Biochemistry **1999,** *38,* 4868−4874

Interdependency of Sequence and Positional Specificities for Cysteine Proteases of the Papain Family[‡]

Dorit K. Nägler,[§] Wendy Tam,[§] Andrew C. Storer,[§] Joanne C. Krupa,[||] John S. Mort,[||] and Robert Ménard*[,§]

Biotechnology Research Institute, National Research Council of Canada, Montreal, Quebec, H4P2R2, Canada, and Joint Diseases Laboratory, Shriners Hospital for Children, 1529 Cedar Avenue, Montreal, Quebec, H3G1A6, Canada

Received November 5, 1998; Revised Manuscript Received February 16, 1999

ABSTRACT: The specificity of cysteine proteases is characterized by the nature of the amino acid sequence recognized by the enzymes (sequence specificity) as well as by the position of the scissile peptide bond (positional specificity, i.e., endopeptidase, aminopeptidase, or carboxypeptidase). In this paper, the interdependency of sequence and positional specificities for selected members of this class of enzymes has been investigated using fluorogenic substrates where both the position of the cleavable peptide bond and the nature of the sequence of residues in P_2−P_1 are varied. The results show that cathepsins K and L and papain, typically considered to act strictly as endopeptidases, can also display dipeptidyl carboxypeptidase activity against the substrate Abz-FRF($4NO_2$)A and dipeptidyl aminopeptidase activity against FR-MCA. In some cases the activity is even equal to or greater than that observed with cathepsin B and DPP-I (dipeptidyl peptidase I), which have been characterized previously as exopeptidases. In contrast, the exopeptidase activities of cathepsins K and L and papain are extremely low when the P_2−P_1 residues are A−A, indicating that, as observed for the normal endopeptidase activity, the exopeptidase activities rely heavily on interactions in subsite S_2 (and possibly S_1). However, cathepsin B and DPP-I are able to hydrolyze substrates through the exopeptidase route even in absence of preferred interactions in subsites S_2 and S_1. This is attributed to the presence in cathepsin B and DPP-I of specific structural elements which serve as an anchor for the C- or N-terminus of a substrate, thereby allowing favorable enzyme−substrate interaction independently of the P_2−P_1 sequence. As a consequence, the nature of the residue at position P_2 of a substrate, which is usually the main factor determining the specificity for cysteine proteases of the papain family, does not have the same contribution for the exopeptidase activities of cathepsin B and DPP-I.

The constantly expanding database of structure−function relationships built from structural and mutational analyses is providing a clearer picture of the fine molecular details defining the specificity of cysteine proteases of the papain family. However, many aspects remain unclear or have not yet been fully explored. In particular, one aspect which has been poorly investigated is the interdependency of binding subsites and the cooperativity of enzyme−ligand interactions. As is the case for many proteases, the specificity of cysteine proteases is determined by the nature of the amino acid sequence recognized by the enzyme (sequence specificity) as well as by the position of the peptide bond being cleaved (positional specificity, i.e., endopeptidase, aminopeptidase, or carboxypeptidase). The sequence specificity of cysteine proteases is in general considered to be relatively broad, 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 (*1*). It is well-established that many cysteine proteases exhibit a strong preference for hydrophobic residues at this position (*2*, *3*). Members of this family of enzymes also display various positional specificities. While enzymes such as papain and cathepsins K, L, and S are usually considered to be strictly endopeptidases, cathepsin B has been shown to have both exo- and endopeptidase activities (*4*, *5*). However, studies using extended IQF substrates have shown that the endopeptidase activity of cathepsin B is much lower than that observed with cathepsin L or papain (*6*, *7*). DPP-I (dipeptidyl peptidase I, also called cathepsin C) and cathepsin H display aminopeptidase activities (*8*, *9*), but can also hydrolyze substrates through endopeptidase activity (*10*, *11*). Therefore, even though members of the papain family of cysteine proteases have been classified as endo- or exopeptidases, it appears that, similarly to sequence specificity, the positional specificity could be relatively broad.

[‡] NRCC Publication Number 41490.

* Address correspondence to this author at the Biotechnology Research Institute, 6100 Royalmount Avenue, Montreal, Canada H4P2R2. E-mail: robert.menard@nrc.ca.

[§] National Research Council of Canada.

[||] Shriners Hospital for Children.

[1] Abbreviations: DPP-I, dipeptidyl carboxypeptidase I; 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, *ortho*-aminobenzoyl-L-alanyl-L-phenylalanyl-L-arginyl-L-seryl-L-alanyl-L-alanyl-L-glutamine *N*-(ethylenediamine)-2,4-dinitrophenyl amide; Abz-FRF($4NO_2$)A, *ortho*-aminobenzoyl-L-phenylalanyl-L-arginyl-4-nitro-L-phenylalanyl-L-alanine; Abz-AAF($4NO_2$)A, *ortho*-aminobenzoyl-L-alanyl-L-alanyl-4-nitro-L-phenylalanyl-L-alanine; FR-MCA, L-phenylalanyl-L-arginine 4-methylcoumarinyl-7-amide; AA-MCA, L-alanyl-L-alanine 4-methylcoumarinyl-7-amide; E-64, 1-[[(L-*trans*-epoxysuccinyl)-L-leucyl]amino]-4-guanidino)butane; IQF substrate, fluorogenic substrate with intramolecularly quenched fluorescence.

10.1021/bi982632s CCC: $18.00

Due to their involvement in several pathological conditions (*12*−*14*), cysteine proteases constitute attractive targets for the development of inhibitors as therapeutic agents. For this family of proteases, due to the existence of a large database of structural and functional information, a rational approach to inhibitor design can be considered. To maximize the chances of successfully designing potent and selective inhibitors, it is essential to have in place a well-defined picture of enzyme−ligand interactions. The substrate binding site of cysteine proteases is characterized by its ability to recognize up to seven residues of a substrate (*15*). Of the seven corresponding subsites, at least five appear to be structurally well-defined (*16*). The consequence of the existence of such an extended binding site for the substrate is that the requirements for interaction in a given subsite can, in the case of a polypeptide substrate, be overruled by interactions of the polypeptide chain with other subsites of the enzyme. This dominance at one site can possibly affect both sequence and positional specificities. In particular, a very important issue that has not yet been adressed is the potential interdependency of sequence and positional specificities. Such considerations are particularly relevant to current strategies for the design of potent and selective cysteine protease inhibitors. For example, inhibitors are being designed to span the extended binding site of cysteine proteases (*17*, *18*). Such inhibitors can utilize the structural features responsible for the positional specificities (e.g., dipeptidyl carboxypeptidase activity of cathepsin B), as well as other interactions with, for example, the primary determinant of sequence specificity, the S_2 subsite.

In this paper, we present specificity profiles for mammalian cathepsins B, K, and L, DPP-I, as well as for the archetypal cysteine protease, papain. To study the interdependency between sequence and positional specificities, we have used fluorogenic substrates where both the position of the cleavable peptide bond and the nature of the sequence of residues in P_2−P_1 are varied. It is shown that each cysteine protease under study can exhibit both endo- and exopeptidase activities, given the particular choice of amino acids in positions P_2 and P_1. However, cathepsin B and DPP-I are able to hydrolyze substrates through the exopeptidase route even in the absence of preferred interactions in subsites S_2 and S_1. This interdependency of sequence and positional specificities in a physiological environment has important implications for inhibitor design.

MATERIALS AND METHODS

Restriction endonucleases and T4 DNA ligase were purchased from New England Biolabs. Taq DNA polymerase was from Promega (Madison, WI). The vector (pPIC9) and *Pichia pastoris* strain GS115 were purchased from Invitrogen Corporation (San Diego, CA). The substrate Cbz-FR-MCA and the irreversible inhibitor E-64 were purchased from IAF Biochem International Inc. (Laval, Quebec). The substrates Cbz-AA-MCA, FR-MCA, and AA-MCA were obtained from Enzyme Systems Products (Livermore, CA). Abz-AFR-SAAQ-EDDnp was obtained from Luiz Juliano (Escola Paulista de Medicina, São Paulo, Brazil). Human cathepsins B and L were prepared as described previously (*6*, *19*). Papain (crystallized suspension in sodium acetate) was purchased from Sigma-Aldrich Canada Ltd. (Oakville, Ontario), further purified, and activated and the active site titrated as described previously (*20*). Bovine DPP-I was from Calbiochem (San Diego, CA). The cDNA for human cathepsin K was amplified by PCR from a human spleen cDNA library (CLONTECH Laboratories, Inc., Palo Alto, CA) using gene-specific primers (5′-GCATCCGCTC-GAGAAAAGAGAGGCTGAAGCTCTGTACCCTGAGGAG-ATACTGGAC-3′ and 5′-CCGGCGGCCGCTCACATCT-TGGGGAAGCTGGCCAG-3′) and Expand DNA polymerase (Boehringer Mannheim, Laval, Québec). The PCR product was cloned into the vector pPIC9 and expressed in the yeast *P. pastoris* as a prepro-α-factor fusion construct using the culture conditions recommended by Invitrogen.

Procathepsin K Activation and Purification. Human procathepsin K was expressed and secreted at levels of approximately 3−5 mg/L of initial culture medium. Yeast cells in suspension culture 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 7.0, and the proenzyme was then converted to the mature form at 30 °C in 50 mM sodium acetate, pH 4.0, containing 5 mM DTT. Activated cathepsin K was purified using a CM-Sepharose (Pharmacia Biotech Inc.) column (2.5 × 4 cm) equilibrated in 50 mM sodium acetate buffer, pH 5.0, and eluted with 0.7 M NaCl. Purified cathepsin K was stored at 4 °C in 25 mM sodium acetate, pH 5.0, containing 0.35 M NaCl and 100 μM reversible inhibitor MMTS (methyl methanethiolsulfonate).

Synthesis and Purification of the Substrates Abz-FRF-(4NO$_2$)A and Abz-AAF(4NO$_2$)A. The IQF substrates were synthesized on an Applied Biosystems 431A solid-phase synthesizer, using standard Fastmoc chemistry. The amino group of Abz and the side chain of amino acids were protected during synthesis of the substrates. Purification was done on a Perkin-Elmer PREP-10 octyl reverse-phase column (1.0 × 25 cm) using a linear gradient from 0% CH3CN (0.1% TFA) to 80% CH3CN (0.01% TFA) at a flow rate of 4.0 mL/min. The expected molecular weight of each peptide was confirmed by mass spectral analysis using a MALDI-TOF spectrometer (LaserMat 2000, Finnigan Mat).

Kinetic Measurements. Kinetic experiments were performed as previously described (*6*). Fluorescence was monitored on a SPEX Fluorolog-2 spectrofluorometer. For MCA substrates, the excitation and emission wavelengths were set at 380 and 440 nm, respectively. Hydrolysis of the IQF substrates resulted in an increase in Abz fluorescence which was monitored at 420 nm using an excitation wavelength of 320 nm (*21*). The assays with the IQF substrates were calibrated as previously described (*6*). The enzymes, stored in inhibited form, were activated by incubation in 50 mM sodium phosphate, 0.2 M NaCl, 5 mM EDTA, 2 mM DTT, pH 6.0. The concentration of active cathepsins B, L, and K was determined by titration with E-64 (*22*) while the concentration of DPP-I was estimated on the basis of specifications from the supplier. 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), and 5 mM EDTA. Initial rates were determined at substrate concentrations much lower than K_M, and k_{cat}/K_M values were obtained by dividing

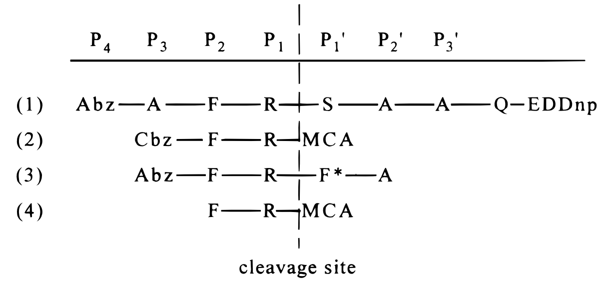

FIGURE 1: Schematic representation of subsite occupancy by endopeptidase and exopeptidase substrates. Substrate nomenclature is as follows: (1) extended endopeptidase substrate; (2) unprimed subsite directed substrate; (3) dipeptidyl carboxypeptidase substrate; (4) dipeptidyl aminopeptidase substrate. Residues P_1 to P_4 interact in the unprimed subsites S_1 to S_4, and residues P_1' to P_3' interact in the primed subsites S_1' to S_3'. The cleavage site is represented by a dashed vertical line. F*: 4-nitro-L-phenylalanyl residue.

the initial rates by enzyme and substrate concentrations. For pH−activity profiles, the data were fit to empirical equations that are not linked to specific models.

Identification of Cleavage Products for IQF Substrates. Substrates (Abz-FRF($4NO_2$)A and Abz-AAF($4NO_2$)A (100 μM) and enzymes were incubated under assay conditions for 15−30 min. The hydrolysis fragments were analyzed by reverse-phase analytical HPLC on a C-18 column (Vydac) with a linear gradient from 0% CH_3CN to 100% CH_3CN, all in 0.1% TFA, in 30 min with a flow rate of 1 mL/min using a Waters HPLC system equipped with both UV and fluorescence detectors. Single fluorescent products with retention times of 15.5 and 12.0 min were detected, corresponding to Abz-FR and Abz-AA, respectively. Identification of the cleavage site with Abz-AFRSAAQ-EDDnp has been done previously (*6*).

RESULTS

Selecting Substrates for Investigation of Positional Specificity. Cbz-FR-MCA is the most widely used substrate for cysteine proteases of the papain family. Since substrates of this general type interact mainly in the unprimed subsites of the enzymes (S_1 to S_3), and only the MCA group is positioned in the S_1' subsite (Figure 1), they are mainly useful to evaluate the basic catalytic activity and unprimed subsite specificity of these proteases. This substrate can therefore be considered as an unprimed subsite-directed substrate. By comparison, the substrate Abz-AFRSAAQ-EDDnp can make use of the extended binding site of cysteine proteases. Since it can interact in and provide information on both the S and S′ subsites of the enzymes, it can be referred to as an extended endopeptidase substrate. This substrate has been used previously to investigate the role of the occluding loop in the endopeptidase activity of cathepsin B (*6*). To monitor dipeptidyl carboxypeptidase activity, we used the IQF substrate Abz-FRF($4NO_2$)A. This compound is similar to substrates previously used to assay the exopeptidase activity of cathepsin B (*23*), except that a dansyl fluorophore is replaced by the Abz moiety, which leads to lower background fluorescence in the nonhydrolyzed substrate. This substrate interacts in subsites S_3 to S_2' of the cysteine proteases (Figure 1). Finally, the substrate FR-MCA can be used to monitor dipeptidyl aminopeptidase activity. For all of these substrates, the sequence of amino acids was chosen in order to position 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. Substrates were also prepared where these two residues are replaced by alanines, to observe the effect of alteration in S_2 and S_1 subsite interactions on positional specificity.

Investigation of Positional Specificity with F−R in P_2−P_1. The specificity profiles of papain, cathepsins B, K, and L, and DPP-I against the substrates Cbz-FR-MCA, Abz-FRF($4NO_2$)A, and FR-MCA are presented in the top panels in Figure 2. Most enzymes have relatively high activity against the unprimed subsite-directed substrate Cbz-FR-MCA, the lowest k_{cat}/K_M being 1.2×10^4 M^{-1} s^{-1} with DPP-I. Cathepsin L has been reported to be the most potent cysteine protease of the papain family (*24*). Accordingly, this enzyme displays significantly higher activity for the hydrolysis of Cbz-FR-MCA compared to the other proteases. This pattern of activity is also observed with the extended endopeptidase substrate Abz-AFRSAAQ-EDDnp (Figure 3). The lowest activities with the extended endopeptidase substrate are observed for cathepsin B and DPP-I. The low endopeptidase activity of cathepsin B can been attributed to the presence of an occluding loop in the S′ subsites of the enzyme, based on crystallographic data and mutational analysis (*6*, *25*, *26*). For DPP-I, although the molecular basis for the low endopeptidase activity has not been fully defined, it has been shown that a segment of the proregion remains associated with the enzyme (*27*) and could therefore play a role in defining activity.

A relatively broad pattern of specificity is also observed with the exopeptidase substrates Abz-FRF($4NO_2$)A and FR-MCA when the P_2−P_1 residues are F−R. As expected, cathepsin B displays excellent dipeptidyl carboxypeptidase activity against Abz-FRF($4NO_2$)A, with a k_{cat}/K_M value of 6.0×10^5 M^{-1} s^{-1}. It must be noted, however, that significant levels of activity are observed with papain, cathepsin L, and cathepsin K while once again DPP-I displays the lowest activity against this substrate. In fact, the dipeptidyl carboxypeptidase activity of cathesin L is higher than that of cathepsin B, with a k_{cat}/K_M value of 1.3×10^6 M^{-1} s^{-1}, even though cathepsin L is considered to act strictly as an endopeptidase. In fact, cathepsin L is listed by the IUBMB (International Union of Biochemistry and Molecular Biology) as having no dipeptidyl peptidase activity. The k_{cat}/K_M for hydrolysis of the substrate Abz-FRF($4NO_2$)A by cathepsin L is only 38-fold lower than that observed for hydrolysis of the extended endopeptidase substrate Abz-AFRSAAQ-EDDnp. For dipeptidyl aminopeptidase activity, k_{cat}/K_M values range from 3.0×10^2 M^{-1} s^{-1} with cathepsin K to 3.8×10^5 M^{-1} s^{-1} with DPP-I. DPP-I has been reported to be a dipeptidyl aminopeptidase and accordingly displays its highest activity against the substrate FR-MCA. It must be noted, however, that cathepsin L can hydrolyze FR-MCA as efficiently as DPP-I, with $k_{cat}/K_M = 4.0 \times 10^5$ M^{-1} s^{-1}, and that the activity with papain is 75-fold lower than that observed with DPP-I. From these results it becomes apparent that many cysteine proteases can display both endo- and exopeptidase activities. The relatively broad sequence specificity of cysteine proteases can therefore be extended also to positional specificity.

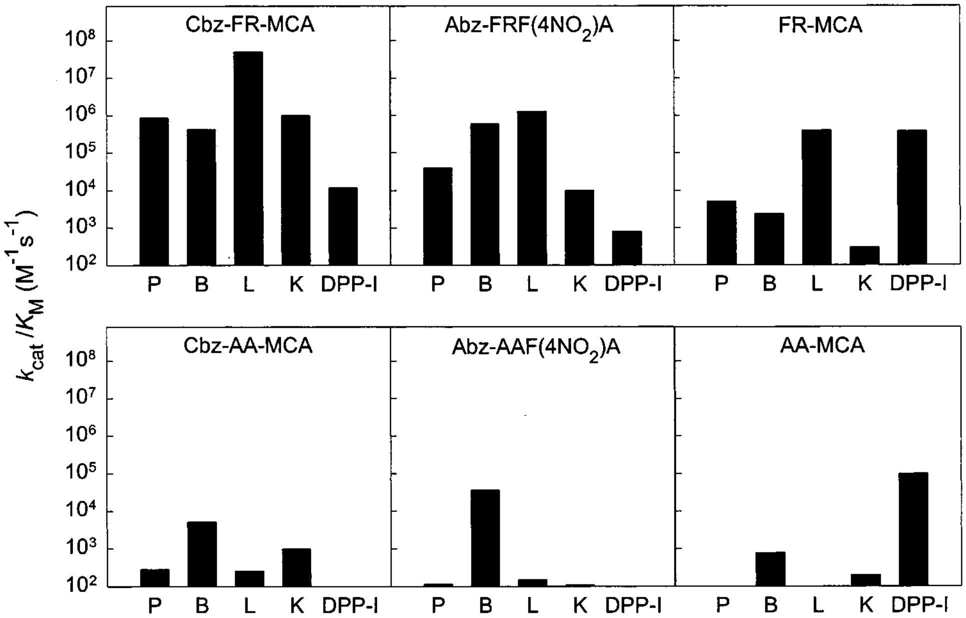

FIGURE 2: Positional and sequence specificity profiles for papain (P), cathepsins B, L, and K, and DPP-I at pH 6.0. The P_2-P_1 sequence is F−R in the top panels, and A−A in the bottom panels. Each value is the result of at least duplicate measurements. The *y*-axis cutoff is set at 10^2 M^{-1} s^{-1}.

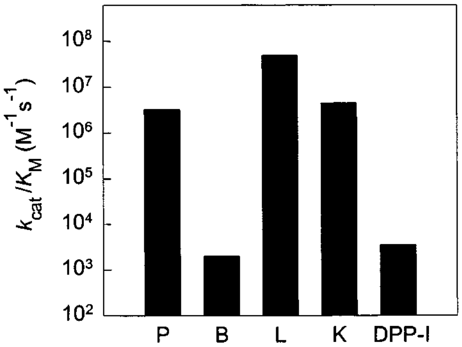

FIGURE 3: Second-order rate constant k_{cat}/K_M for hydrolysis of the extended endopeptidase substrate Abz-AFRSAAQ-EDDnp by papain, cathepsins B, L, and K and DPP-I at pH 6.0.

Investigation of Positional Specificity with A−A in P_2-P_1. Very different specificity profiles are obtained with the substrate bearing alanine residues in positions P_2 and P_1 (Figure 2, bottom panels). With Cbz-AA-MCA, k_{cat}/K_M values are lower than those obtained with the Cbz-FR-MCA by 3−5 orders of magnitude, with the exception of cathepsin B for which $k_{cat}/K_M = 5.1 \times 10^3$ M^{-1} s^{-1}, a value 80-fold lower than observed with Cbz-FR-MCA. Even though these data could be interpreted as an indication that cathepsin B can hydrolyze substrates with alanine residues in P_2-P_1 better than the other cysteine proteases in this study, an alternative explanation for the activity of cathepsin B against Cbz-AA-MCA is that the MCA moiety can interact better in the S_1' subsite of the enzyme than is the case for other cysteine proteases. The S_1' subsite of cathepsin B is defined mainly by hydrophobic residues and has been shown to display a preference for hydrophobic residues in substrates and inhibitors (*1*). For the substrate Abz-AAF(4NO$_2$)A, cathepsin B is the only enzyme with significant dipeptidyl carboxypeptidase activity (Figure 2). On going from Abz-FRF(4NO$_2$)A to Abz-AAF(4NO$_2$)A, k_{cat}/K_M decreased by only 17-fold with cathepsin B, while for cathepsin L this parameter decreased 8700-fold. Similarly, DPP-I is clearly the best dipeptidyl aminopeptidase on the basis of data for the hydrolysis of the substrate AA-MCA (Figure 2). DPP-I can hydrolyze AA-MCA almost as effectively as FR-MCA, with a k_{cat}/K_M of 9.8×10^4 M^{-1} s^{-1}. Among the other enzymes, cathepsin B has the highest activity ($k_{cat}/K_M = 775$ M^{-1} s^{-1}), possibly again reflecting the favorable interaction of the MCA group in the S_1' subsite of cathepsin B. Taken together, the results presented in Figure 2 indicate that exopeptidase activity of cysteine proteases is highly dependent on the nature of the sequence of residues in P_2-P_1.

pH Dependency of Exopeptidase Activities. Contrary to the substrates Cbz-FR-MCA and Abz-AFRSAAQ-EDDnp, the exopeptidase substrates possess groups at the N- or C-terminus which ionize in the pH range where the enzymes are active and which could therefore modulate activity in a pH-dependent manner. Figure 4 illustrates the pH dependency of activity (k_{cat}/K_M) for hydrolysis by cathepsin K of the exopeptidase substrates Abz-FRF(4NO$_2$)A (a) and FR-MCA (b), as well as that of Cbz-FR-MCA (c). For the substrate Cbz-FR-MCA, a bell-shaped curve is observed with maximal activity at pH 6. This profile is typical of many cysteine proteases of the papain family, the acid and basic limbs reflecting mainly ionization of the active site cysteine and histidine residues, with the thiolate−imidazolium ion pair being the active form of the enzyme (*28*). With the exopeptidase substrates, additional transitions are observed in the profiles which reflect variations in protonation state of the N- and C-terminus of these substrates. These additional transitions are localized in the expected pH ranges for ionization of the C-terminal carboxylic acid in Abz-FRF(4NO$_2$)A and for deprotonation of the N-terminal ammonium group in FR-MCA. Similar results were obtained with papain and cathepsin L (data not shown), except that cathepsin L is unstable at pHs higher than 6. Due to difficulty in fitting pH profiles containing very narrow peaks of activity (*29*) and to the relative instability of certain enzymes at extremes

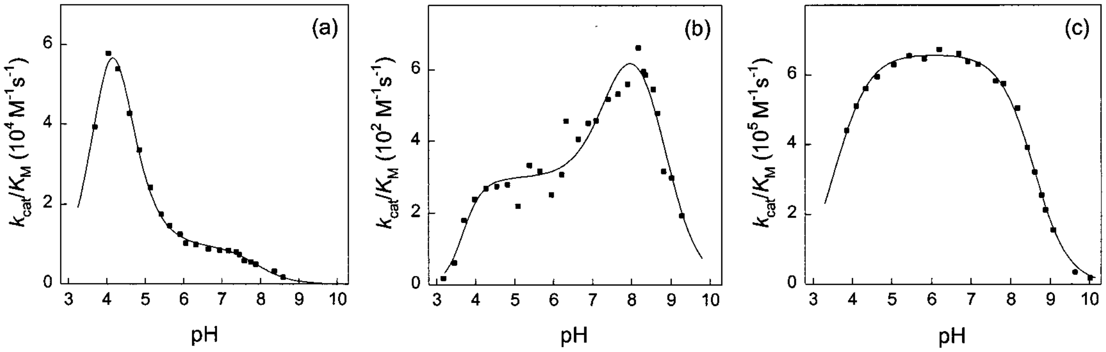

FIGURE 4: pH dependence of k_{cat}/K_M for hydrolysis of Abz-FRF(4NO$_2$)A (a), FR-MCA (b), and Cbz-FR-MCA (c) by human cathepsin K.

Table 1: Second-Order Rate Constant k_{cat}/K_M for Hydrolysis of the Exopeptidase Substrates by Cathepsin B, Cathepsin L, and DPP-I

	k_{cat}/K_M (M^{-1} s^{-1})					
	Abz-FRF(4NO$_2$)A			Abz-AAF(4NO$_2$)A		
enzyme	pH 4.0	pH 5.0	pH 6.0	pH 4.0	pH 5.0	pH 6.0
cathepsin B	467 000	800 000	600 000	120 000	102 000	35 600
cathepsin L	15 000 000	6 000 000	1 300 000	4 000	840	150
ratio B/L	0.03	0.13	0.46	30	120	230

	k_{cat}/K_M (M^{-1} s^{-1})					
	FR-MCA			AA-MCA		
enzyme	pH 4.0	pH 5.0	pH 6.0	pH 4.0	pH 5.0	pH 6.0
DPP-I	118 000	235 000	380 000	30 000	80 000	98 000
cathepsin L	12 600	90 000	400 000	81	105	62
ratio DPP-I/L	9.4	2.6	0.95	370	760	1 580

of pH, reliable kinetic constants (i.e., limiting k_{cat}/K_M values for the enzyme and the substrate in a particular protonic state) cannot be obtained from nonlinear regression analysis of the data to equations based on specific models. However it is clear from Figure 4 that the hydrolysis of the exopeptidase substrates Abz-FRF(4NO$_2$)A and FR-MCA by cathepsin K is faster when these substrates are nonionized (i.e., neutral C-terminal carboxylate and N-terminal amine, respectively). In the case of cathepsin L, for example, activity against the neutral C-terminal carboxylic acid form of Abz-FRF(4NO$_2$)A is much higher than the activity toward the negatively charged carboxylate form and contributes in large part to the cleavage of the C-terminal dipeptide. Even though at low pH the activity is higher, a substantial level of dipeptidyl carboxypeptidase activity is still observed at pH values where the substrate C-terminus is mostly deprotonated. The same comment applies to aminopeptidase activity against the unprotonated substrate. It must be noted that the activity of cathepsin L against FR-MCA greatly increases between pH 4 and 6 (Table 1) as compared to what is observed with cathepsin K (Figure 4), due to the fact that the difference in activity against the nonionized and ionized forms of FR-MCA is greater with cathepsin L than with cathepsin K.

The pH profiles for the dipeptidyl carboxypeptidase activity of cathepsin B and the dipeptidyl aminopeptidase activity of DPP-I are presented in Figure 5. Due to the presence of unique structural features and numerous ionizable groups near the active site in cathepsin B (*6*, *25*, *30*) and to the absence of a crystal structure for DPP-I, it is impossible to unambiguously assign pK_a's for the observed transitions in the pH−activity profiles to particular residues on both enzymes. However, it can be seen that the pH−activity profile for the hydrolysis of Abz-FRF(4NO$_2$)A by cathepsin B is very different from that observed with Cbz-FR-MCA or Abz-AFRSAAQ-EDDnp (*6*, *31*), and maximum activity is observed at pH 5.0. With DPP-I, maximum activity for hydrolysis of FR-MCA is found at near-neutral pH (6.8), and the profile is slightly different from that obtained with GF-MCA (unpublished data) or GF-4MbNA (*32*), for which activity is maximum at pH 5.8.

The specificity constants presented in Figure 2 were determined at pH 6.0. At that pH, the dipeptidyl carboxypeptidase activity of cathepsin B with the substrate Abz-FRF(4NO$_2$)A is 2-fold lower than that observed with cathepsin L. If the comparison is made at lower pH values, the ratio of exopeptidase activity of cathepsin B relative to cathepsin L decreases to 0.13 at pH 5.0 (the pH of maximum activity with cathepsin B) and 0.03 at pH 4.0 (Table 1). At pH 4.0, cathepsin L is thus able to hydrolyze Abz-FRF(4NO$_2$)A 33-fold faster than cathepsin B. When the P$_2$−P$_1$ residues are AA (i.e., with the substrate Abz-AAF(4NO$_2$)A), the situation is reversed. The dipeptidyl carboxypeptidase activity of cathepsin B is 30-fold higher than that of cathepsin L at pH 4.0 and 230-fold higher at pH 6.0 (Table 1). This reversal of specificity is mainly due to large decreases in k_{cat}/K_M with cathepsin L on going from Abz-FRF(4NO$_2$)A to Abz-AAF(4NO$_2$)A, indicating a strong interdependence of positional and sequence specificities with this enzyme. A similar situation is observed with DPP-I (Table 1). With the substrate FR-MCA, the ratio of k_{cat}/K_M values for DPP-I over cathepsin L varies from 9.4 at pH 4.0 to 0.95 at pH 6.0. With AA-MCA, however, DPP-I is a much better dipeptidyl aminopeptidase than cathepsin L, with ratios varying from 370 to 1580. The effect is particularly dramatic at pH 6.0, near the optimum pH for DPP-I, and can be attributed again largely to a decrease in k_{cat}/K_M with cathepsin L when replacing FR by AA in P$_2$−P$_1$ of the substrate.

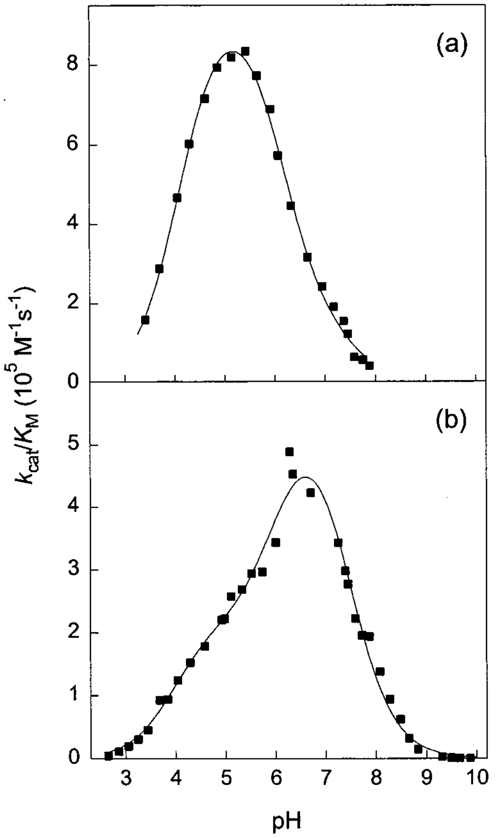

FIGURE 5: pH dependence of k_{cat}/K_M for hydrolysis of Abz-FRF-(4NO$_2$)A by cathepsin B (a) and for hydrolysis of FR-MCA by DPPI (b).

DISCUSSION

The results obtained in the present study clearly demonstrate that cathepsins K and L and papain, typically considered to act strictly as endopeptidases, can also hydrolyze substrates through exopeptidase activities. The positional specificity, however, is highly dependent on the P_2-P_1 sequence specificity. For these enzymes, exopeptidase activities are observed when the sequence of residues in P_2-P_1 of the substrate is F−R. Cathepsin L in particular, a powerful endopeptidase, was shown to possess exopeptidase activities equal to or greater than those observed with DPP-I and cathepsin B with the substrates FR-MCA and Abz-FRF-(4NO$_2$)A. Sequence specificity is therefore "over-ruling" positional specificity, most likely due to the existence of a strong S_2-P_2 interaction. This is not the case with cathepsin B and DPP-I, where replacement of F−R in P_2-P_1 of the substrates by A−A has little effect on dipeptidyl carboxypeptidase or dipeptidyl aminopeptidase activity, respectively. As a consequence, cathepsin B is the only enzyme with significant activity against the substrate Abz-AAF-(4NO$_2$)A, while DPP-I is clearly the best dipeptidyl aminopeptidase for hydrolysis of the substrate AA-MCA. By comparison, all enzymes display endopeptidase activity greatly dependent on the nature of the P_2 residue. While these results were obtained in vitro using small well-defined synthetic substrates, the contribution of the exopeptidase activity of enzymes such as cathepsin L against protein substrates in vivo is difficult to predict due to the existence of a strong competion from the endopeptidase activity.

The shape of the pH−activity profiles is very indicative of the positional specificity preference of the various proteases investigated in this work. From Figure 4 it is clear that cathepsin K displays a much higher level of exopeptidase activity when the substrate C- or N-terminus is nonionized. This is also true for cathepsin L and papain, and indicates that these enzymes are not designed to interact specifically with the charged C- or N-terminus of a substrate, but can still very efficiently hydrolyze terminal dipeptides from a substrate. It must be noted that, by definition, an exopeptidase activity is observed when a reaction occurs at the free N- or C-terminus of substrates, independent of the charged state of that terminal group. For cathepsin B and DPP-I, even though the relative activities against the neutral and ionized forms of the substrates cannot be dissected out, it is obvious from the profiles in Figure 5 that these enzymes are able "by design" to function as exopeptidases. Most cysteine proteases of the papain family are endopeptidases and, on the basis of the currently available information, it is probable that all exopeptidases within this family have evolved from the basic endopeptidase platform (*16*). Cathepsins K and L and papain do not possess any structural features to specifically interact with the C- or N-terminus of an exopeptidase substrate and therefore rely heavily on interactions in subsite S_2 (and possibly S_1) for exopeptidase activity. Exopeptidases such as cathepsin B, cathepsin H, and bleomycin hydrolase, for which 3-D structures are available (*25*, *33*, *34*), consist of a papain-like two-domain catalytic platform. However, each of these three enzymes has additional novel structural elements that confer upon them their exopeptidase capabilities. Cathepsin B has an insert that forms a surface loop that partially occludes the binding cleft beyond the S_2' subsite (*25*). Two histidine residues in that loop interact with the C-terminus of the substrate, thus defining the exopeptidase specificity. Cathepsin H, an aminopeptidase, retains an eight amino acid fragment (minichain) derived from its prosequence that is covalently linked via a disulfide bond to the endopeptidase platform (*33*). The side chain of the N-terminal glutamic acid of the minichain interacts with the N-terminus of the aminopeptidase substrate. Bleomycin hydrolase, also an aminopeptidase, has a C-terminal extension that loops around the molecule such that the terminal carboxylate binds in the active site and interacts with the N-terminus of the aminopeptidase substrate (*34*). The modification of the cysteine protease endopeptidase platform utilized by DPP-I to incorporate an exopeptidase activity remains to be determined, but it is clear that this enzyme also contains structural features beyond those of the simple papain framework.

As shown in this study, the relatively broad specificity of cysteine proteases includes not only sequence specificity but also positional specificity. For enzymes such as cathepsin B and DPP-I, for example, the sequence specificity for exopeptidase activity might even be lower than that observed for endopeptidase activities with cathepsin B and other cysteine proteases. The nature of the residue at position P_2 of a substrate, which is usually the main factor determining specificity for cysteine proteases of the papain family, does not make the same contribution to the exopeptidase activity of cathepsin B. In addition, positional and sequence specificities can be interdependent. These considerations can be important when designing inhibitors for cysteine proteases. For example, since the extended binding site of cysteine proteases provides numerous possibilities for the formation of enzyme−ligand interactions, it is tempting to optimize the number of interactions between an inhibitor and the protein template in order to maximize potency. Ideally, each enzyme−ligand interaction introduced should be specific to the particular enzyme target. If a given interaction is common to other enzymes, the gain in potency might be at the expense of inhibitor selectivity. Results obtained in this study with substrates can be used to illustrate this problem. The substrate Abz-AAF(NO_2)A can interact with the occluding loop in the primed subsites of cathepsin B but does not possess optimal residues for interaction in the unprimed subsites. This compound is a much better substrate of cathepsin B than cathepsin L. If we introduce F−R in P_2−P_1, the resulting compound is a 10-fold better substrate of cathepsin B, but becomes an even better substrate for cathepsin L, corresponding to a reversal of specificity of up to 1 000-fold. By analogy, to design a specific cathepsin B inhibitor, a good strategy might be to exploit mainly the primed subsite loop of cathepsin B, which is absent in most cysteine proteases of the papain family, and avoid interactions in the unprimed subsites which are usually common to many enzymes. This has the additional advantage that the resulting compound would be of lower molecular weight than if the entire extended binding site is utilized. If extension of the inhibitor to the unprimed subsites is attempted, the interactions introduced should be specific for cathepsin B (e.g., make use of the particular S_2 subsite specificity (*1*)). The existence of a common basic structural template is one of the reasons why cysteine proteases are particulary challenging targets for the design of specific inhibitors.

The functional properties of cysteine proteases, and in particular substrate specificity, have been extensively studied over the years. However, certain aspects remain that have not been fully explored. The present work shows that the positional specificity of these enzymes is broader than is normally believed. Enzymes that are classified as acting strictly as endopeptidases can display very important levels of exopeptidase activity. More importantly, it was shown that the S_2 subsite specificity of cysteine proteases, the hallmark of this group of enzymes, is much less important for enzymes specifically designed to act as exopeptidases. Very little has been reported regarding the sequence specificity for the exopeptidase versus endopeptidase activities; thus the novel findings of this work significantly extend our understanding of the concepts underlying the specificity of cysteine proteases.

ACKNOWLEDGMENT

We would like to thank Dr. Luiz Juliano for providing one of the substrates, Manon Lalumière for the cathepsin K construct, Elisa de Miguel for peptide synthesis, and Boehringer Ingelheim Canada (BioMéga Research Division) for the use of a MALDI-TOF mass spectrometer.

REFERENCES

1. Storer, A. C., and Ménard, R. (1996) *Perspect. Drug Discovery Des. 6*, 33−46.
2. Berger, A., and Schechter, I. (1970) *Philos. Trans. R. Soc. London, Ser. B 257*, 249−264.
3. Barrett, A. J. (1980) *Biochem. J. 187*, 909−912.
4. Barrett, A. J., and Kirschke, H. (1981) *Methods Enzymol. 80*, 535−561.
5. Koga, H., Yamada, H., Nishimura, Y., and Imoto, T. (1991) *J. Biochem. 110*, 179−188.
6. Nägler, D. K., Storer, A. C., Portaro, F. C. V., Carmona, E., Juliano, L., and Ménard, R. (1997) *Biochemistry 36*, 12608−12615.
7. Lalmanach, G., Serveau, C., Brillard-Bourdet, M., Chagas, J. R., Mayer, R., Juliano, L., and Gauthier, F. (1995) *J. Protein Chem. 14*, 645−653.
8. Fruton, J. S. (1960) in *The Enzymes* (Boyer, P. D., Ed.) pp 233−241, Academic Press, New York.
9. Kirschke, H., and Barrett, A. J. (1987) in *Lysosomes: Their Role in Protein Breakdown* (Glaumann, H., and Ballard, F. J., Eds.) pp 193−238, Academic Press, London, U.K.
10. Kirschke, H., Langner, J., Wiederanders, B., Ansorge, S., Bohley, P., and Hanson, H. (1977) *Acta Biol. Med. Ger. 36*, 185−199.
11. Kuribayashi, M., Yamada, H., Ohmori, T., Yanai, T., and Imoto, T. (1993) *J. Biochem. 113*, 441−449.
12. Müller-Ladner, U., Gay, R. E., and Gay, S. (1996) *Perspect. Drug Discovery Des. 6*, 87−98.
13. Chapman, H. A., Riese, R. J., and Shi, G.-P. (1997) *Annu. Rev. Physiol. 59*, 63−88.
14. Mort, J. S., and Buttle, D. J. (1997) *Int. J. Biochem. Cell Biol. 29*, 715−720.
15. Schechter, I., and Berger, A. (1967) *Biochem. Biophys. Res. Commun. 27*, 157−162.
16. Turk, D., Guncar, G., Podobnik, M., and Turk, B. (1998) *Biol. Chem. 379*, 137−147.
17. Thompson, S. K., Halbert, S. M., Bossard, M. J., Tomaszek, T. A., Levy, M. A., Zhao, B., Smith, W. W., Abdel-Meguid, S. S., Janson, C. A., D'Alessio, K. J., McQueney, M. S., Amegadzie, B. Y., Hanning, C. R., DesJarlais, R. L., Briand, J., Sarkar, S. K., Huddleston, M. J., Ijames, C. F., Carr, S. A., Garnes, K. T., Shu, A., Heys, J. R., Bradbeer, J., Zembryki, D., Rykaczewski, L. L., James, I. E., Lark, M. W., Drake, F. H., Gowen, M., Gleason, J. G., and Veber, D. F. (1997) *Proc. Natl. Acad. Sci. U.S.A. 94*, 14249−14254.
18. Schaschke, N., Assfalg-Machleidt, I., Machleidt, W., and Moroder, L. (1998) *FEBS Lett. 421*, 80−82.
19. Carmona, E., Dufour, E., Plouffe, C., Takebe, S., Mason, P., Mort, J. S., and Ménard, R. (1996) *Biochemistry 35*, 8149−8157.
20. 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.
21. Chagas, J. R., Juliano, L., and Prado, E. S. (1991) *Anal. Biochem. 192*, 419−425.
22. 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.
23. Pohl, J., Davinic, S., Blaha, I., Strop, P., and Kostka, V. (1987) *Anal. Biochem. 165*, 96−101.
24. Barrett, A. J., and Kirschke, H. (1981) *Methods Enzymol. 80*, 535−561.
25. 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.
26. Illy, C., Quraishi, O., Wang, J., Purisima, E., Vernet, T., and Mort, J. (1997) *J. Biol Chem. 272*, 1197−1202.
27. Cigić, B., Križaj, I., Kralj, B., Turk, V., and Pain, R. H. (1998) *Biochim. Biophys. Acta 1382*, 143−150.
28. Storer, A. C., and Ménard, R. (1994) *Methods Enzymol. 244*, 486−500.
29. Dixon, H. B. F. (1973) *Biochem. J. 131*, 149−154.
30. Polgár, L., and Csoma, C. (1987) *J. Biol. Chem. 262*, 14448−14453.
31. Khouri, H. E., Plouffe, C., Hasnain, S., Hirama, T., Storer, A. C., and Ménard, R. (1991) *Biochem. J. 275*, 751−757.
32. Dolenc, I., Turk, B., Pungerčič, G., Ritonja, A., and Turk, V. (1995) *J. Biol. Chem. 270*, 21626−21631.
33. Gunčar, G., Podobnik, M., Pungerčar, J., Štrukelj, B., Turk, V., and Turk, D. (1998) *Structure 6*, 51−61.
34. Joshua-Tor, L., Xu, H. E., Johnston, S. A., and Rees, D. C. (1995) *Science 269*, 945−950.

BI982632S