Human Cathepsin X: A Cysteine Protease with Unique Carboxypeptidase Activity†

Dorit K. Nägler,‡ Rulin Zhang, Wendy Tam, Traian Sulea, Enrico O. Purisima, and Robert Ménard*

Biotechnology Research Institute, National Research Council of Canada, 6100 Avenue Royalmount, Montréal, Québec, Canada H4P 2R2

Received June 15, 1999; Revised Manuscript Received August 6, 1999

ABSTRACT: Cathepsin X is a novel cysteine protease which was identified recently from the EST (expressed sequence tags) database. In a homology model of the mature cathepsin X, a unique three residue insertion between the Gln22 of the oxyanion hole and the active site Cys31 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 we have termed the "mini-loop". From the model, it became apparent that this distinctive structural feature might confer exopeptidase activity to the enzyme. To verify this hypothesis, human procathepsin X was expressed in *Pichia pastoris* and converted to mature cathepsin X using small amounts of human cathepsin L. Cathepsin X was found to display excellent carboxypeptidase activity against the substrate Abz-FRF-($4NO_2$), with a k_{cat}/K_M value of 1.23×10^5 M^{-1} s^{-1} at the optimal pH of 5.0. However, the activity of cathepsin X against the substrates Cbz-FR-MCA and Abz-AFRSAAQ-EDDnp was found to be extremely low, with k_{cat}/K_M values lower than 70 M^{-1} s^{-1}. Therefore, cathepsin X displays a stricter exopeptidase activity than cathepsin B. No inhibition of cathepsin X by cystatin C could be detected up to a concentration of 4 μM of inhibitor. From a model of the protease complexed with Cbz-FRF, the bound carboxypeptidase substrate is predicted to establish a number of favorable contacts within the cathepsin X binding site, in particular with residues His23 and Tyr27 from the mini-loop. The presence of the mini-loop restricts the accessibility of cystatin C as well as of the endopeptidase and MCA substrates in the primed subsites of the protease. The marked structural and functional differences of cathepsin X relative to other members of the papain family of cysteine proteases will be of great value in designing specific inhibitors useful as research tools to investigate the physiological and potential pathological roles of this novel enzyme.

Proteases are recognized for their vital roles in the regulation of many biological processes and their participation in a vast array of diseases. Cysteine proteases in particular constitute attractive targets for development of inhibitors as potential therapeutic agents due to their involvement in several pathological conditions. As is the case with many other gene families, the mining of databases from genome-sequencing projects has led to the discovery of several new cysteine proteases of the papain family; e.g., cathepsin K (*1*, *2*), lymphopain (*3*, *4*), cathepsin V (*5*, *6*), cathepsin X (*7*, *8*), and cathepsin F (*9*, *10*) have all been cloned in the past few years. Even though the increasing number of cysteine proteases could possibly lead to functional redundancy, differences in cellular distribution and intracellular localization have been shown to contribute to defining specific functional roles for certain members of this family of proteases [e.g., cathepsin S (*11*) and cathepsin K (*1*, *2*, *12*)]. Although restricted expression patterns might 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, with the nature of the residue at position P_2 of a substrate being in many cases the most significant in terms of determining specificity for cysteine proteases of the papain family (*13*). In addition, it appears that similarly to sequence specificity, the positional specificity of these proteases (endo- or exopeptidase) can also be relatively broad (*14*). This is due to the fact that cysteine proteases of the papain family probably all evolved from the same basic endopeptidase structural platform (*15*). Enzymes able to function "by design" as exopeptidases also consist of a papain-like two-domain endopeptidase platform, with additional structural elements that confer their exopeptidase capability. The best characterized example is cathepsin B, which has an insert that forms a surface loop partially occluding the binding cleft beyond the S_2'subsite (*16*). This loop, which contains two histidine residues able to interact with the C-terminus of a substrate, is the major structural determinant of the exopeptidase specificity and relatively low endopeptidase activity of cathepsin B (*16*−*18*).

Cathepsin X is a novel cysteine protease which was identified from the EST (expressed sequence tags) database and was obtained by PCR amplification from a human ovary cDNA library (*7*). It is clear, based on amino acid sequence alignment of the mature part of the enzyme, that cathepsin X belongs to the papain family of cysteine proteases. However, unlike most of the recently discovered members

† NRCC Publication No. 42924.

* To whom correspondence should be addressed at the Biotechnology Research Institute, National Research Council of Canada, 6100 Royalmount Ave., Montreal, Quebec, Canada H4P 2R2. Telephone: 514-496-6317. Fax: 514-496-5143. E-mail: robert.menard@nrc.ca.

‡ Present address: Institute of Pathology, Otto-von-Guericke University Magdeburg, Leipziger Str.44, 39120 Magdeburg, Germany.

10.1021/bi991371z CCC: $18.00 Published 1999 by the American Chemical Society
Published on Web 09/09/1999

of this family which are closely related to cathepsin L, the primary structure of cathepsin X contains several unique features that clearly distinguish it from other human cysteine proteases (*7*). The predicted proregion is very short (38 amino acid residues) compared to all known cathepsins, where the proregions vary in length from 62 residues in cathepsin B to around 100 residues in cathepsin L-like enzymes (*19, 20*) and 251 residues in cathepsin F (*10*). In addition, the proregion of cathepsin X does not share a significant degree of similarity with that of other cathepsins. Another distinctive feature of cathepsin X is a three amino acid residue insertion before the consensus CGXC (Cys28-Gly29-Ser30-Cys31) sequence. Such an insertion in a highly conserved region between the glutamine (Gln22) of the putative oxyanion hole and the active site cysteine (Cys31) is highly unusual for a papain-like cysteine protease and might confer special properties to the enzyme. These features suggest that cathepsin X may belong to a new subfamily within the papain family of cysteine proteases (*7, 8*).

The large database of structural and functional information available for the papain family of cysteine proteases can be of tremendous value for rapid assessment of the putative functional properties of novel cysteine proteases. In this paper, we have used molecular modeling to gain insights into the possible functional consequences of the different features observed in the primary sequence of the enzyme, and to assist in the development of substrates for this protease. Kinetic characterization and pH-dependency of hydrolysis of the designed substrates have been used to describe the specificity profile of this novel enzyme. The results indicate that cathepsin X exhibits unique properties in terms of both its substrate specificity and its interaction with a protein inhibitor of cysteine proteases. The combination of functional characterization with structural model building has enabled us to propose the fine molecular details defining the specificity of cathepsin X. The relevancy of this information for inhibitor design is discussed.

EXPERIMENTAL PROCEDURES

Materials. Restriction endonucleases, T4 DNA ligase, and T4 DNA polymerase were purchased from New England Biolabs (Mississauga, Ontario). Expand DNA polymerase was from Boehringer Mannheim (Laval, Québec). The vector (pPIC9) and *Pichia pastoris* strain GS115 were purchased from Invitrogen Corp. (San Diego, CA). The substrate Cbz-FR-MCA[1] and the irreversible inhibitor E-64 were purchased from IAF Biochem International Inc. (Laval, Québec). The substrates Cbz-AA-MCA, Abz-FRF($4NO_2$), Abz-FRF($4NO_2$)A, and Abz-AAF($4NO_2$) were from Enzyme Systems Products (Livermore, CA), and Abz-AFRSAAQ-EDDnp was obtained from Luiz Juliano (Escola Paulista de Medicina, São Paolo, Brazil). Human cathepsins B and L were expressed and purified as described previously (*18, 21*). Human cystatin C was prepared as described earlier (*22*).

Expression, Purification, and Activation of Recombinant Human Cathepsin X. The human cathepsin X cDNA was amplified by PCR from a human ovary cDNA library (CLONTECH Laboratories, Inc., Palo Alto, CA) using Expand DNA polymerase and the following primers: 5′-CCGCTCGAGAAAAGAGAGGCTGAAGCTGGCGCGG-CGCAGGGCGGCC-3′ and 5′-TTAAACGATGGGGTC-CCCAAATGTACAG-3′. The obtained 900 bp fragment was cloned into the vector pPIC9, sequenced, and expressed in the yeast *Pichia pastoris* as a prepro-α-factor fusion construct using the culture conditions recommended by Invitrogen. Procathepsin X was expressed and secreted at levels of approximately 5 mg/L of initial culture medium. Yeast cells in suspension culture (1 L) were centrifuged at 3000*g* for 10 min, and the supernatant was concentrated 10-fold using an Amicon YM10 membrane in a stirred cell. The concentrated solution was diafiltered against 50 mM sodium acetate, pH 5.0. The sample was loaded onto a CM-Sephadex column (2.5 × 6.0 cm) equilibrated with 50 mM sodium acetate buffer, pH 5.0. Procathepsin X was eluted with 0.6 M NaCl in the same buffer and used for processing trials. Production of mature cathepsin X from the proenzyme was achieved by incubating procathepsin X (3−4 μM in 50 mM sodium citrate, pH 5.5, containing 0.2 M NaCl, 1 mM EDTA, and 2 mM DTT) at 22 °C with 25 nM human cathepsin L. The conversion to mature enzyme was verified by N-terminal sequence analysis on a model 470A gas-phase sequencer equipped with an on-line model 120A phenylthiohydantoin analyzer (Applied Biosystems Inc.). The solution containing processed cathepsin X was dialyzed against 50 mM potassium phosphate buffer, pH 7.0. The sample was loaded onto a DEAE column (2.5 × 4.0 cm) equilibrated with 50 mM potassium phosphate buffer, pH 7.0, and cathepsin X was eluted with 0.7 M NaCl. Purified cathepsin X was stored at 4 °C in the elution buffer containing 100 μM MMTS.

Kinetic Measurements. Kinetic experiments were performed as previously described (*18, 23*). All kinetic measurements were carried out at 25 °C in the presence of 2 mM DTT, 0.2 M NaCl, and 3% DMSO. The reaction mixtures also contained 50 mM sodium citrate (pH 3.0−5.9), 1 mM EDTA; or 50 mM sodium phosphate (pH 5.8−7.9), 5 mM EDTA; or 50 mM sodium borate (pH 8.0−10.0), 5 mM EDTA. The concentration of active enzyme was determined by titration using E-64 (*24*). 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. Cleavage products resulting from hydrolysis of the IQF substrates were identified as described previously (*14, 18*). Inhibition studies with cystatin C have been carried out by measuring the steady-state rate of hydrolysis of the substrate Abz-FRF($4NO_2$) in the presence of inhibitor, as described previously for inhibition of cathepsin L by its propeptide (*21*).

Molecular Modeling. The homology model of mature cathepsin X (residues Leu1−Pro240) was constructed using the COMPOSER and PROTEIN LOOPS modules in SYBYL 6.4 (Tripos, Inc., St. Louis, MO). Crystal structures of cathepsins B, L, and K, papain, actinidin, and protease omega (PDB codes 1huc, 1cjl, 1mem, 1ppn, 1aec, and 1ppo,

[1] Abbreviations: Abz-FRF($4NO_2$), *o*-aminobenzoyl-L-phenylalanyl-L-arginyl-4-nitro-L-phenylalanine; Abz-AAF($4NO_2$), *o*-aminobenzoyl-L-alanyl-L-alanyl-4-nitro-L-phenylalanine; Abz-FRF($4NO_2$)A, *o*-aminobenzoyl-L-phenylalanyl-L-arginyl-4-nitro-L-phenylalanyl-L-alanine; Cbz-FR-MCA, carbobenzoxy-L-phenylalanyl-L-arginine 4-methylcoumarinyl-7-amide; Cbz-AA-MCA, carbobenzoxy-L-alanyl-L-alanine 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; MMTS, methyl methanethiolsulfonate; IQF substrate, fluorogenic substrate with intramolecularly quenched fluorescence.

respectively) were used as templates. Structural refinements were carried out in SYBYL 6.4 by energy minimization to an RMS gradient of 0.05 kcal/(mol·Å) using the AMBER all-atom force-field (*25*) with the Powel minimizer, a distance-dependent (4r) dielectric constant, and an 8 Å nonbonded cutoff. The structure was refined in three stages: (i) the loop regions only; (ii) including all side chains; and (iii) allowing all atoms to relax. A conformational search was carried out for the loop Gln22-His23-Ile24-Pro25-Gln26-Tyr27-Cys28 using a Monte Carlo with energy minimization procedure (*26*). Starting conformations for minimization were generated by randomly perturbing one or more dihedral angles in the loop. The perturbations involved random changes in the side-chain dihedral angles of His23, Ile24, Gln26, and Tyr27 as well as crankshaft rotations of peptide units Gln22-His23, His23-Ile24, Ile24-Pro25, Pro25-Gln26, Gln26-Tyr27, and Tyr27-Cys28. The Cys28−Gly29 amide bond and the Cys28−Cys71 disulfide bond were broken prior to backbone dihedral angle rotations and re-formed for energy minimization. A total of 800 Monte Carlo minimization steps were carried out in which all residues in the loop as well as those within 6 Å were allowed to move. The lowest-energy structure obtained was 14.4 kcal/mol lower than the starting conformation. This structure was used to dock the substrate Cbz-FRF in the putative binding site of cathepsin X. The substrate binding mode in the unprimed (for Cbz-FR part) and primed (for C-terminal Phe) enzyme binding sites was inferred from the crystal structures of the papain−Cbz-FA-CMK (PDB code 6pad) and cathepsin B−CA030 (PDB code 1csb) complexes, respectively. The complex was modeled in the thiohemiketal oxyanion state formed by the attack of the Arg backbone carbonyl of the substrate by the active site Cys31 of the protease. Partial atomic charges of the oxyanion species were obtained through ESP-fit calculation on a model compound. The substrate was energy-minimized in the binding site while keeping the protein residues fixed, followed by energy-minimization in which the substrate, the Gln22−Cys28 loop, and all residues within 6 Å were relaxed up to a root-mean-square gradient of 0.01 kcal/(mol·Å).

RESULTS AND DISCUSSION

Production of the Mature Form of Cathepsin X. Human procathepsin X was expressed in *Pichia pastoris* at a level of approximately 5 mg/L of culture medium. Attempts to produce mature cathepsin X by autocatalytic processing of the proenzyme were unsuccessful. Procathepsin X did not autoprocess at low pH (4.0 or 5.0), on addition of DTT (2−5 mM), and on increasing the temperature (from room temperature to 30 or 60 °C). This is contrary to what is normally observed for most cysteine proteases of the papain family, where processing has been shown to occur autocatalytically under acidic conditions (*27*−*29*). A small level of processing was observed with pepsin; however, the most efficient conversion of procathepsin X to the mature form was obtained using small amounts of human cathepsin L, as described under Experimental Procedures. The cathepsin L cleavage site was found to be four residues upstream of the predicted N-terminus for mature cathepsin X (*7*), and processing of the 43 kDa glycosylated precursor by cathepsin L yielded a mature enzyme form of approximately 39 kDa by SDS−PAGE analysis. Upon deglycosylation with endoH, these bands appear at approximately 36 and 32 kDa, respectively. It must be noted that several cysteine proteases have been found to contain a short N-terminal extension upon in vitro activation, with no measurable consequence on the activity of the enzyme (*30*−*33*). The observation that procathepsin X does not readily autoprocess in vitro, coupled to the fact that the proregion is by far the shortest one for this group of enzymes and does not share any significant similarity with other cathepsins, suggests that the structural basis for regulation of activity and processing of the zymogen might be different from that of other members of the papain superfamily (*34*).

Modeling of the Cathepsin X Structure and Prediction of Substrate Specificity. The identification of distinctive features in the cathepsin X primary sequence (*7*, *8*) prompted us to model the three-dimensional structure of the mature part of the protease. The structural alignment of cathepsin X onto the six template structures (not shown) corrected the previously reported primary sequence-based alignments (*7*, *8*). Among the changes, the new alignment does not predict a 15 residue insertion to be aligned with the occluding loop sequence in cathepsin B (*7*) nor does it contain the 14 residue insertion in the C-terminal end of the molecule (*8*) which was realigned thus allowing formation of the Cys153−Cys235 disulfide. Despite the differences, the three residue insertion between the Gln22 of the putative oxyanion hole and the active site Cys31, highlighted in the previous reports (*7*, *8*), is preserved in the structural alignment. This insertion is located in a loop corresponding to residues His23−Tyr27, which we term the "mini-loop". After the preliminary model building, it became apparent that only this unique insertion may directly affect the substrate specificity, whereas the many other differences are located in exposed loops distant from the active center. Conformational search in the Gln22−Cys28 sequence resulted in the His23 residue facing the Trp202 in a position that partially occludes the S_2' subsite. The superposition of the cathepsin X model onto the cathepsin B crystal structure was striking in that the His23 of cathepsin X occupied a region in space which partially overlapped with His110 of cathepsin B, a residue which is considered to be responsible for the exopeptidase activity of the latter enzyme (*16*). Complexation of cathepsin X with the substrate Cbz-FRF slightly refined the position of the histidine to optimally interact with the free carboxylate of the substrate P_1' residue. The superposition of the modeled structure of mature cathepsin X (from the complex with the Cbz-FRF substrate) onto the X-ray structure of cathepsin B is given in Figure 1. This finding suggests that cathepsin X may possess exopeptidase activity and motivated the choice of substrates for kinetic studies on substrate specificity.

Functional Characterization of Cathepsin X. To investigate the proteolytic activity and specificity of cathepsin X, substrates designed to interact in various subsites of the enzyme were prepared. To monitor the putative carboxypeptidase activity of cathepsin X, the IQF substrate Abz-FRF-($4NO_2$) was used. This substrate interacts in subsites S_3 to S_1' of the cysteine protease. For endopeptidase activity, the extended IQF substrate Abz-AFRSAAQ-EDDnp was tested. The activity of cathepsin X against Cbz-FR-MCA, the most widely used substrate for cysteine proteases of the papain family, was also determined. For these substrates, the sequence of amino acids was chosen in order to position

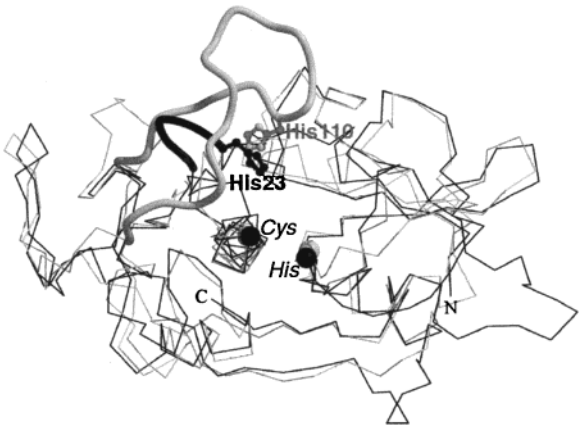

FIGURE 1: Superposition of the modeled structure of mature human cathepsin X and the crystal structure of mature human cathepsin B (PDB code 1huc). The molecules are depicted as a Cα trace shown in black for cathepsin X and in gray for cathepsin B. The mini-loop of cathepsin X (His23−Tyr27) and the occluding loop of cathepsin B (Ser104−Thr125) are represented as tubes. Cα atoms of the catalytic Cys and His residues are indicated as CPK. The side chains of the His23 of cathepsin X and His110 of cathepsin B are shown in ball-and-stick representation.

Table 1: Second-Order Rate Constants k_{cat}/K_M for Substrate Hydrolysis by Human Cathepsin X at pH 5.0

substrate	k_{cat}/K_M ($\times 10^3$ M^{-1} s^{-1})
Abz-F-R-F($4NO_2$)	123.2 ± 3.9
Cbz-F-R-MCA	0.064 ± 0.011
Abz-A-F-R-S-A-A-Q-EDDnp	0.030 ± 0.009
Abz-A-A-F($4NO_2$)	12.45 ± 0.78

phenylalanyl in P_2, corresponding to the preferred residue for interaction in the S_2 subsite of most cysteine proteases, and arginyl in P_1, a well-accepted residue to interact in S_1 and commonly found in many cysteine protease substrates (*14*). The results are presented in Table 1. It can be seen that cathepsin X displays excellent carboxypeptidase activity against Abz-FRF($4NO_2$), with a k_{cat}/K_M value of 1.23×10^5 M^{-1} s^{-1} at pH 5.0. With the substrate Abz-FRF($4NO_2$)A, normally used to characterize the dipeptidyl carboxypeptidase activity of cathepsin B (*14*), a lag phase was observed in the fluorescence vs time progress curve, suggesting that hydrolysis might occur through a sequential carboxypeptidase pathway with an initial cleavage at the F($4NO_2$)−A bond followed by cleavage of the R−F($4NO_2$) bond with increase in fluorescence. This was confirmed by HPLC analysis (data not shown). Surprisingly, the activity of cathepsin X against the substrate Cbz-FR-MCA was found to be extremely low, with a k_{cat}/K_M value of only 64 M^{-1} s^{-1}. Most human cysteine proteases of the papain family have relatively high activity against the unprimed subsite-directed substrate Cbz-FR-MCA, with k_{cat}/K_M values ranging from approximately 1.2×10^5 M^{-1} s^{-1} for cathepsin S up to values close to 10^7 M^{-1} s^{-1} with cathepsins L and F (*9*, *14*). With this substrate, only the MCA portion interacts in the S′ subsites (mainly subsite S_1'). Santamaria et al. (*8*) reported previously that cathepsin X is able to degrade Cbz-FR-MCA. It must be noted, however, that the cathepsin X used in the work of Santamaria et al. consists of a fusion protein with glutathione *S*-transferase, and the kinetic characterization is only qualitative. The activity of cathepsin X against the extended endopeptidase substrate Abz-AFRSAAQ-EDDnp was also found to be very low, with $k_{cat}/K_M = 30$ M^{-1} s^{-1} (Table 1).

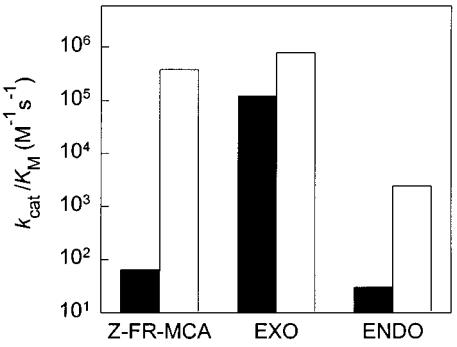

FIGURE 2: Positional specificity profile for cathepsin X (black bars) and cathepsin B (white bars). Z-FR-MCA, Cbz-FR-MCA substrate; EXO, carboxypeptidase substrate Abz-FRF($4NO_2$) for cathepsin X, and dipeptidyl carboxypeptidase substrate Abz-FRF($4NO_2$)A for cathepsin B; ENDO, endopeptidase substrate Abz-AFRSAAQ-EDDnp.

The positional specificity profiles of cathepsins X and B are compared in Figure 2. Cathepsins X and B both possess low endopeptidase activity against the extended IQF substrate, in comparison with their exopeptidase activities against Abz-AAF($4NO_2$) and Abz-AAF($4NO_2$)A, respectively. The difference, however, is bigger for cathepsin X (4100-fold) than for cathepsin B (330-fold). The most striking difference is observed with the substrate Cbz-FR-MCA, which is as good as the exopeptidase substrate with cathepsin B, but is hydrolyzed 2000-fold slower than Abz-FRF($4NO_2$) by cathepsin X. It is therefore clear that cathepsin X displays a much stricter exopeptidase activity than cathepsin B.

Substrates were also prepared where the P_2-P_1 residues (FR) are replaced by alanines, to evaluate the contribution of these residues to exopeptidase activity. On going from Abz-FRF($4NO_2$) to Abz-AAF($4NO_2$), k_{cat}/K_M decreased by only 10-fold with cathepsin X. This is similar to what was observed for the dipeptidyl carboxypeptidase activity of cathepsin B, where replacement of FR by AA in positions P_2-P_1 of the substrate lead to only 17-fold lower exopeptidase activity (*14*). By comparison, the activity of cathepsin L, cathepsin K, and papain against the substrate Cbz-AA-MCA is lower than with Cbz-FR-MCA by 3−5 orders of magnitude (*14*). As expected, Cbz-AA-MCA is a very poor substrate of cathepsin X, with a k_{cat}/K_M barely measurable and estimated to be lower than 2 M^{-1} s^{-1} (data not shown). Therefore, although only two substrates were tested in the present study, the carboxypeptidase activity of cathepsin X seems to be relatively independent of the nature of the residues interacting in subsites S_1 and S_2. The nature of the residue at position P_2 of the carboxypeptidase substrate does not seem to be a determining factor in the specificity of cathepsin X.

To understand the substrate specificity of cathepsin X from a structural viewpoint, we modeled the complex of the protease with the substrate Cbz-FRF which closely mimics the tested IQF substrate Abz-FRF($4NO_2$). From this simulation, the bound substrate is predicted to establish a number of favorable contacts within the cathepsin X binding site, and some of these interactions are shown in Figure 3. The fact that the C-terminal free carboxylate of the modeled substrate is engaged in hydrogen bond interactions with the imidazole ϵNH group of the protonated His23, the phenolic hydroxyl of Tyr27, and the indole NH group of Trp202

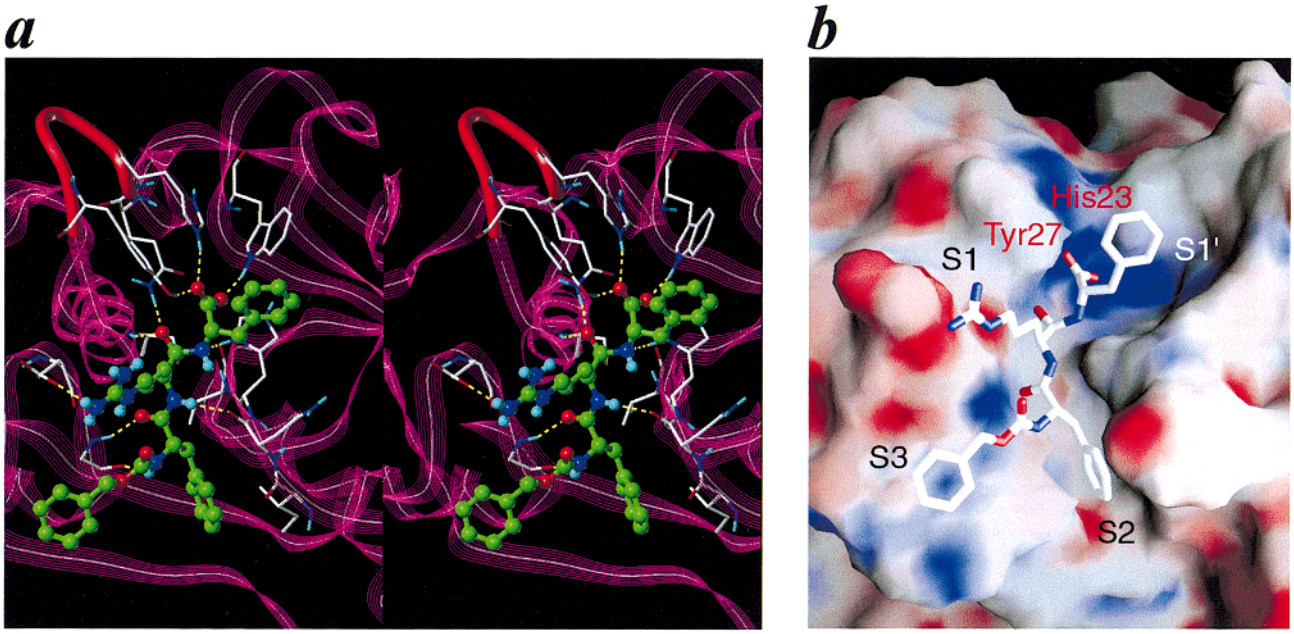

FIGURE 3: Model of the Cbz-FRF substrate in the binding site of cathepsin X. (a) Stereoview showing the intermolecular interactions between the substrate (ball-and-stick with green carbons) and selected binding site residues (capped sticks). Hydrogen bonds are shown as yellow dashed lines. The mini-loop of the enzyme (His23−Tyr27) is highlighted as a red tube. The C-terminal carboxylate of the substrate makes H-bond interactions with residues His23, Tyr27, and Trp202 of the enzyme. (b) Electrostatic and shape complementarity between the substrate (capped sticks) and the enzyme binding site (molecular surface). The electrostatic potential of the uncomplexed enzyme is mapped between −15.0 kT (deep red) and +15.0 kT (deep blue) levels.

(Figure 3a) in the primed region of the binding groove offers a valid explanation for the carboxypeptidase activity of the enzyme. The interactions with His23 and Tyr27 from the mini-loop cannot be formed in any other member of the papain family of cysteine proteases. The spatial localization of these groups leads to the formation of a positively charged protein surface which accommodates the negatively charged carboxylate (Figure 3b). This is in agreement with the demonstrated excellent monopeptidyl exopeptidase activity of cathepsin X. In addition, it appears that His23 and Tyr27 form a "wall" which occludes the S_2' subsite and prevents binding of an extra C-terminal residue of the substrate. Indeed, attempts to manually build the Cbz-FRFA and Cbz-FR-MCA substrates in the modeled binding site resulted in serious collisions (not shown) of the P_2' Ala and MCA moieties with the protein, mostly due to the steric restrictions in the S_2' region. This may explain the poor activity of cathepsin X against the endopeptidase substrate Abz-AFR-SAAQ-EDDnp, and against the MCA substrate. Even though only one extended substrate has been used in this study, the model of cathepsin X indicates that this enzyme is unlikely to function as an endopeptidase. A significant amount of additional work will be necessary to obtain a more intimate understanding of the subsite preferences and potential endopeptidase activity of cathepsin X.

pH-Dependency of Substrate Hydrolysis by Cathepsin X. Figure 4 illustrates the pH-dependency of activity (k_{cat}/K_M) for hydrolysis of the substrates Abz-FRF(4NO$_2$) and Cbz-FR-MCA by cathepsin X. For the carboxypeptidase activity of cathepsin X against the substrate Abz-FRF(4NO$_2$), maximal activity is observed at pH 5.0. Due to difficulty in fitting pH profiles containing very narrow peaks of activity (*35*), reliable kinetic constants cannot be obtained from nonlinear regression analysis of the data. However, it can be seen that the activity decreases at higher pH with a

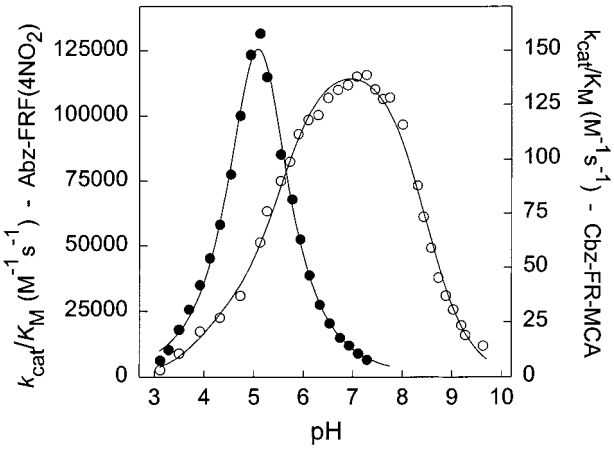

FIGURE 4: pH dependence of k_{cat}/K_M for hydrolysis of the substrates Abz-FRF(4NO$_2$) (filled circles) and Cbz-FR-MCA (open circles) by cathepsin X.

midpoint value near pH 5.5. This is consistent with a contribution from the protonated His23 to the carboxypeptidase activity of cathepsin X. In addition, the exopeptidase substrate possesses a C-terminal group which is expected to ionize in the pH range where the enzyme is active. Deprotonation of this residue possibly contributes to the increase in activity observed in the acid limb of the profile. For the substrate Cbz-FR-MCA, a bell-shaped curve typical of many cysteine proteases of the papain family is observed, with maximal activity near pH 7. The data can be fitted to an equation derived from a model where three ionizable groups modulate activity (*36*). In the acid limb, activity increases with a pK_a of 3.9, with a limiting k_{cat}/K_M value of 34 M^{-1} s^{-1}. Deprotonation of a second residue with a pK_a of 5.6 leads to an enzyme with a limiting k_{cat}/K_M value of 145 M^{-1} s^{-1}. At higher pH values, the activity decreases with a pK_a of 8.5, most likely reflecting the ionization of the active site

Table 2: Inhibition of Human Cathepsins X, B, and L by Cystatin C at pH 5.0

enzyme	K_i (nM)
cathepsin X	>4000
cathepsin B	1.1 ± 0.1
cathepsin L	0.0012 ± 0.0002

His180. Further experiments using site-directed mutagenesis are necessary to confirm the nature of the residues that modulate the proteolytic activity of cathepsin X.

Inhibition of Cathepsin X by Cystatin C. To investigate further the accessibility of the active site of cathepsin X, the inhibitory activity of cystatin C against cathepsin X was monitored (Table 2). Cystatin C is a very potent inhibitor of cathepsin L, with $K_i = 1.2$ pM under the conditions of our assay. With cathepsin B, the inhibitory activity is lower ($K_i = 1.1$ nM) due to the presence of the occluding loop in this enzyme (*17*). With cathepsin X, no inhibition by cystatin C could be detected up to a concentration of 4 μM inhibitor. The dissociation constant of a cathepsin X−cystatin C complex must therefore be at least (3×10^6)-fold higher than for cathepsin L. The tripartite inhibitory wedge of cystatins interacts with the primed subsites of cysteine proteases of the papain family mainly via hydrophobic contacts (*37*, *38*). Our model shows that the S_2' subsite of cathepsin X is much more restricted as compared with the other family members. The lower inhibitory activity of cystatin C against cathepsin X may thus be explained by a steric (and electrostatic) misfit upon binding, which in turn has to induce high-cost conformational rearrangements in both interacting molecules.

Conclusion. The results of this study demonstrate that cathepsin X has a unique carboxypeptidase activity and can be considered to belong to a new subfamily within the papain family of cysteine proteases. Unlike many of the recently discovered enzymes which display similar specificity patterns, cathepsin X exhibits a very strong preference for hydrolyzing substrates through a carboxypeptidase pathway. The model of the cathepsin X structure provides a molecular basis for its unique substrate specificity. According to this model, cathepsin X consists of a papain-like two-domain catalytic platform, similar to other exopeptidases such as cathepsin B, cathepsin H, and bleomycin hydrolase, for which 3-D structures are available (*16*, *39*, *40*). Within this shared framework, functional diversity is generated by the addition of structural elements that confer unique specificities. In cathepsin X, the three residue insertion including His23 provides an anchor for the C-terminus of a substrate, thereby allowing favorable enzyme−substrate interaction independently of the P_2-P_1 sequence. This relatively small modification to the endopeptidase platform results in a tremendous effect on the specificity of the enzyme. In addition, it is becoming clear that the S_2 subsite specificity, which has been considered for many years to be the hallmark of the papain family of cysteine proteases, is much less important for enzymes specifically designed to act as exopeptidases, e.g., cathepsins B and X (*14*). The physiological role of cathepsin X and potential involvement in diseases are unknown at the moment. The results presented here will certainly be applied in further investigations aimed at defining the role of cathepsin X in both normal and pathological conditions. The substrate specificity as well as the unusual resistance to autocatalytic activation and cystatin C inhibition can be used to discriminate between cathepsin X and other known cysteine proteases of the papain family. In addition, the marked differences between the primed subsites of cathepsin X relative to other members of the papain family of cysteine proteases will be of great value in designing specific inhibitors useful both as lead compounds for drug design and as research tools to investigate the physiological or pathological roles of the enzyme. The novel findings of this work greatly extend our understanding of the concepts underlying the specificity of cysteine proteases and provide insights into possible strategies for specific inhibition of cathepsin X.

ACKNOWLEDGMENT

We thank Irena Ekiel for providing the purified cystatin C sample.

REFERENCES

1. Inaoka, T., Bilbe, G., Ishibashi, O., Tezuka, K., Kumegawa, M., and Kokubo, T. (1995) *Biochem. Biophys. Res. Commun. 206*, 89−96.
2. Brömme, D., Okamoto, K., Wang, B. B., and Biroc, S. (1996) *J. Biol. Chem. 271*, 2126−2132.
3. Linnevers, C., Smeekens, S. P., and Brömme, D. (1997) *FEBS Lett. 405*, 253−259.
4. Brown, J., Matutes, E., Singleton, A., Price, C., Molgaard, H., Buttle, D., and Enver, T. (1998) *Leukemia 12*, 1771−1781.
5. Adachi, W., Kawamoto, S., Ohno, I., Nishida, K., Kinoshita, S., Matsubara, K., and Okubo, K. (1998) *Invest. Ophthalmol. Visual Sci. 39*, 1789−1796.
6. Santamaría, I., Velasco, G., Cazorla, M., Fueyo, A., Campo, E., and López-Otín, C. (1998) *Cancer Res. 58*, 1624−1630.
7. Nägler, D. K., and Ménard, R. (1998) *FEBS Lett. 434*, 135−139.
8. Santamaría, I., Velasco, G., Pendás, A. M., Fueyo, A., and López-Otín, C. (1998) *J. Biol. Chem. 273*, 16816−16823.
9. Wang, B., Shi, G. P., Yao, P. M., Li, Z., Chapman, H. A., and Brömme, D. (1998) *J. Biol. Chem. 273*, 32000−32008.
10. Nägler, D. K., Sulea, T., and Ménard, R. (1999) *Biochem. Biophys. Res. Commun. 257*, 313−318.
11. Riese, R. J., Wolf, P. R., Brömme, D., Natkin, L. R., Villandagos, J. A., Ploegh, H. L., and, Chapman, H. A. (1996) *Immunity 4*, 357−366.
12. Inui, T., Ishibashi, O., Inaoka, T., Origane, Y., Kumegawa, M., Kokubo, T., and Yamamura, T. (1997) *J. Biol. Chem. 272*, 8109−8112.
13. Storer, A. C., and Ménard, R. (1996) *Perspect. Drug Discovery Des. 6*, 33−46.
14. Nägler, D. K., Tam, W., Storer, A. C., Krupa, J. C., Mort, J. S., and Ménard, R. (1999) *Biochemistry 38*, 4868−4874.
15. Turk, D., Guncar, G., Podobnik, M., and Turk, B. (1998) *Biol. Chem. Hoppe-Seyler 379*, 137−147.
16. 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.
17. Illy, C., Quraishi, O., Wang, J., Purisima, E., Vernet, T., and Mort, J. S. (1997) *J. Biol. Chem. 272*, 1197−1202.
18. Nägler, D. K., Storer, A. C., Portaro, F. C. V., Carmona, E., Juliano, L., and Ménard, R. (1997) *Biochemistry 36*, 12608−12615.
19. Karrer, K. M., Peiffer, S. L., and DiThomas, M. E. (1993) *Proc. Natl. Acad. Sci. U.S.A. 90*, 3063−3067.
20. Coulombe, R., Grochulski, P., Sivaraman, J., Ménard, R., Mort, J. S., and Cygler, M. (1996) *EMBO J. 15*, 5492−5503.
21. Carmona, E., Dufour, E., Plouffe, C., Takebe, S., Mason, P., Mort, J. S., and Ménard, R. (1996) *Biochemistry 35*, 8149−8157.
22. Ekiel, I., Abrahamson, M., Fulton, D. B., Lindahl, P., Storer, A. C., Levadoux, W., Lafrance, M., Labelle, S., Pomerleau,

Y., Groleau, D., LeSauteur, L., and Gehring, K. (1997) *J. Mol. Biol. 271*, 266−277.
23. Ménard, R., Khouri, H. E., Plouffe, C., Dupras, R., Ripoll, D., Vernet, T., Tessier, D. C., Laliberté, F., Thomas, D. Y., and Storer, A. C. (1990) *Biochemistry 29*, 6706−6713.
24. 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.
25. Weiner, S. J., Kollman, P. A., Nguyen, D. T., and Case, D. A. (1986) *J. Comput. Chem. 7*, 230−252.
26. Li, Z., and Scheraga, H. A. (1987) *Proc. Natl. Acad. Sci. U.S.A. 84*, 6611−6615.
27. Rowan, A. D., Mason, P., Mach, L., and Mort, J. S. (1992) *J. Biol. Chem. 267*, 15993−15999.
28. Mason, R. W., Gal, S., and Gottesman, M. M. (1987) *Biochem. J. 248*, 449−454.
29. 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.
30. Vernet, T., Tessier, D. C., Richardson, C., Laliberte, F., Khouri, H. E., Bell, A. W., Storer, A. C., and Thomas, D. Y. (1990) *J. Biol. Chem. 265*, 16661−16666.
31. Mach, L., Schwihla, H., Stuwe, K., Rowan, A. D., Mort, J. S., and Glossl, J. (1993) *Biochem. J. 293*, 437−442.
32. Brömme, D., Bonneau, P. R., Lachance, P., Wiederanders, B., Kirschke, H., Peters, C., Thomas, D. Y., Storer, A. C., and Vernet, T. (1993) *J. Biol. Chem. 268*, 4832−4838.
33. 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.
34. Cygler, M., and Mort, J. S. (1997) *Biochimie 79*, 645−652.
35. Dixon, H. B. F. (1973) *Biochem. J. 131*, 149−154.
36. Khouri, H. E., Vernet, T., Ménard, R., Parlati, F., Laflamme, P., Tessier, D. C., Gour-Salin, B., Thomas, D. Y., and Storer, A. C. (1991) *Biochemistry 30*, 8929−8936.
37. Bode, W., Engh, R., Musil, D., Thiele, U., Huber, R., Karshikov, A., Brzin, J., Kos, J., and Turk, V. (1988) *EMBO J. 7*, 2593−2599.
38. Stubbs, M. T., Laber, B., Bode, W., Huber, R., Jerala, R., Lenarčič, B., and Turk, V. (1990) *EMBO J. 9*, 1939−1947.
39. Gunčar, G., Podobnik, M., Pungerčar, J., Štrukelj, B., Turk, V., and Turk, D. (1998) *Structure 6*, 51−61.
40. Joshua-Tor, L., Xu, H. E., Johnston, S. A., and Rees, D. C. (1995) *Science 269*, 945−950.

BI991371Z