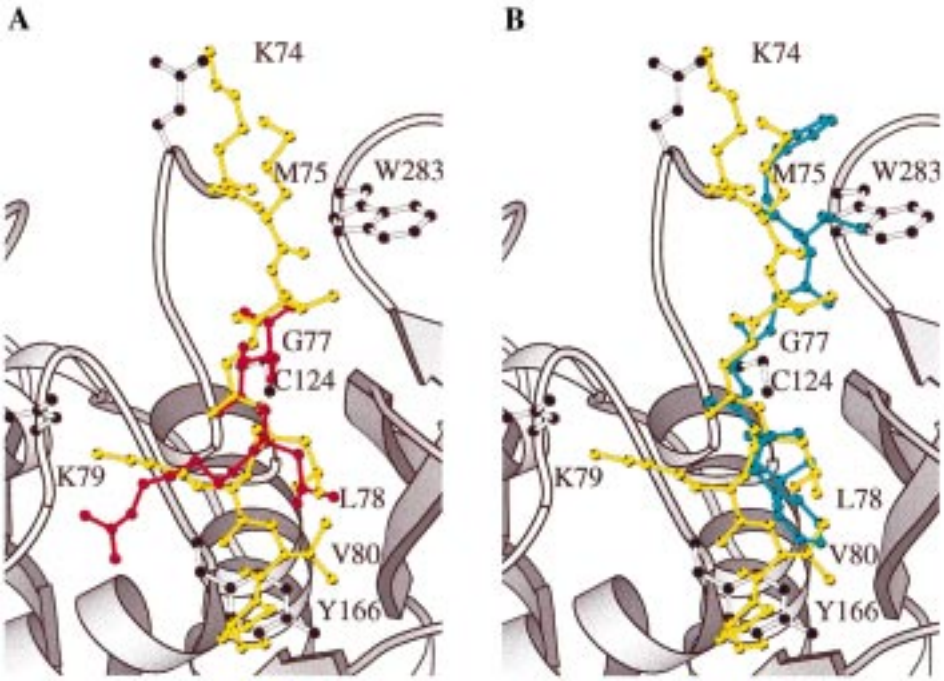

FIGURE 5: Comparison of cathepsin K−inhibitor complexes with pCatK. The inhibitors (A) E-64, red, and (B) 1,3-bis[[*N*$^{(a)}$-[(phenylmethoxy)-carbonyl]-L -leucyl]amino]-2-propanone, blue, are superimposed on the structure of the propeptide, colored yellow, in the active site. Figures drawn with BobScript (*26*).

direction to that postulated for a natural substrate. The leucine side chain from E-64 in the S2-subsite occupies an identical position to that of Leu78 in the propeptide. However, the guanidinium groups of E-64 and the amino group of Lys79 have slightly different positions. While both groups form a salt bridge with Asp160, the guanidinium group of E-64 also forms a hydrogen bond to the main-chain carbonyl of Glu158. Another difference at the active site between pCatK and the cathepsin K−inhibitor complexes is the conformation of the active site residue Tyr166. In the complex of cathepsin K with the symmetric inhibitor, Tyr166 points into the active site forming an aromatic stacking interaction with the Cbz group of the inhibitor. This tyrosine occupies the same position in the cathepsin K/E64 complex. In procathepsin K, Tyr166 is rotated 90° about $\chi 1$ to form a hydrophobic packing surface with Pro81 of the propeptide.

Differences between the Structures of Cathepsin K and Procathepsin K. The rms deviations between cathepsin K and the structurally equivalent atoms of pCatK are 0.47 and 1.56 Å for main-chain and side-chain atoms, respectively. There is no change in the overall conformation of the mature portion of the enzyme in procathepsin K. There are some changes in side-chain positions that are induced by interactions with the propeptide in the active site. Gln120 shifts so that it forms a hydrogen bond in the active site with Lys74; Tyr166 rotates to pack against Pro81; Asp160 shifts to form a salt bridge with Lys79, whereas in the structure of the cathepsin K/E-64 complex Asp160 forms hydrogen bonds with E-64 (*13*). Near the C-terminal segment of the propeptide, Arg210, Arg222, and Arg226 move to accommodate residues 90−95, and there is a shift in the position of Met313.

Comparison of the Procathepsins B, L, and K. The three-dimensional structures of human pCatB (9), rat pCatB (*10*), and human pCatL (*11*) have been reported. Procathepsin L and pCatK belong to the same subfamily, while pCatB belongs to a different subfamily of the cysteine protease family (*25*). The structure of pCatB differs from that of pCatL and pCatK in that is has a shorter propeptide that does not form a globular domain on its own. Procathepsin L and pCatK share higher amino acid sequence homology (68%) in their propeptides, and the propeptides have similar structures. Comparison of the similarities and differences between pCatL and pCatK in the region of the propeptide could further elucidate the mechanism of autocatalysis. Both pCatL and pCatK have the same three-dimensional fold. The rms difference between the main-chain atoms of the two proteins is 1.38 Å. The rms differences computed separately for the propeptide and mature peptide between the two proteins are 2.2 Å (residues 5−99) and 0.92 Å (residues 100−314), respectively. Procathepsin K and pCatL differ structurally in three regions: residues 17−27 (corresponding to a shift in the loop between helix 2 and helix 3), residues 67−73 (corresponding to a shift in the position of helix 3), and residues 80−91 (corresponding to a different conformation of the turn structure, with pCatL having four fewer residues). These three regions in pCatK and pCatL also exhibit higher B-factors.

The propeptide residues that lie in the active sites of pCatL and pCatK have the same main-chain conformation and form the same set of main-chain hydrogen bonds. Bound in the active site of pCatL is the sequence Val-Met-Asn-Gly-Leu-Gln-Asn compared to Lys-Met-Thr-Gly-Leu-Lys-Val in pCatK. In the S2′-, S1-, and S2-subsites the same side chains are found: Met, Gly, and Leu, respectively. In the S3-subsites both proenzymes use a positively charged side chain, Lys or Arg to form a salt bridge to Asp160. In the S1′-subsite of pCatL, the side chain of Asn hydrogen bonds to Asp260, whereas, in pCatK, the side chain from Thr76 hydrogen bonds with Val72.

Aside from the three regions 17−27, 67−73, and 80−91, the conformations of the propeptides of pCatK and pCatL are quite similar. This allows the comparison of the salt

bridges in pCatK and pCatL. Procathepsin K has five salt bridges in its globular domain compared with six in pCatL. The common salt bridges are Asp8 to Arg32 (Glu9-Arg32 in pCatL), Glu48 to His54 (Glu48-His54 in pCatL), Asp65 to Lys20 (Asp65-Arg20 in pCatL), and Glu70 to Arg31 (Glu70-Arg31 in pCatL). Procathepsin L does not contain the salt bridge Glu36 to Arg32. Instead, it has salt bridges Glu28 to Lys16 and Glu51 to Lys53. Near the C-terminus of the propeptide of pCatL, salt bridges are formed between Glu95 and Lys218 and between Glu95 and Lys222. In pCatK the salt bridges at Glu95 are not possible because one carboxylate oxygen of Glu95 forms a hydrogen bond with a carboxylate oxygen of Glu97 from a symmetry-related molecule in the crystal. The lack of salt bridges between Glu95 and Lys218 or Arg222 in pCatK may account for some of the disorder observed in this region of the molecule.

DISCUSSION

Structure of Procathepsin K. The structure of wild-type procathepsin K reveals a common fold observed with other procathepsins, pCatL and pCatB. A segment of the propeptide sits in the active site mimicking the substrate by positioning Leu and Met in the S2- and S2′-subsites, respectively. Cleavage of the propeptide is inhibited because the propeptide is bound in reverse orientation to that of a natural substrate. One of the four cleavage sites, Glu4, is not observed in the pCatK structure reported here. The sites Ser83 and Glu95 occur in the C-terminal portion of the propeptide. This region exhibits higher B-factors, consistent with flexibility (Figure 1) which might allow cleavage to occur readily at these sites. However, the fourth site, Ala59, occurs at the end of the β-strand that forms the β-sheet between the propeptide and mCatK. Cleavage at Ala59 places Leu58 in the S2 specificity pocket. Similarly, the cleavage site at Ser83 places Leu82 in the S2-subsite. If leucine is required in the S2-subsite, why does cleavage not occur at the other leucines in the propeptide sequence? Five other Leu residues occur in the propeptide sequence (Leu7, Leu33, Leu47, Leu51, and Leu63); however, cleavage does not occur at any of these sites. This can be explained in terms of secondary structure; these five Leu residues are located on α-helices, which could not be easily contorted to provide an extended conformation to fit in the active site. In contrast, Ala59 resides on a stretch of β-strand that, once rotated away from mCatK, could fit in the active site.

Activation of Procathepsin K. Autoactivation of pCatK requires a low-pH environment; this suggests the use of a pH-sensing mechanism to trigger autocatalysis. Several salt bridges are observed within the pCatK globular domain. Most of these salt bridges are conserved in the pCatL structure. However, pCatK and pCatL do not share any common salt bridges in the C-terminal domain of the propeptide. This is consistent with the data presented by Coulombe et al. (*11*) suggesting that the pH-sensing mechanism in pCatL occurs only in the region spanning residues 21−81. Furthermore, Coulombe et al. suggest that two residues, Asp65 and Glu70, are likely to participate in the pH-sensing mechanism. The conservation of salt bridges formed by these two residues in the pCatK and pCatL structures supports this hypothesis. In pCatK, the salt bridge between Asp65 and Lys20 is a part of several electrostatic and hydrophobic interactions in the globular domain of the propeptide (Figure 4). The protonation and disruption of one or more of these salt bridges at low pH could be quickly translated to the entire domain of the propeptide.

Autocatalysis is believed to proceed *in trans* where one pCatK molecule is cleaved by another activated molecule (*7*). The mechanism of activation for the first pCatK molecule is unknown, but it could be self-activated or cleaved by another enzyme. Nonetheless, activation of the proenzyme by mCatK would require the cleavage sites of the propeptide to be accessible to the active site of a neighboring mCatK molecule. Once cleavage occurs the propeptide must dissociate from mCatK. Preferred cleavage sites are observed during auto activation *in vitro* resulting in a 6.7 kDa fragment corresponding to residues 4−59, a 9.4 kDa fragment corresponding to residues 4−83, and a 27 kDa intermediate corresponding to residues 83−314. Fragments corresponding to residues 60−83 and 60−314 are not observed, probably as a result of proteolytic instability (M. S. McQueney, unpublished results). We can postulate the events that occur during activation of pCatK based on the location of these cleavage sites in our structure. One scenario is that an initial cut occurs at Ser83 allowing the propeptide to dissociate from the active site. This region of the propeptide appears flexible in our structure. Once the site is free, cleavage at Ala59 would occur if the globular domain shifts to allow access to Ala59. Once cleaved, two fragments, residues 4−59 and residues 60−83, could dissociate from mCatK. Another possible scenario would involve initial cleavage at Ala59. This would require a loosening of the structure, perhaps as a result of lowering the pH to weaken salt bridges. Further biochemical studies combined with site-directed mutagenesis will be required to determine the order of cleavage of the propeptide.

In regard to the mechanism of pH activation, one theory supposes that breaking salt bridges at low pH is enough to disrupt the tertiary fold and allow the propeptide to dissociate. In the structure of human pCatB, Podobnik et al. observe acidic residues at the N-termini of the two propeptide helices, suggesting that a change in pH affects the acidic residues on the surface and destabilizes the propeptide secondary structure triggering autoactivation (*9*). However, from the structure of rat pCatB, Cygler et al. suggest there are no specific interactions with acidic residues that could be responsible for the pH dependence of its autoactivation (*10*). Instead, Cygler et al. propose an overall loosening of the structure at low pH, thus increasing the mobility of the propeptide (*10*). The higher mobility of the C-terminal segment of the propeptide containing two of the four cleavage sites is consistent with this proposal. However, for the cleavage site at Ala59, the propeptide is closely associated with mCatK. A larger force would be needed to sever this hydrophobic contact. In the case of pCatK, we observe several networks in the globular domain of the propeptide, between salt bridges, hydrogen bonds, and aromatic stacking. The linkage between the electrostatic and hydrophobic interactions allows the disruption of electrostatic interactions to have disruptive effects on the hydrophobic packing of the propeptide. This would enhance the dissociation of the globular domain from mCatK.

Given that pCatB is from a different subfamily than pCatL and pCatK, more than one pH-sensing mechanism may have evolved among the subfamilies. This is consistent with the

difference in the overall structures of the propeptide globular domains in pCatB versus pCatL and pCatK. Procathepsin L and pCatK have larger globular domains with more helical character than pCatB. A concerted process consisting of disrupting salt bridges, which in turn disrupt a network of electrostatic and hydrophobic interactions, may be required to disrupt the large globular domain of pCatK permitting access to the Ala59 cleavage site and dissociation of the propeptide from mCatK.

REFERENCES

1. Li, Y. P., Alexander, M. B., Wucherpfennig, A. L., Chen, W., Yelick, P., and Stashenko, P. (1994) *Mol. Biol. Cell 5*, S335a.
2. Inaoka, T., Bilbe, G., Ishibashi, O., Tezuka, K., Kumegawa, M., and Kokubo, T. (1995) *Biochem. Biophys. Res. Commun. 206*, 89−96.
3. Tezuka, K., Tezuka, Y., Maejima, A., Sato, T., Nemoto, K., Kamioka, H., Hakeda, Y., and Kumegawa, M. (1994) *J. Biol. Chem. 269*, 1106−1109.
4. Brömme, D., and Okamoto, K. (1995) *Biol. Chem. Hoppe-Seyler 376*, 379−384.
5. 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.
6. Shi, G.-P., Chapman, H. A., Bhairi, S. M., DeLeeuw, C., Reddy, V. Y., and Weiss, S. J. (1995) *FEBS Lett. 357*, 129−134.
7. McQueney, M. S., Amegadzie, B. Y., D'Alessio, K., Hanning, C. R., McLaughlin, M. M., McNulty, D., Carr, S. A., Ijames, C., Kurdyla, J., and Jones, C. S.(1997) *J. Biol. Chem. 272*, 13955−13960.
8. Baron, R. (1989) *Anat. Rec. 224*, 317−324.
9. Podobnik, M., Kuhelj, R., Dolinar, M., Turk V., and Turk, D. (1997) *J. Mol. Biol. 271*, 774−88.
10. Cygler, M., Sivaraman, J., Grochulski, P., Coulombe, R., Storer, A. C., and Mort, J. S. (1996) *Structure 4*, 405−416.
11. Coulombe, R., Grochulski, P., Sivaraman, J., Menard, R., Mort, J. S., and Cygler, M. (1996) *EMBO J. 15*, 5492−5503.
12. Berti, P. and Storer, A. C.(1995) *J. Mol. Biol. 246*, 273−283.
13. Zhao, B., Janson, C. A., Amegadzie, B. Y., D'Alessio, K., Griffin, C., Hanning, C. R., Jones, C., Kurdyla, J., McQueney, M., Qiu, X., Smith, W. W. and Abdel-Meguid, S. S. (1997) *Nat. Struct. Biol. 4*, 109−111.
14. Yamashita, D. S., Smith, W. W., Zhao, B., Janson, C. A., Tomaszek, T. A., Bossard, M. J., Levy, M. A., Marquis, R. W., Oh, H.-J., Ru, Y., Carr, T. J., Thompson, S. K., Ijames, C. F., Carr, S. A., McQueney, M., D'Alessio, K. J., Amegadzie, B. Y., Hanning, C. R., Abdel-Meguid, S. S., DesJarlais, R. L., Gleason, J. G., and Veber, D. F. (1997) *J. Am. Chem. Soc. 119*, 11351−11352.
15. Brunger, A. T., Kuriyan, J., and Karplus, M. (1987) *Science 235*, 458−460.
16. Collaborative Computational Project, Number 4 (1994) *Acta Crystallogr. Sect. D 50*, 760−763.
17. Rice, L. M., and Brunger, A. T. (1994) *Proteins 19*, 277−290.
18. Brunger, A. T. (1992) *Nature 355*, 472−474.
19. Read, R. J. (1986) *Acta Crystallogr., Sect. A 42*, 140−149.
20. Kleywegt, G. J., and Brunger, A. T. (1996) *Structure 4*, 897−904.
21. Nicholls, A., Sharp, K. A., and Honig, B. (1991) *Proteins 11*, 281−296.
22. Drenth, J., Kalk, K. H., and Swen, H. M. (1976**)** *Biochemistry 15*, 3731−3738.
23. Bossard, M. J., Tomaszek, T. A., Thompson, S. K., Amegadzie, B. Y., Hanning, C. R., Jones, C., Kurdyla, J. T., McNulty, D. E., Drake, F. H., Gowen, M., and Levy, M. A. (1996) *J. Biol. Chem. 271*, 12517−12524.
24. Vernet, T., Berti, P. J., de Montigny, C., Musil, R., Tessier, D. C., Ménard, R., Magny, M.-C., Storer, A. C., and Thomas, D. Y. (1995) *J. Biol. Chem. 270*, 10838−10846.
25. Karrer, K. M., Peiffer, S. L., and DiTomas, M. E. (1993) *Proc. Natl. Acad. Sci. U.S.A. 90*, 3063−3067.
26. Esnouf, R. M. (1997*) J. Mol. Graphics 15*, 133−138.

BI9822271