Cathepsins X and B Can Be Differentiated through Their Respective Mono- and Dipeptidyl Carboxypeptidase Activities†

Christian Therrien,‡ Paule Lachance,‡ Traian Sulea,‡ Enrico O. Purisima,‡ Hongtao Qi,‡ Edmund Ziomek,‡ Alejandro Alvarez-Hernandez,§ William R. Roush,§ and Robert Ménard*,‡

Biotechnology Research Institute, National Research Council of Canada, Montreal, Quebec H4P 2R2, Canada, and Department of Chemistry, University of Michigan, Ann Arbor, Michigan 48109-1055

Received October 24, 2000; Revised Manuscript Received January 2, 2001

ABSTRACT: Several new cysteine proteases of the papain family have been discovered in the past few years. To help in the assignment of physiological roles and in the design of specific inhibitors, a clear picture of the specificities of these enzymes is needed. One of these novel enzymes, cathepsin X, displays a unique specificity, cleaving single amino acid residues at the C-terminus of substrates very efficiently. In this study, the carboxypeptidase activities and substrate specificity of cathepsins X and B have been investigated in detail and compared. Using quenched fluorogenic substrates and HPLC measurements, it was shown that cathepsin X preferentially cleaves substrates through a monopeptidyl carboxypeptidase pathway, while cathepsin B displays a preference for the dipeptidyl pathway. The preference for one or the other pathway is about the same for both enzymes, i.e., approximately 2 orders of magnitude, a result supported by molecular modeling of enzyme−substrate complexes. Cleavage of a C-terminal dipeptide of a substrate by cathepsin X can become more important under conditions that preclude efficient monopeptidyl carboxypeptidase activity, e.g., nonoptimal interactions in subsites S_2-S_1. These results confirm that cathepsin X is designed to function as a monopeptidyl carboxypeptidase. Contrary to a recent report [Klemencic, I., et al. (2000) *Eur. J. Biochem. 267,* 5404−5412], it is shown that cathepsins X and B do not share similar activity profiles, and that reagents are available to clearly distinguish the two enzymes. In particular, CA074 was found to inactivate cathepsin B at least 34000-fold more efficiently than cathepsin X. The insights obtained from this and previous studies have been used to produce an inhibitor designed to exploit the unique structural features responsible for the carboxypeptidase activity of cathepsin X. Although of moderate potency, this E-64 derivative is the first reported example of a cathepsin X-specific inhibitor.

Cysteine proteases (cathepsins) constitute an important group of enzymes involved in a number of physiological processes. Several novel sequences have been discovered in recent years, which brings the total number of human cathepsins to eleven. With the novel sequences available, and the resulting expanded information database for this group of enzymes, there is now a fair level of diversity that can be found in tissue distribution and physiological roles (*1*−*3*). Evidence has been accumulating also that cysteine proteases can play a role in several pathological conditions (*4*). Of these, cathepsin K represents at the moment the best validated drug design target for the pharmaceutical industry. This is due to the clear demonstration of the role of cathepsin K in bone remodeling, and the finding that a genetic bone disorder, pycnodysostosis, results from mutations in the cathepsin K gene (*5*). Cathepsins have also been linked or speculated to be involved in a number of degenerative diseases, in cancer, and in diseases of the immune system (*4*). In the cases where the contribution of the cysteine protease to a disease condition results from elevated or uncontrolled proteolytic activity, these enzymes constitute attractive targets for development of inhibitors as potential therapeutic agents. Although restricted expression patterns exist, the increasing number of cysteine proteases is likely to render more challenging the task of designing inhibitors specific for a given enzyme. This is particularly important considering that the information available to date indicates relatively broad, overlapping specificities for the mature enzymes (*6*). It is therefore of major importance to uncover and better define the structural and functional features that could allow these various proteases to be differentiated.

Unlike many of the recently discovered cathepsins of the papain family which display (or are predicted to display) similar specificity patterns (*1*), cathepsin X exhibits a very strong preference for hydrolyzing substrates through a carboxypeptidase[1] pathway as opposed to endopeptidase activity (*7*). Even though cathepsin X has been discovered only recently, it has already been relatively well characterized at the gene (*8*−*10*), functional (*7*) and structural (*11*, *12*)

† NRCC Publication no. 42969.

* Author to whom correspondence should be addressed. E-mail: robert.menard@nrc.ca.

‡ Biotechnology Research Institute, National Research Council of Canada.

§ Department of Chemistry, University of Michigan.

[1] In the text, dipeptidyl carboxypeptidase refers to the cleavage of a C-terminal dipeptide from a substrate, while the cleavage of a single C-terminal residue is referred to either as a carboxypeptidase step, or a monopeptidyl carboxypeptidase step. In certain cases, for greater clarity, an arrow is used to indicate the scissile peptide bond.

10.1021/bi002460a CCC: $20.00
Published on Web 02/06/2001

levels. Cathepsin X contains unique features that clearly distinguish it from other human cysteine proteases (*8*, *9*). One of these distinctive features is a three amino acid residue insertion in a highly conserved region between the glutamine (Gln22) of the putative oxyanion hole and the active site cysteine (Cys31). This motif is present in a number of species homologues of cathepsin X (*13*) and can be considered as a "signature" for this subgroup of enzymes. On the basis of an homology model built from available 3D structures, the three residue insertion was found to be located in the primed region of the binding cleft as part of a surface loop corresponding to residues His23 to Tyr27, which was termed the "mini-loop" (*7*). From the model, it was predicted that the mini-loop may confer exopeptidase activity to cathepsin X, which was confirmed by kinetic studies using synthetic substrates. More striking was the finding that the enzyme displays extremely low activity against the extended endopeptidase substrate Abz-AFRSAAQ-EDDnp,[2] as well as against the substrate Cbz-FR-MCA, the most commonly used substrate for cysteine proteases of the papain family. The model, and in particular the nature and exact positioning of the mini-loop, was later confirmed by the determination of the three-dimensional structures of procathepsin X (*11*) and mature cathepsin X (*12*) at resolutions of 1.7 and 2.7 Å, respectively.

The superposition of the cathepsin X and cathepsin B structures indicates that His23 of cathepsin X occupies a region in space which partially overlaps with His110 of cathepsin B, a residue which is considered to be responsible for the exopeptidase activity of the latter enzyme (*14*). Cathepsin B is known to hydrolyze substrates through a dipeptidyl carboxypeptidase pathway, and also displays a lower but significant endopeptidase activity (*15*, *16*). In a recent study, it was reported that cathepsin X can also display dipeptidyl carboxypeptidase activity similar to cathepsin B (*12*, *17*). A very significant finding also reported is the observation that cathepsin X is apparently readily inhibited by CA074, an inhibitor considered to be specific for cathepsin B. This is a very important issue, taking into account that putative physiological or pathological roles for cathepsin B have often been assigned based on the dipeptidyl carboxypeptidase activity of the enzyme and the use of the CA074 inhibitor. Overall, it was indicated that cathepsins X and B share similar activity profiles, making it difficult to differentiate the two enzymes (*12*). However, these results are difficult to reconcile with the structure of cathepsin X, where the mini-loop clearly occludes the S_2' binding site (*7*, *11*, *12*).

In this study, the carboxypeptidase activities and substrate specificity of both cathepsins X and B have been investigated in detail and compared. The results confirm the uniqueness of the cathepsin X carboxypeptidase activity. Moreover, it will be shown that contrary to earlier reports (*12*, *17*), cathepsins X and B can be differentiated through their activity profile against substrates and inhibitors, and that reagents are available to clearly distinguish the two enzymes. The insights obtained from this and previous studies have been used to produce the first generation of a cathepsin X-specific inhibitor designed to exploit the unique structural features underlying the carboxypeptidase activity of the enzyme.

EXPERIMENTAL PROCEDURES

Materials. Recombinant cathepsin B and L were expressed and purified as described previously (*16*, *18*). The vector (pPic9) and *P. pastoris* strain GS115 were purchased from Invitrogen Corp. (San Diego, CA). The substrates containing the Abz/F(4NO2) quenched pair and hydrolysis products used for controls in HPLC [Abz-FR, F(4NO$_2$)F, F(4NO$_2$), F] were all obtained from Enzyme Systems Products (Livermore, CA). Cbz-FR-MCA and the irreversible inhibitor E-64 were purchased from IAF Biochem International Inc. (Laval, Québec). CA-074 was obtained from Peptides International (Louisville, KY).

Expression and Purification of Recombinant Human Cathepsin X. Human procathepsin X was expressed in *P. pastoris* as an α-factor fusion construct as described previously (*7*). The culture medium was centrifuged at 4500*g* for 30 min at 4 °C, and the supernatant (1.2 L) concentrated to 100 mL using a TCF-10 diafiltration system equipped with a YM-10 membrane (Amicon). The concentrated crude culture filtrate was applied on a Q-Sepharose column (2.6 × 10 cm) equilibrated with 20 mM Tris-HCl buffer pH 8.1, and elution was carried out with a linear gradient of 0 to 1.0 M NaCl in the same buffer. Fractions containing the proenzyme were concentrated 10-fold and the pH of the solution was adjusted to 5.0 by diluting this sample in a 20 mM acetate buffer. The sample was applied on a SP-Sepharose column (2.6 × 10 cm) equilibrated with 20 mM acetate buffer pH 5.0. The column was washed with 100 mL of 20 mM acetate buffer pH 5.0 and elution was carried out using a linear 0 to 1.0 NaCl gradient. The presence of procathepsin X in the fractions was detected by SDS−PAGE and measurement of activity under conditions for processing to the mature enzyme. The fractions containing proenzyme were concentrated and the buffer exchanged to 20 mM phosphate pH 7.0 in 1.7 M ammonium sulfate. Approximately 10 mg of partially purified protein was applied on an Alkyl-Superose HR10/10 column (Pharmacia). The proenzyme was eluted with a linear gradient of 1.7 to 0 M ammonium sulfate in 20 mM phosphate buffer pH 7.0. The fractions eluting at approximately 1.2 M ammonium sulfate were analyzed by SDS−PAGE and pooled according to purity and protein concentration. The sample was concentrated to 2 mL using Centriprep-10 (Amicon) and further purified on a Superdex-75 column (1.6 × 60 cm) equilibrated with 150 mM NaCl in 50 mM phosphate buffer pH 7.0. The purified procathepsin X fractions were analyzed by SDS−PAGE and pooled according to their purity.

Production of mature cathepsin X from the proenzyme was achieved by incubating procathepsin X with small amounts

[2] Abbreviations: Abz-FRF(4NO$_2$), *o*-aminobenzoyl-L-phenylalanyl-L-arginyl-4-nitro-L-phenylalanine; Abz-RRF(4NO$_2$), *o*-aminobenzoyl-L-arginyl-L-arginyl-4-nitro-L-phenylalanine; Abz-RFF(4NO$_2$), *o*-aminobenzoyl-L-arginyl-L-phenylalanyl-4-nitro-L-phenylalanine; Abz-AAF(4NO$_2$), *o*-aminobenzoyl-L-alanyl-L-alanyl-4-nitro-L-phenylalanine; Abz-FRF(4NO$_2$)F, *o*-aminobenzoyl-L-phenylalanyl-L-arginyl-4-nitro-L-phenylalanyl-L-phenylalanine; Abz-FRFF(4NO$_2$), *o*-aminobenzoyl-L-phenylalanyl-L-arginyl-L-phenylalanyl-4-nitro-L-phenylalanine; Ac-FRF, acetyl-L-phenylalanyl-L-arginyl-L-phenylalanine; Ac-FRFF, acetyl-L-phenylalanyl-L-arginyl-L-phenylalanyl-L-phenylalanine; Cbz-FR-MCA, carbobenzoxy-L-phenylalanyl-L-arginine 4-methylcoumarinyl-7-amide; Abz-AFRSAAQ-EDDnp, *o*-aminobenzoyl-L-alanyl-L-phenylalanyl-L-arginyl-L-seryl-L-alanyl-L-alanyl-L-glutamine N-(ethylenediamine)-2,4-dinitrophenyl amide; E-64, 1-[[(L-*trans*-epoxysuccinyl)-L-leucyl]amino]-4-guanidinobutane; *n*PrNH-(2S,3S)-*t*Eps-Ile-OH, [L-3-*trans*-(propylcarbamoyl)oxirane-2-carbonyl]-L-isoleucine; *n*PrNH-(2S,3S)-*t*Eps-Ile-Pro-OH or CA074, [L-3-*trans*-(propylcarbamoyl)oxirane-2-carbonyl]-L-isoleucyl-L-proline; MMTS, methyl methanethiolsulfonate.

of cathepsin L as described previously (*7*). Typically, 2 mL of purified procathepsin X (2.5 mg/mL) was incubated with 100 mM acetate buffer pH 5.0 containing 2 mM DTT and 25 nM cathepsin L. The reaction was stopped by decreasing the temperature to 4 °C, followed by concentration of the sample 5-fold using a Biomax-10K system (Millipore). The pH of the sample was adjusted to 7.0 with phosphate buffer (approximately 50 mM final concentration) and the sample was loaded on a DEAE-Sepharose column (2.6 × 10 cm) equilibrated with 20 mM phosphate buffer pH 7.0. The mature form of enzyme was eluted with a linear gradient (0 to 1.0) of NaCl in the same buffer. Since cathepsin L is highly unstable at neutral pH, adjusting the pH to 7.0 coupled to the ion exchange purification step ensures that the small amounts of cathepsin L used to process procathepsin X will not be present in the purified mature cathepsin X preparations. The eluted fractions were analyzed by SDS−PAGE and tested for activity against the substrate Abz-FRF($4NO_2$). Enzyme concentration was determined from the initial rates of substrate hydrolysis, and by titration with E-64. Purified cathepsin X was stored at 4 °C in the elution buffer containing 100 μM of MMTS. The yield of mature cathepsin X was approximately 5 mg/L of BMGY (buffered glycerol-complex medium) culture medium.

Kinetic Measurements. To monitor enzyme activity and evaluate the substrate specificity of cathepsin X, quenched fluorogenic substrates containing the Abz (fluorescent group) − F($4NO_2$) (quencher) pair were used. The rate of substrate hydrolysis was measured by monitoring the increase in Abz fluorescence (λ_{exc}= 320 nm, λ_{em}= 420 nm) as a function of time on a SPEX Fluorolog-2 spectrofluorometer as previously described (*19*). The kinetic measurements were carried out at 25 °C in the presence of 50 mM sodium citrate (pH 5.0), 2 mM DTT, 0.2 M NaCl, 1 mM EDTA, and 3% DMSO. The concentration of active enzyme was determined by titration using E-64 (*20*). Initial rates were determined at substrate concentrations much lower than K_M, and k_{cat}/K_M values were obtained by dividing the initial rates by enzyme and substrate concentrations. For the multiple nonlinear regression analysis of the product vs time progress curves, the program Grafit v 3.0 (R. J. Leatherbarrow) was used. Second-order rate constants for enzyme inactivation by the irreversible inhibitors were determined as described previously (*21*). The same experimental conditions were used for cathepsins B and L.

Identification and Monitoring of Substrate Hydrolysis Products by HPLC. The reaction mixture was composed of either 15 μM or 150 μM Abz-FRF($4NO_2$)F, 54 nM cathepsin X, 2 mM BME in 100 mM citrate buffer pH 5.0 containing 400 mM NaCl, 2 mM EDTA, and 3% DMSO. Samples (500 μL) were withdrawn after various periods of time and mixed with 50 μL of 5% TFA. The reaction products were separated by reverse-phase HPLC on a Water Spherisorb ODS-2 C18 column (5 μm, 25 × 0.46 cm) equilibrated with 0.1% TFA and a combination of linear and step-gradient (5 min 0.1% TFA, 15 min 0 to 60% acetonitrile gradient, 5 min 60% acetonitrile, 10 min 60 to 100% acetonitrile gradient and 5 min 100% acetonitrile). The identity of the substrate and reaction products was assigned by performing control runs using the intact substrate and the following products obtained by synthesis: Abz-FRF($4NO_2$), Abz-FR, F($4NO_2$)F, F($4NO_2$), and F. Substrate and products were detected by UV (200−400 nm, Waters Photodiode Array Detector, Waters, Mississauga, Ontario, Canada) and fluorescence (420-AC Fluorescence Detector, Waters). The retention times of the products and subtrate were such that they could easily be separated and differentiated, and calibration curves were determined to accurately calculate the concentration of the substrate and each of the reaction products in the course of the hydrolysis reaction.

Molecular Modeling. The crystal structures of human procathepsin X (*11*) (PDB code 1deu) and human cathepsin B (*22*) (PDB code 1csb) were used for substrate docking. The proregion segment of cathepsin X, the CA030 inhibitor cocrystallized with cathepsin B and the water molecules were removed, and the hydrogen atoms were added using SYBYL 6.6 (Tripos, Inc., St. Louis, MO). The histidine residues were protonated and the catalytic cysteine was modeled in the thiolate form. The monopeptidyl carboxypeptidase substrate Ac-FR↓F and the dipeptidyl carboxypeptidase substrate Ac-FR↓FF were individually docked in the binding sites of human cathepsins X and B as noncovalent Michaelis complexes. The substrate binding mode was assigned as described previously (*7*, *11*). Energy minimization was carried out in SYBYL 6.6 using AMBER 4.1 all-atom force field (*23*) with the Powel minimizer, a 4*R* distance dependent dielectric constant and an 8 Å nonbonded cutoff up to an rms gradient of 0.05 kcal/(mol·Å). Due to unrealistically high steric repulsion between the Phe(P_2') residue of the Ac-FR↓FF substrate and the His23 and Tyr27 side chains of cathepsin X, substrate docking was carried out in this case by a conformational search using a Monte Carlo with energy minimization procedure (*24*). Starting conformations for minimization were generated by randomly perturbing one or more dihedral angles of the side chains of the His23, Tyr27, and Phe(P_2') residues as well as crank-shaft rotation of the peptide unit Phe(P_1')-Phe(P_2'). A total of 400 Monte Carlo steps were carried out, each followed by an energy minimization in which the substrate and the His23 and Tyr27 residues were allowed to relax.

The enzyme−substrate complexes were used for binding free energy calculations. The following equation was used to evaluate the relative binding free energy of two different substrates that bind separately to a given enzyme:

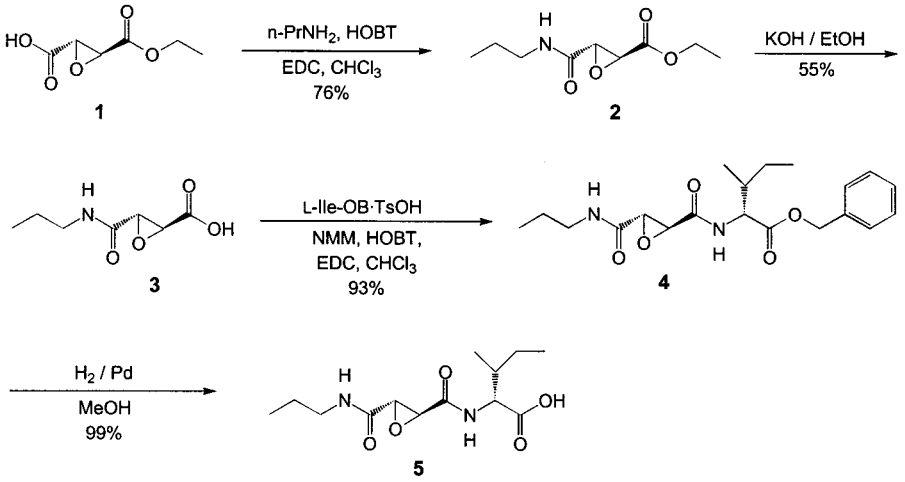
$$\Delta\Delta G_{binding} = \Delta SIE + \Delta SCE_{protein} + T\Delta\Delta S_{ligand} \quad (1)$$

where ΔSIE is the relative solvated interaction energy between the protein and each ligand, $\Delta SCE_{protein}$ is the relative solvated conformational energy of distinct protein conformations, and $\Delta\Delta S_{ligand}$ is the relative binding configurational entropy of the ligands. ΔSIE is calculated as a scaled sum of the relative van der Waals intermolecular interaction energy (ΔE_{vdw}), relative Coulomb electrostatic intermolecular interaction energy (ΔE_{coul}), relative change in the electrostatic component of solvation between the bound and free states ($\Delta\Delta G_{elst}^{solvation}$) and relative change in the nonelectrostatic component of solvation between the bound and free states ($\Delta\Delta G_{non\text{-}elst}^{solvation}$).

$$\Delta SIE = \alpha(\Delta E_{vdw} + \Delta E_{coul} + \Delta\Delta G_{elst}^{solvation} + \Delta\Delta G_{non\text{-}elst}^{solvation}) \quad (2)$$

E_{vdw} and E_{coul} were computed using the AMBER force field

Scheme 1



with infinite cutoff and a dielectric constant of 2. $\Delta G_{elst}^{solvation}$ was taken as the difference in the reaction field energies between the bound and free states. Reaction field energies were calculated with the boundary element method (BEM) within the continuum dielectric model using the BRI BEM program (*25*, *26*), SIMS molecular surface meshing algorithm (*27*), AMBER van der Waals radii with a few modifications, AMBER partial atomic charges, and interior and exterior dielectric constants of 2 and 78.5, respectively. $\Delta G_{non\text{-}elst}^{solvation}$ was computed as a sum of cavity and dispersion energies compatible with the AMBER force field and BEM reaction field energies. Linear surface area relationships with single and atom type dependent coefficients were used to derive the cavity and dispersion terms, respectively (unpublished parametrization). α is an empirical scaling factor with the value of 0.23 which was determined from a least-squares-fit of ΔSIE against the experimental relative binding free energies of a diverse set of ligand-protein complexes (unpublished experiments). $\Delta SCE_{protein}$ is calculated as a scaled sum of the relative internal energy ($\Delta E_{protein}^{intra}$) and relative solvation free energy ($\Delta G_{protein}^{solvation}$) for different conformations of the protein.

$$\Delta SCE_{protein} = \alpha(\Delta E_{protein}^{intra} + \Delta G_{protein}^{solvation}) \qquad (3)$$

$E_{protein}^{intra}$, which includes bond stretching, angle bending, and torsional energies as well as van der Waals and electrostatic energies, was calculated with the AMBER force field using an infinite cutoff and a dielectric constant of 2. $G_{protein}^{solvation}$ was taken as the sum of the reaction field energy and the nonelectrostatic solvation energy of a given protein conformation. The same α scaling factor of 0.23 was used to calculate $\Delta SCE_{protein}$. $T\Delta\Delta S_{ligand}$ is an empirical term which describes the relative loss of configurational entropy between two different ligands due to the loss of main chain (ΔS_{mc}) and side chain (ΔS_{sc}) torsional freedom upon complexation to an enzyme:

$$T\Delta\Delta S_{ligand} = T\Delta\Delta S_{mc} + T\Delta\Delta S_{sc} \qquad (4)$$

$T\Delta S_{mc}$ was assumed to be 2 kcal/mol/residue and $T\Delta S_{sc}$ was given 0.54 kcal/mol/rotor (*28*).

Synthesis of nPrNH-(2S,3S)-tEps-Ile-OH. ^{1}H NMR spectra were measured at 500 MHz on a Varian Inova 500 instrument. Residual Me_4Si in DMSO-d_6 was used as internal reference. ^{13}C NMR spectra were measured at 100 MHz using residual DMSO (39.5 ppm) as internal reference. Infrared spectra were recorded on a Perkin-Elmer Spectrum 1000 spectrophotometer. High-resolution mass spectra (HRMS) were measured on a Micromass Corp. VG 70-250-S spectrometer at the University of Michigan Mass Spectrometry Laboratory. Chemical ionization (CI) mass spectra were obtained using NH_3 as the reagent gas. Optical rotations were measured on a Rudolph Autopoll III polarimeter using a quartz cell of 1 mL capacity and a 10 cm path length. Melting points were determined in a Thomas-Hoover apparatus and are uncorrected. L-Isoleucine benzyl ester·*p*-toluenesulfonate salt, 1-hydroxybenzotriazole hydrate (HOBT·H_2O), and 1-(3-dimethylaminopropyl)-3-ethylcarbodiimide hydrochloride (EDC·HCl) were purchased from Advanced ChemTech (Kentucky). *N*-methylmorpholine (NMM) and chloroform (stabilized with amylene) were dried and stored over 4 Å molecular sieves. All reaction intermediates in the synthesis scheme were characterized in detail, but only the results for the final compound [the inhibitor *n*PrNH-(2S,3S)-*t*Eps-Ile-OH, **5**] are presented in Scheme 1.

To a solution of 2.00 g (12.5 mmol) of (2S,3S) epoxysuccinate monoethyl ester **1** (*29*, *30*), 1.04 mL of propylamine (12.5 mmol) and 1.91 g of HOBT (12.5 mmol) in 20 mL of chloroform at 0 °C, was added EDC (2.64 g, 13.7 mmol) slowly in five portions. The reaction mixture was stirred at 0 °C for 1.5 h then allowed to warm to 23 °C and stirred for 12 h. The reaction mixture was partitioned between ethyl acetate and water. The organic portion was consecutively washed with 10% HCl, satd. $NaHCO_3$ and brine. After drying with Na_2SO_4 the solvent was evaporated in vacuo and the residue purified by flash chromatography (SiO_2, hexane:ethyl acetate 7:3) to give 1.91 g of **2** as a white solid (76%). Ethanolic KOH (8.9 mL of a 1N solution, 8.9 mL) was added to a solution of 1.78 g of ester **2** (8.9 mmol) in 100 mL of absolute ethanol at 0 °C. After 15 min the reaction mixture was allowed to warm to 23 °C and stirred for 1.5 h. The solvent was evaporated in vacuo and the residue dissolved in cold water. After acidification to pH 2 with 10% HCl, the product was extracted with ethyl acetate and the solvent dried over Na_2SO_4 and partially evaporated. Recrystallization of the acid in hexanes−ethyl acetate gave 0.84 g (55%) of **3** as white flakes. To prepare derivative **4**, EDC (0.63 g, 3.3

mmol) was added to a solution of the acid **3** (0.52 g, 3.0 mmol), L-isoleucine benzyl ester·*p*-toluene sulfonic acid salt (1.18 g, 3.0 mmol), NMM (0.33 mL, 3.0 mmol) and HOBT (0.46 g, 3.0 mmol) in 10 mL of chloroform at 0 °C. The reaction mixture was stirred at 0 °C for 1.5 h then allowed to warm to 23 °C and stirred for 12 h. The reaction mixture was diluted with 50 mL of ethyl acetate and consecutively washed with 10% HCl, saturated $NaHCO_3$, and brine. After drying with Na_2SO_4 the solvent was evaporated in vacuo and the residue purified by flash chromatography (SiO_2, hexane:ethyl acetate 4:6) to give 1.05 g (93%) of **4** as a clear oil that solidifies on standing. Finally, the inhibitor *n*PrNH-(2S,3S)-*t*Eps-Ile-OH (**5**) was obtained by bubbling hydrogen through a suspension of **4** (0.68 g, 1.8 mmol), and 0.1 g of 10% palladium on carbon in 30 mL of methanol under vigorous magnetic stirring. After 1 h the catalyst was filtered through a pad of Celite and the solvent evaporated in vacuo. After drying the residue under high vaccum, 0.51 g (>99%) of **5** was obtained as a white solid which required no further purification: mp 57 °C; $[\alpha]_D = +94°$ ($c = 2.0$; CH_3OH); ^{1}H NMR (DMSO-d_6) δ 12.78 (s, 1H), 8.66 (d, $J = 7.5$ Hz, 1H), 8.38 (t, $J = 5.5$ Hz, 1H), 4.23 (dd, $J_a = 8.5$ Hz, $J_b = 6$ Hz, 1H), 3.72 (d, $J = 1.5$ Hz, 1H), 3.48 (d, $J = 1.5$ Hz, 1H), 3.04 (m, 2H), 1.82 (m, 1H), 1.42 (m, 3H), 1.21 (m, 1H), 0.86 (m, 9H); ^{13}C NMR (DMSO-d_6) δ 172.50, 165.88, 165.70, 56.46, 52.78, 52.27, 40.39, 36.37, 24.64, 22.16, 15.58, 11.36, 11.33; IR (KBr) cm^{-1} 3300, 2968, 1734, 1655, 1542, 1459, 1247, 1209, 1148, 896; HRMS, CI (NH_3) [M + H]$^+$ calcd for $C_{13}H_{23}N_2O_5$, 287.1601; found, 287.1619.

RESULTS AND DISCUSSION

Hydrolysis of Quenched Fluorogenic Substrates by Cathepsin X. The proteolytic activity of human cathepsin X has been investigated in a previous study, where it was found that cathepsin X can display excellent carboxypeptidase activity against the Abz-tripeptide substrate Abz-FRF($4NO_2$), but almost negligeable endopeptidase activity against Abz-AFRSAAQ-EDDnp or Cbz-FR-MCA (*7*). Cathepsin X was also found to hydrolyze an Abz-tetrapeptide substrate, similar to the substrates normally used to characterize the dipeptidyl carboxypeptidase activity of cathepsin B. However, the preliminary data suggested that hydrolysis might occur through a sequential carboxypeptidase pathway (*7*). In a recent paper, cathepsin X was reported to display both mono- and dipeptidyl carboxypeptidase activities (*12*). In this study, we performed a more in-depth investigation of the relative contribution of the two hydrolysis pathways at the substrate C-terminus by cathepsin X. For that purpose, we used mainly two Abz-tetrapeptide substrates that differ only by the position of a $4NO_2$ group on a phenylalanine side chain, i.e., Abz-FRF($4NO_2$)F and Abz-FRFF($4NO_2$). The mechanism describing the overall proteolytic process with the substrate Abz-FRF($4NO_2$)F is illustrated in Figure 1. The substrate can be hydrolyzed through two parallel reactions: a dipeptidyl carboxypeptidase pathway, and/or through two consecutive monopeptidyl carboxypeptidase steps. Cleavage of the C-terminal dipeptide will result in an increase in Abz fluorescence since the Abz and F($4NO_2$) residues will be separated. However, if the substrate is hydrolyzed by two consecutive monopeptidyl carboxypeptidase reactions, fluorescence will increase only after the second cleavage and a lag phase in the fluorescence vs time progress curve will be observed if $k_{1a} < k_{1b}$ (Figure 1). As illustrated in Figure 2a, a lag phase is observed in the progress curve for hydrolysis of Abz-FRF($4NO_2$)F by cathepsin X, indicating that the

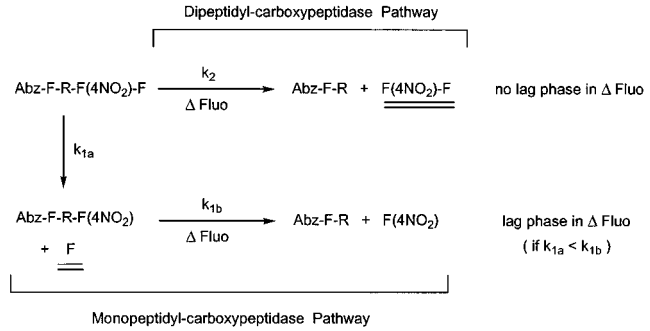


FIGURE 1: Kinetic model for hydrolysis of Abz-FRF($4NO_2$)F by cathepsin X. The rate constants k represent the apparent first-order rate constants (k_{cat}/K_M[E]) for each step under conditions where [S] $\ll K_M$. k_{1a} and k_{1b} represent the rates for the two sequential monopeptidyl carboxypeptidase steps, while k_2 corresponds to the dipeptidyl carboxypeptidase cleavage rate. The steps that are accompanied by a change in fluorescence of the substrate are indicated by the symbol ΔFluo. The reaction products that allow the evaluation of the relative rates of mono- vs dipeptidyl carboxypeptidase activities by HPLC measurements are double-underlined.

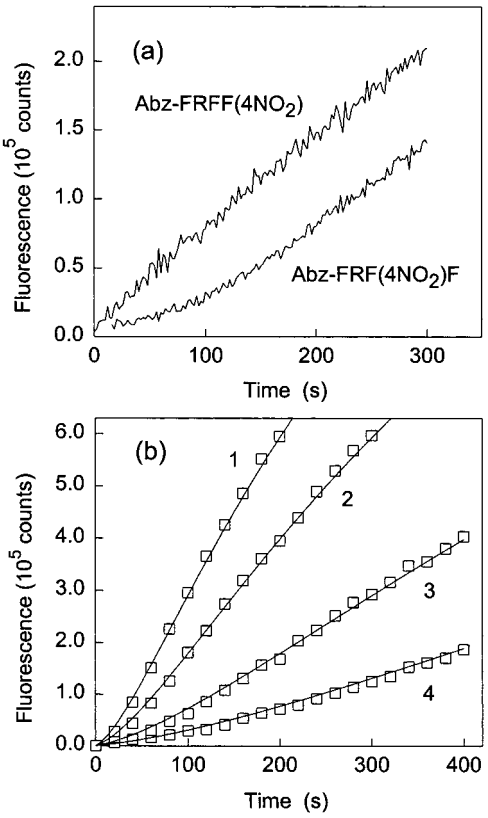


FIGURE 2: Fluorescence vs time progress curves for substrate hydrolysis by cathepsin X. (a) Hydrolysis of the substrates Abz-FRF($4NO_2$)F ([cathepsin X] = 45 nM) and Abz-FRFF($4NO_2$) ([cathepsin X] = 90 nM) at pH 5.0. Different cathepsin X concentrations were used for better graphical visualization of the presence or absence of a lag phase. (b) Progress curves obtained for hydrolysis of Abz-FRF($4NO_2$)F at various enzyme concentrations (1, 47 nM; 2, 90 nM; 3, 180 nM; 4, 270 nM). For greater clarity, only one out of 10 data points is displayed, even though all points have been used for the nonlinear regression analysis. Substrate concentration was 1.0 μM in all cases. The curves corresponding to the best fit to eq 5 are represented.

monopeptidyl carboxypeptidase pathway contributes very significantly to the overall hydrolysis. With the substrate Abz-FRFF($4NO_2$), no lag phase is observed since hydrolysis by either pathway will cause the separation of the Abz − F($4NO_2$) pair. From initial rate measurements with this substrate, an apparent k_{cat}/K_M value of 4.24×10^3 M^{-1} s^{-1} (which corresponds to the sum of the individual k_{cat}/K_M for the mono- and dipeptidyl carboxypeptidase reactions) was obtained. By comparison, cathepsin B can hydrolyze this substrate with $k_{cat}/K_M = 2.87 \times 10^6$ M^{-1} s^{-1}, a value 680-fold higher than for cathepsin X.

The presence and extent of the lag phase is a reflection of the relative contribution of mono- and dipeptidyl carboxypeptidase activities and can be used to obtain the rate constants for the two reactions by nonlinear regression to the proper equation. By working under conditions where the substrate concentration is much lower than K_M for each step involved, the appearance of product (fluorescence) as a function of time is described by eq 5.

$$[P] = [S_0]\left[1 - \left(\frac{(k_{1b} - k_2)e^{-(k_{1a}+k_2)[E]t} - k_{1a}\, e^{-k_{1b}[E]t}}{k_{1b} - (k_{1a} + k_2)}\right)\right] \quad (5)$$

where [P], [E], and [S_0] are the product (expressed in fluorescence units), enzyme and initial substrate concentrations, t is the time, and k_{1a}, k_{1b}, and k_2 are pseudo-first-order rate constants corresponding to the value of (k_{cat}/K_M[E]) for each step in the mechanism described in Figure 1. The kinetic parameters for the various proteolytic steps involved in hydrolysis of Abz-FRF($4NO_2$)F by cathepsin X were determined by nonlinear regression analysis of the data to eq 5. To obtain more reliable kinetic parameters from the data fitting procedure, a series of progress curves were recorded at various enzyme concentrations and a multiple nonlinear regression analysis was performed. Typical progress curves obtained at varying concentrations of enzyme are shown in Figure 2b, and the curves representing the best fit are shown.

To further verify the relative contributions of the mono- and dipeptidyl carboxypeptidase pathways to hydrolysis of Abz-FRF($4NO_2$)F, HPLC was used to monitor the substrate and products in the time course of the hydrolysis by cathepsin X. The results are shown in Figure 3. Even though all products could be monitored in this assay, only the concentrations of F and F($4NO_2$)F are presented in Figure 3. It can be shown that the relative concentration of these two products ([F]/[F($4NO_2$)F]) will reflect the ratio of k_{cat}/K_M for the mono- and dipeptidase steps (i.e., k_{1a}/k_2) if $k_{1b} \gg k_{1a}$, a condition that is met in the present case. As seen in Figure 3, a ratio of 2.3 for [F]/[F($4NO_2$)F] was obtained, in agreement with the ratio of 2.9 for k_{1a}/k_2 determined by the curve-fitting procedure described earlier. Therefore, a value of 2.6 was used as the ratio of k_{cat}/K_M values for the mono- and dipeptidase steps. The kinetic data are reported in Table 1. For the hydrolysis of a single C-terminal residue in Abz-FRF($4NO_2$)F (the monopeptidyl carboxypeptidase step), $k_{cat}/K_M = 5600$ M^{-1} s^{-1}, as compared to $k_{cat}/K_M = 2150$ M^{-1} s^{-1} for hydrolysis of the C-terminal dipeptide. This latter value is much lower (by almost 2000-fold) than for the analogous reaction with cathepsin B.

Positional vs Sequence Specificity of Cathepsin X. It must be noted that in the hydrolysis of Abz-FRF($4NO_2$)F by cathepsin X, the nature of the residues that interact in subsites S_2 and S_1 vary depending on the position of the bond being cleaved. As illustrated in Table 1, when a dipeptide is cleaved from the C-terminus of the substrate, the P_2-P_1 residues are F-R (Table 1). However, in the first monopeptidyl carboxypeptidase step, R and F($4NO_2$) are found in P_2 and P_1. A better evaluation of the relative ability of cathepsin X to cleave a substrate by a mono- vs dipeptidyl carboxypeptidase pathway can be done if most of the substrate residues interacting with the enzyme are kept unchanged [i.e., Abz-F-R-F($4NO_2$) in P_3-P_2-P_1-P_1'). This is achieved by comparing the hydrolysis of the C-terminal dipeptide in Abz-FRF($4NO_2$)F

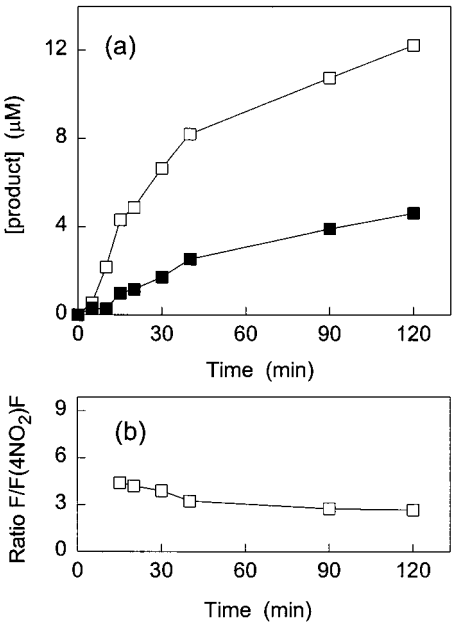


FIGURE 3: HPLC analysis of the hydrolysis products of Abz-FRF($4NO_2$)F by cathepsin X at pH 5.0. (a) Concentration of the products F($4NO_2$)F (black squares) and F (white squares) as a function of time. (b) Ratio of the concentration of the two products as a function of time. The concentration of substrate used was 15 μM. Similar results were obtained at 150 μM substrate (data not shown).

Table 1: Second Order Rate Constant (k_{cat}/K_M) for Substrate Hydrolysis by Cathepsin X

Substrate [a]	k_{cat}/K_M (x 10^3 $M^{-1}s^{-1}$)	
P4-P3-P2-P1-↓-P1′-P2′	cathepsin X	Cathepsin B
Abz-F--R-↓-F*-F	2.15	4,170
Abz-F--R--F*↓-F	5.60	- [b]
Abz-F--R-↓-F* [c]	123.2	20.7
Abz-R--R-↓-F*	14.6	1.74
Abz-R--F-↓-F*	3.63	5.72
Abz-A--A-↓-F* [c]	12.45	6.18

[a] Position of hydrolysis indicated by the arrow ↓. F* is used to represent F($4NO_2$). [b] On the basis of the rates obtained with other substrates presented in this table, the hydrolysis of Abz-FRF*F by cathepsin B is considered to occur exclusively through the dipeptidyl carboxypeptidase pathway. [c] From ref *7*. For the hydrolysis of Abz-FR↓F* by cathepsin X, the k_{cat}/K_M value obtained from multiple nonlinear regression analysis (this work) is identical to the value obtained in the previous study (*7*).

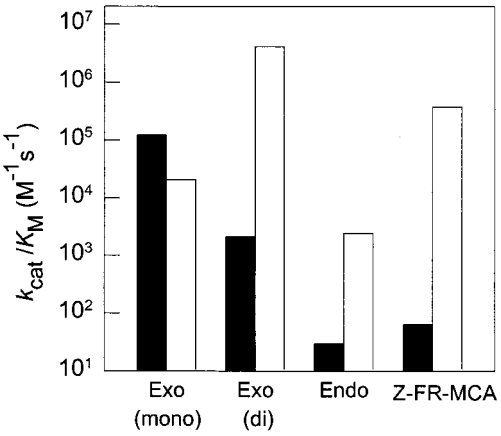


FIGURE 4: Positional specificity profile for cathepsins X (black bars) and cathepsin B (white bars). Exo (mono), carboxypeptidase substrate Abz-FRF($4NO_2$); Exo (di), dipeptidyl carboxypeptidase substrate Abz-FRF($4NO_2$)F; Endo, endopeptidase substrate Abz-AFRSAAQ-EDDnp; Z-FR-MCA, Cbz-FR-MCA.

($k_{cat}/K_M = 2150\ M^{-1}\ s^{-1}$) to removal of a single C-terminal residue in Abz-FRF($4NO_2$) ($k_{cat}/K_M = 123\ 200\ M^{-1}\ s^{-1}$). It is clear then that the enzyme is a much better carboxypeptidase than dipeptidyl carboxypeptidase (by 57-fold in this case). The observation of a direct cleavage of a dipeptide from Abz-FRF($4NO_2$)F can therefore be attributed to a relatively low dipeptidyl carboxypeptidase activity, which is able to contribute to the overall hydrolysis mainly due to the presence of nonoptimal P_2-P_1 residues in the carboxypeptidase step. The influence of the nature of the P_2-P_1 residues on the carboxypeptidase step has been confirmed by using a series of substrates Abz-P_2-P_1-F($4NO_2$), where residues P_2 and P_1 are varied. The results are presented in Table 1. It can be seen that in this tripeptide series, the pair R-F in P_2-P_1 [the closest analogy to placing R-F($4NO_2$) in P_2-P_1 for the hydrolysis of a single C-terminal residue of Abz-FRF($4NO_2$)F] leads to the weakest hydrolysis by cathepsin X. The k_{cat}/K_M value for hydrolysis of Abz-RFF-($4NO_2$) is 34-fold lower than for hydrolysis of Abz-FRF-($4NO_2$). By comparison, R-R and A-A are better tolerated in P_2-P_1, with $k_{cat}/K_M = 14\ 600\ M^{-1}\ s^{-1}$ and $12\ 450\ M^{-1}\ s^{-1}$ for hydrolysis of Abz-RRF($4NO_2$) and Abz-AAF($4NO_2$), respectively.

The positional specificity (mono-, di-, or endopeptidase activity) for cathepsins X and B are compared in Figure 4. For the exopeptidase activities, cathepsin X preferentially cleaves substrates through a monopeptidyl carboxypeptidase pathway, while cathepsin B displays a preference for the dipeptidyl pathway. It must be noted that the preference for one or the other pathway is about the same for both enzymes, i.e., approximately 2 orders of magnitude. Therefore, the dual carboxypeptidase activity is not a feature unique of cathepsin X, since cathepsin B also displays a relatively low monopeptidase activity in addition to its normal dipeptidase activity. What best differentiates the two enzymes is their endopeptidase activity, and especially the activity against Cbz-FR-MCA (Figure 4). As previously noted (*7*), cathepsin X displays a much stricter exopeptidase activity (relative to endopeptidase activity) than cathepsin B. The relative exo- and endopeptidase activities of cathepsin B are "gated" by the occluding loop, a large insertion loop which occupies the S′ subsites of the enzyme (*16*). We have confirmed using site-directed mutagenesis experiments that the occluding loop is important to provide the structural basis defining the exopeptidase activity (data not shown), and also decreases the endopeptidase activity of cathepsin B (*16*). The ability of the long, flexible occluding loop of cathepsin B to move or adjust upon binding of an extended ligand (*31*, *32*) might explain the greater endopeptidase activity of cathepsin B in comparison with cathepsin X. In the latter enzyme, the short, more rigid mini-loop seems to control more tightly the preference for a carboxypeptidase over an endopeptidase activity.

Molecular Basis Underlying the Carboxypeptidase Specificity of Cathepsins X and B. The available structural data at atomic resolution combined with theoretical calculations can assist in perceiving the structural and energetic factors responsible for the carboxypeptidase specificity of cathepsins X and B. Therefore, the monopeptidyl carboxypeptidase substrate Ac-FR↓F and the dipeptidyl carboxypeptidase substrate Ac-FR↓FF were evaluated comparatively at the binding site of each enzyme. All modeled substrates were readily accommodated in the experimentally determined conformation of the binding sites of cathepsins B and X with the exception of the dipeptidyl carboxypeptidase substrate in the cathepsin X binding site. This is due to the presence of the side chains of the mini-loop residues His23 and Tyr27 which occlude the S_2' subsite of cathepsin X and thus prevent binding of a P_2' substrate residue (*7*, *11*). It has been suggested that the C-terminal P_2' residue may be accommodated by rotation of a single side chain belonging to His23 (*12*, *17*). As shown in Figure 5, our conformational search-driven docking of the Ac-FR↓FF substrate in the binding site of cathepsin X predicts that the side chains of both mini-loop residues His23 and Tyr27 have to undergo a conformational change in order to accommodate the C-terminal Phe(P_2') residue of the substrate. The δNH group of His23 is hydrogen bonded to the C-terminal free carboxylate at the P_2' position. This replaces the two hydrogen bonds formed by the C-terminal free carboxylate at the P_1' position in the monopeptidyl carboxypeptidase substrate Ac-FR↓F with the ϵNH group of His23 and the phenolic hydroxyl of Tyr27 (*7*, *11*) (Figure 5).

The modeled complexes of cathepsin X with the monopeptidyl carboxypeptidase substrate Ac-FR↓F and with the dipeptidyl carboxypeptidase substrate Ac-FR↓FF indicate that there are at least three factors that need to be taken into account in order to comparatively evaluate the energetics of binding of the two substrates (eq 1). First, there are additional intermolecular binding contacts afforded by the C-terminal P_2' substrate residue. Second, a conformational transition is required in the protein in order to accommodate the P_2' residue. Third, a number of rotatable bonds in the P_2' residue are frozen upon complexation. The results of our theoretical evaluation of the relative binding free energies are presented in Table 3. Overall, the binding affinity of the dipeptidyl carboxypeptidase substrate Ac-FR↓FF to cathepsin X was found to be weaker by 2.91 kcal/mol than that of the monopeptidyl carboxypeptidase substrate Ac-FR↓F. In contrast, in the case of the binding of the same two substrates to cathepsin B, the Ac-FR↓FF substrate was predicted to have a binding affinity 2.52 kcal/mol stronger than that of the Ac-FR↓F substrate. Therefore, the theoretical calculations predict opposite carboxypeptidase specificities for the two enzymes, in agreement with the experimental observations. Moreover,

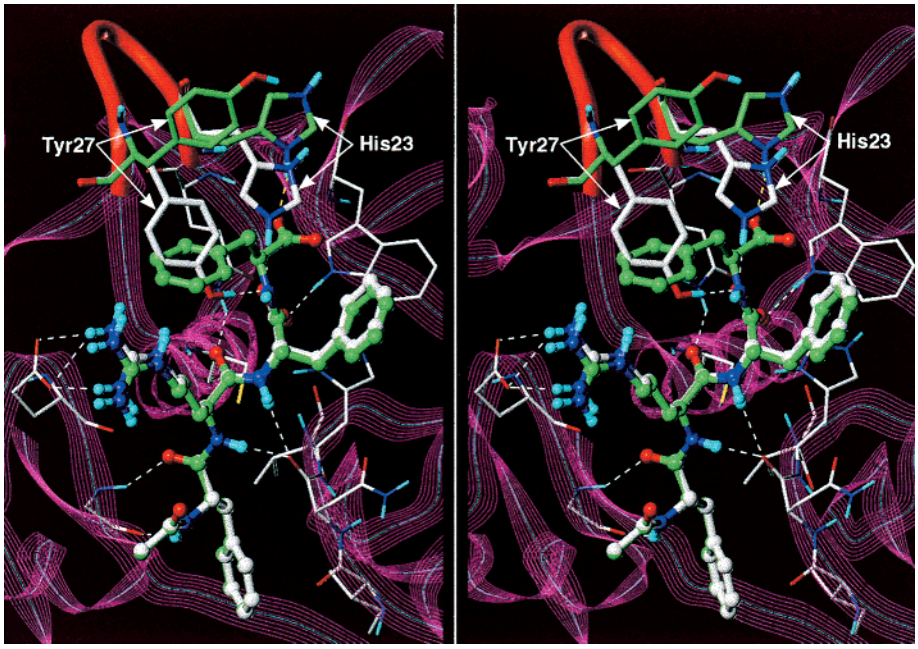


FIGURE 5: Stereoview of the superimposed models of the human cathepsin X in complex with the monopeptidyl carboxypeptidase substrate Ac-FR↓F and with the dipeptidyl carboxypeptidase substrate Ac-FR↓FF. The substrate molecules are shown in ball-and-stick representation with the carbon atoms of Ac-FR↓FF colored in green. Binding of Ac-FR↓FF induces a conformational change in the side chains of the His23 and Tyr27 residues of the enzyme. These two residues are depicted as thick capped-sticks with white carbon atoms in the free, monopeptidyl carboxypeptidase characteristic conformation, and with green carbon atoms in the ligand-induced, dipeptidyl carboxypeptidase characteristic conformation. Selected residues lining the binding site are also shown. Hydrogen bonds between the Ac-FR↓F substrate and enzyme are represented with white dashed lines. The hydrogen bond between the C-terminal free carboxylate of the Ac-FR↓FF substrate and His23 is indicated with a yellow dashed line. The backbone of the mini-loop is represented as a red tube. Nonpolar hydrogen atoms are omitted for clarity.

Table 2: Second Order Rate Constant (k_{inac}/K_i) for Inhibition of Cathepsins X, B, and L by E-64 Analogues

	k_{inac}/K_i (M^{-1} s^{-1})		
inhibitor	cathepsin X	cathepsin B	cathepsin L
*n*PrNH-(2S,3S)-*t*Eps-Ile-OH	225	33	21
CA074	4	136 500	31
E-64	775	55 700	37 700

Table 3: Summary of the Results Obtained from the Calculation of the Relative Free Energy of Binding of Mono- and Dipeptidyl Carboxypeptidase Substrates against Cathepsins X and B[a]

enzyme	cathepsin X	cathepsin B
ΔSIE	**−0.96**	**−5.60**
ΔE_{vdw}	−2.38	−4.74
$\Delta E_{coul} + \Delta\Delta G^{solvation}$	1.42	−0.87
$\Delta SCE_{protein}$	**0.78**	**0.00**
$\Delta E^{intra}_{protein}$	1.36	0.00
$\Delta G^{solvation}_{protein}$	−0.58	0.00
$T\Delta\Delta S_{ligand}$	**3.08**	**3.08**
$\Delta\Delta G_{binding}$	**2.91**	**−2.52**
$\Delta\Delta G^{\ddagger}_{obs}$	**2.40**	**−3.14**

[a] All energies (in kcal/mol) are for the binding of the dipeptidyl carboxypeptidase substrate Ac-FR↓FF relative to those for the binding of the monopeptidyl carboxypeptidase substrate Ac-FR↓F against the corresponding enzyme. The components of ΔSIE and $\Delta SCE_{protein}$ incorporate the scaling factor shown in eqs 2 and 3. $\Delta\Delta G^{solvation}$ includes the electrostatic and nonelectrostatic components of solvation. Experimental values calculated from $\Delta\Delta G^{\ddagger}_{obs} = -RT \ln[(k_{cat}/K_M - AbzFRF(NO_2)F)/(k_{cat}/K_M - AbzFRF(NO_2))]$.

the calculated relative binding free energies agree favorably with the $\Delta\Delta G^{\ddagger}_{obs}$ for transition-state stabilization (i.e. binding) in the hydrolysis of the di- vs monopeptidyl carboxypeptidase substrates, as calculated from the measured k_{cat}/K_M values (Table 3). This suggests that the difference in substrate binding affinity may account for most of the carboxypeptidase specificity of each of the two cathepsins.

The comparative analysis of the energy terms contributing to the overall relative binding affinity allows a deeper insight into the energetic factors that govern the opposite carboxypeptidase specificities of cathepsins X and B (see Table 3). The dipeptidyl carboxypeptidase substrate has more favorable solvated interaction energy with each of the enzymes relative to the monopeptidyl carboxypeptidase substrate (i.e., ΔSIE < 0). However, this energetic gain is predicted to be larger for cathepsin B (−5.60 kcal/mol) than cathepsin X (−0.96 kcal/mol). Although the addition of the Phe(P_2') residue results in significantly stronger van der Waals interactions with each enzyme, these interactions are almost doubled for cathepsin B in comparison with cathepsin X. This is qualitatively expected from the well-defined shape of the S_2' subsite of cathepsin B, which is designed to accept a P_2' residue. In addition, the combined Coulomb electrostatic and solvation terms favor the binding to cathepsin B of the Ac-FR↓FF substrate relative to Ac-FR↓F. The reverse is true for cathepsin X. The reason is that with cathepsin B, the Ac-FR↓FF introduces a salt-bridge interaction not present in the shorter Ac-FR↓F. With cathepsin X, Ac-FR↓F already contains a salt-bridge interaction and extension of this peptide to Ac-FR↓FF results in less well-formed salt-bridge interactions.

There is a further cost to cathepsin X for accommodating Ac-FR↓FF. This is the cost of the induced conformational change in the side chains of His23 and Tyr27. The energetic penalty was calculated to be 0.78 kcal/mol relative to the free enzyme. There is no penalty associated with the binding

of Ac-FR↓F to cathepsin X since there is no significant induced conformational change relative to the free enzyme. Further decomposition of the energetic cost associated with this conformational transition shows that it arises from an increase in the internal energy that is only partially compensated by a better solvation free energy (the polar His23 and Tyr27 side chains become more solvent exposed) for the dipeptidyl carboxypeptidase characteristic conformation. There is no change in the solvated conformational energy associated with the binding of the two substrates to cathepsin B, since the larger, dipeptidyl carboxypeptidase substrate is readily accommodated in the free enzyme conformation. Finally, there is the cost due to the loss of configurational entropy by freezing the rotatable bonds of the additional Phe-(P_2') residue upon substrate binding. This amounts to approximately 3 kcal/mol (Table 3).

Design of a Cathepsin X-Specific Inhibitor. The marked differences between the primed subsites of cathepsin X relative to other members of the papain family of cysteine proteases can be used to design inhibitors specific to cathepsin X. To demonstrate the potential of this approach, we have synthesized an E-64 derivative that should place a single amino acid residue in the S′ subsite of cathepsin X. This inhibitor is based on the observation that CA074, a cathepsin B specific inhibitor, interacts in the primed subsites of cathepsin B where it positions two amino acid residues to interact optimally in subsites S_1' and S_2' of the enzyme (*22*). Since the cathepsin X structure is designed to accommodate a single amino acid residue in the primed subsites, the inhibitor *n*PrNH-(2S,3S)-*t*Eps-Ile-OH was synthesized and tested against cathepsin X. Selectivity for cathepsin X was evaluated by testing the inhibitor against cathepsins B and L. Inhibition of the three enzymes by E-64 and CA074 was also determined for comparison and the results are presented in Table 2. It can be seen that *n*PrNH-(2S,3S)-*t*Eps-Ile-OH is indeed a specific inhibitor of cathepsin X, with approximately 10-fold preference for cathepsin X over cathepsins B or L. However, the potency of the inhibitor is relatively low, with $k_{inac}/K_i = 225\ M^{-1}\ s^{-1}$. By comparison, CA074 inactivates cathepsin B with a second-order rate constant of 136,500 $M^{-1}\ s^{-1}$. This is believed to be due to an inherent poor reactivity of epoxysuccinyl derivatives with cathepsin X, as illustrated by the low inhibitory activity observed for E-64 ($k_{inac}/K_i = 775\ M^{-1}\ s^{-1}$). Work is in progress to investigate this hypothesis. A very interesting finding of the present study is that cathepsin X is virtually not inhibited by CA074, with an estimated k_{inac}/K_i value of $4\ M^{-1}\ s^{-1}$. This result does not agree with the findings of Guncar et al. (*12*) who reported that CA074 readily inhibited both cathepsins B and X. The relative inhibitory activities of *n*PrNH-(2S,3S)-*t*Eps-Ile-OH and CA074 differ by a factor of 56, a value identical to the relative mono- and dipeptidyl carboxypeptidase activities observed for substrate hydrolysis. The results obtained with the inhibitors are therefore consistent with the data for substrate hydrolysis.

Conclusion. With the availability now of several crystal structures of exopeptidases and numerous structures of endopeptidases, a clearer picture of the structural basis of positional specificity has emerged. As pointed out previously (*2*, *19*), all these enzymes have a basic endopeptidase platform composed of a papain-like two-domain protein, with the substrate binding site located in the interdomain cleft. Enzymes designed to act as exopeptidases have additional structural features that (i) obstruct the substrate binding cleft to prevent or diminish the endopeptidase activity and (ii) position groups that can stabilize either the N- or C-terminus of the bound substrate for exopeptidase activity (*2*). The mini-loop in cathepsin X can very efficiently restrict the access of substrates to and beyond the S_2' subsite, and the enzyme is clearly designed to act as a monopeptidyl carboxypeptidase. However, as stated previously (*19*), the existence of a common catalytic platform and a generally broad sequence specificity can explain why the cathepsins (including cathepsin X) do not exhibit "near-absolute" specificity. For example, it was shown that given the proper sequence of residues in P_2-P_1, substrates can be hydrolyzed by cathepsin L through a dipeptidyl carboxypeptidase pathway more efficiently than with cathepsin B (*19*). A similar situation is observed here with cathepsin X, where the positional specificity (mono- vs dipeptidyl carboxypeptidase) can be controlled or "overruled" by the sequence specificity. In this case, significant dipeptidase activity was observed for an Abz-tetrapeptide substrate mainly because the substrate sequence favors binding in the dipeptidase mode (F-R in P_2-P_1) as compared to the monopeptidase mode (R-F in P_2-P_1). It is important to remember that these results were obtained from in vitro experiments using small well-defined synthetic substrates. The contribution of a putative dipeptidyl carboxypeptidase activity for cathepsin X against protein substrates in vivo is difficult to predict, but it is very likely that a strong competition from the monopeptidyl carboxypeptidase pathway will exist. It is also clear that very little if any endopeptidase cleavage should occur, unless the substrate has certain features that allow endoproteolytic cleavage despite the presence of the mini-loop.

The results of this study strongly support the view that cathepsin X is designed to function as a monopeptidyl carboxypeptidase. The fact that a number of cathepsins are designed to act as exopeptidases (i.e., cathepsins B, C, H, and X) provides a clear strategy to develop reagents (substrates and/or inhibitors) that can be used to differentiate these enzymes between themselves and from the endopeptidases. The inhibitor CA074 is one example, and it has been used extensively over the years to confirm the involvement of cathepsin B in specific processes. Contrary to a recent report from Klemencic et al. (*17*), our results indicate that CA074 is a cathepsin B-specific inhibitor, and in particular does not inhibit cathepsin X. It was also indicated that cathepsins X and B display similar specificity profiles, making it difficult to differentiate the two enzymes (*12*, *17*). To address this point, the specificity of a number of existing and recently designed substrates and inhibitors is reported in Figure 6. It can be seen that both CA074 and cystatin C inhibit selectively cathepsin B over cathepsin X. Cystatin C in particular can readily inhibit cathepsin B with K_i in the low nanomolar range, while no inhibition of cathepsin X was detected up to 4 μM of inhibitor (*7*). It must be noted that in their work, Klemencic et al. used cathepsin X purified from human liver (*17*). Therefore, there is a high risk that this preparation might have been contaminated with small amounts of cathepsin B. This would explain the differences in results between the two studies. We have estimated that a cathepsin B contamination of 1% or less would be sufficient to account for the results obtained by Klemencic et al. (*17*),

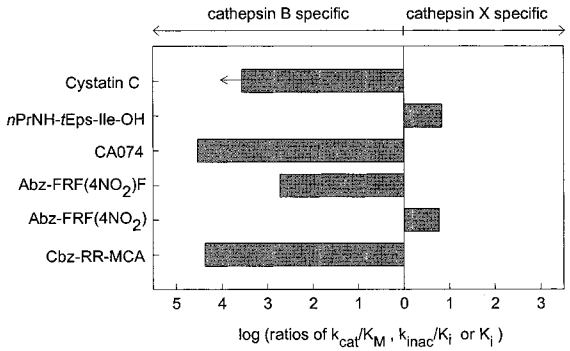


FIGURE 6: Selectivity of various substrates and inhibitors toward cathepsins X and B. The log of the ratios of k_{cat}/K_M values for substrates, k_{inac}/K_i for the irreversible inhibitors CA074 and *n*PrNH-(2S,3S)-*t*Eps-Ile-OH, and K_i for the reversible inhibitor cystatin C are displayed on the side of the graph corresponding to the enzyme to which the compound is selective.

especially considering that the substrate used in their study, Cbz-FR-MCA, shows 6000-fold greater activity against cathepsin B than cathepsin X. In our study, glycosylated recombinant cathepsin X was produced in *Pichia pastoris*, therefore eliminating the risk of contamination by cathepsin B. The glycosylation of cathepsin X does not appear to be a factor in inhibition by cystatin C, since no inhibition was observed after deglycosylation of cathepsin X by endoH (data not shown). A substrate designed to fit the sequence requirements of cathepsin B for dipeptidyl carboxypeptidase activity is hydrolyzed ca. 10^3-fold better by cathepsin B than by cathepsin X. Cbz-RR-MCA represents another prototype of a cathepsin B-specific substrate that combines the relatively poor activity of cathepsin X against MCA substrate with the recognized ability of cathepsin B to hydrolyze substrates containing R-R in P_2-P_1. At the moment, reagents such as the substrate Abz-FRF($4NO_2$) and the inhibitor *n*PrNH-(2S,3S)-*t*Eps-Ile-OH are only mildly selective for cathepsin X over cathepsin B (i.e., about 1 order of magnitude preference for cathepsin X). Nevertheless, from the data presented in Figure 6 it is clear that reagents do exist that can be used to differentiate cathepsin X from cathepsin B.

With the increasing number of cysteine proteases of the papain family, the possibility of functional redundancy is becoming an important issue. This is possibly best exemplified by cathepsins X and B, which are both exopeptidases with a lower level of endopeptidase activity. A clear picture of the specificities of these enzymes is needed, which can help in the assignment of physiological roles and in the design of specific inhibitors. The results of this study confirm our original findings that unlike many of the recently discovered enzymes, cathepsin X exhibits a very strong preference for hydrolyzing substrates through a carboxypeptidase pathway. More importantly, and contrary to recent reports (*12*, *17*), cathepsins X and B can be differentiated through their activity profile against substrates and inhibitors. A number of substrates and inhibitors specific to cathepsin B are known. However, a potent and specific inhibitor of cathepsin X useful for investigating the physiological role of this enzyme, has not been identified yet. Even though our preliminary results indicate that targeting the S_1' subsite of cathepsin X is a valid approach toward selectivity, additional work is required to increase the inhibitory activity.

REFERENCES

1. Bühling, F., Fengler, A., Brandt, W., Welte, T., Ansorge, S., and Nägler, D. K. (2000) *Adv. Exp. Med. Biol. 477*, 241−254.
2. McGrath, M. E. (1999) *Annu. Rev. Biophys. Biomol. Struct. 28*, 181−204.
3. Barret, A. J., Rawlings, N. D., and Woessner, J. F., Jr. (1998) In *Handbook of Proteolytic Enzymes* (Barret, A. J., Rawlings, N. D., and Woessner, J. F., Eds.) Academic Press, London.
4. Brömme, D. (1999) *Drug News Perspect. 12*, 73−82.
5. Gowen, M. (1997) *Exp. Opin. Invest. Drugs 6*, 1199−1202.
6. Storer, A. C., and Ménard, R. (1996) *Perspect. Drug Discovery Des. 6*, 33−46.
7. Nägler, D. K., Zhang, R., Tam, W., Sulea, T., Purisima, E. O., and Ménard, R. (1999) *Biochemistry 38*, 12648−12654.
8. Nägler, D. K., and Ménard, R. (1998) *FEBS Lett. 434*, 135−139.
9. Santamaría, I., Velasco, G., Pendás, A. M., Fueyo, A., and López-Otín, C. (1998) *J. Biol. Chem. 273*, 16816−16823.
10. Deussing, J., von Olshausen, I., and Peters, C. (2000) *Biochim. Biophys. Acta 1491*, 93−106.
11. Sivaraman, J., Nägler, D. K., Zhang, R., Ménard, R., and Cygler, M. (2000) *J. Mol. Biol. 295*, 939−951.
12. Guncar, G., Klemencic, I., Turk, B., Turk, V., Karaoglanovic-Carmona, A., Juliano, L., and Turk, D. (2000) *Struct. Folding Des. 8*, 305−313.
13. Falcone, F. H., Tetteh, K. K. A., Hunt, P., Blaxter, M. L., Loukas, A., and Maizels, R. M. (2000) *Exp. Parasitol. 94*, 201−207.
14. Musil, D., Zucic, D., Turk, D., Engh, R. A., Mayr, I., Huber, R., Popovic, T., Turk, V., Towatari, T., Katunuma, N., and Bode, W. (1991) *EMBO J. 10*, 2321−2330.
15. Barrett, A. J., and Kirschke, H. (1981) *Methods Enzymol. 80*, 535−561.
16. Nägler, D. K., Storer, A. C., Portaro, F. C. V., Carmona, E., Juliano, L., and Ménard, R. (1997) *Biochemistry 36*, 12608−12615.
17. Klemencic, I., Carmona, A. K., Cezari, M. H. S., Juliano, M. A., Juliano, L., Guncar, G., Turk, D., Krizaj, T., Turk, V., and Turk, B (2000) *Eur. J. Biochem. 267*, 5404−5412.
18. Carmona, E., Dufour, E., Plouffe, C., Takebe, S., Mason, P., Mort, J. S., and Ménard, R. (1996) *Biochemistry 35*, 8149−8157.
19. Nägler, D. K., Tam, W., Storer, A. C., Krupa, J. C., Mort, J. S., and Ménard, R. (1999) *Biochemistry 38*, 4868−4874.
20. 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.
21. Gour-Salin, B. J., Lachance, P., Magny, M.-C., Plouffe, C., Ménard, R., and Storer, A. C. (1994) *Biochem. J. 299*, 389−392.
22. Turk, D., Podobnik, M., Popovic, T., Katunuma, N., Bode, W., Huber, R., and Turk, V. (1995) *Biochemistry 34*, 4791−4797.
23. Cornell, W. D., Cieplak, P., Bayly, C. I., Gould, I. R., Merz, K. M., Ferguson, D. M., Spellmeyer, D. C., Fox, T., Caldwell, J. W., and Kollman, P. A. (1995) *J. Am. Chem. Soc. 117*, 5179−5197.
24. Li, Z., and Scheraga, H. A. (1987) *Proc. Natl. Acad. Sci. U.S.A. 84*, 6611−6615.
25. Purisima, E. O. (1998) *J. Comput. Chem. 19*, 1494−1504.
26. Purisima, E. O., and Nilar, S. H. (1995) *J. Comput. Chem. 16*, 681−689.
27. Vorobjev, Y. N., and Hermans, J. (1997) *Biophys. J. 73*, 722−731.
28. Froloff, N., Windemuth, A., and Honig, B. (1997) *Protein Sci. 6*, 1293−1301.
29. Tamai, M., Yokoo, C., Murata, M., Oguma, K., Sota, K., Sato, E., and Kanaoka, Y. (1987) *Chem. Pharm. Bull. 35*, 1098−1104.
30. Mori, K., and Iwasawa, H. (1980) *Tetrahedron 36*, 87−90.
31. Cygler, M., Sivaraman, J., Grochulski, P., Coulombe, R., Storer, A. C., and Mort, J. S. (1996) *Structure 4*, 405−416.
32. Turk, D., Podobnik, M., Kuhelj, R., Dolinar, M., and Turk, V. (1996) *FEBS Lett. 384*, 211−214.

BI002460A