Potency and Selectivity of the Cathepsin L Propeptide as an Inhibitor of Cysteine Proteases†

Euridice Carmona,‡,§ Éric Dufour,‡,|| Céline Plouffe,‡ Sachiko Takebe,‡ Patrizia Mason,⊥ John S. Mort,⊥,# and Robert Ménard*,‡

Pharmaceutical Biotechnology Sector, Biotechnology Research Institute, National Research Council of Canada, 6100 Royalmount Avenue, Montreal, Quebec, Canada H4P 2R2, Joint Diseases Laboratory, Shriners Hospital for Crippled Children, 1529 Cedar Avenue, Montreal, Quebec, Canada H3G 1A6, and Department of Surgery, McGill University, Montreal, Quebec, Canada

Received November 16, 1995; Revised Manuscript Received March 15, 1996⊗

ABSTRACT: The cathepsin L propeptide (phcl-2) was expressed in *Saccharomyces cerevisiae* using a human procathepsin L/α-factor fusion construct containing a stop codon at position −1 (the C-terminal amino acid of the proregion). Since the yield after purification was very low, the cathepsin L propeptide was also obtained by an alternate procedure through controlled processing of an inactive mutant of procathepsin L (Cys25Ser/Thr110Ala) expressed in *Pichia pastoris*, by small amounts of cathepsin L. The peptide resulting from the cleavage of the proenzyme (phcl-1) was then purified by HPLC. The purified propeptides were characterized by N-terminal sequencing and mass spectrometry and correspond to incomplete forms of the proregion (87 and 81 aa for phcl-1 and phcl-2 respectively, compared to 96 aa for the complete cathepsin L propeptide). The two peptides were found to be potent and selective inhibitors of cathepsin L at pH 5.5, with K_i values of 0.088 nM for phcl-1 and 0.66 nM for phcl-2. The K_i for inhibition of cathepsin S was much higher (44.6 nM with phcl-1), and no inhibition of cathepsin B or papain could be detected at up to 1 μM of the propeptide. The inhibitory activity was also found to be strongly pH-dependent. Two synthetic peptides of 75 and 44 aa corresponding to N-terminal truncated versions of the propeptide were also prepared by solid phase synthesis and displayed K_i values of 11 nM and 2900 nM, respectively, against cathepsin L. The data obtained for the 4 propeptide derivatives of various lengths indicate that the first 20 residues in the N-terminal region of the propeptide are more important for inhibition than the C-terminal region which contributes little to the overall inhibitory activity.

Cysteine proteases of the papain superfamily have been implicated in a number of degradative and invasive processes. For example, cathepsins B and L appear to play a role in arthritis (Trabandt et al., 1991; Esser et al., 1994), tumor invasion and metastasis (Sheahan et al., 1989; Duffy, 1992), and muscular dystrophy (Katunuma & Kominami, 1987). These enzymes therefore constitute attractive targets for the development of inhibitors as potential therapeutic agents. A number of compounds are known which can inhibit cysteine proteases (Rich, 1986; Demuth, 1990; Shaw, 1990). These inhibitors have been developed mainly on the basis of affinity labeling of the active site cysteine residue and make use of the known substrate specificities of the enzymes. Selectivity for the cysteine proteases as opposed to other classes of proteases can be achieved in this manner. However, designing an inhibitor highly selective for a given cysteine protease constitutes a challenge due to the relatively broad substrate specificity of the papain superfamily (Brocklehurst et al., 1987).

A number of studies have indicated that peptides derived from the proregion[1] of various protease zymogens can inhibit their corresponding enzymes (e.g., San Segundo et al., 1982; Fusek et al., 1991; Baker et al., 1992). This is also true for the cysteine proteases, where it has been shown recently that the propeptide of rat cathepsin B is a potent inhibitor of the mature enzyme with $K_i = 0.4$ nM (Fox et al., 1992). The propeptides of two other cysteine proteases, papain and papaya proteinase IV (PPIV), were also shown to be relatively good inhibitors for cysteine proteases (Taylor et al., 1995). Interestingly, the cathepsin B propeptide is highly selective for cathepsin B, displaying a relatively weak inhibitory activity against papain ($K_i = 5.6$ μM), while enzyme selectivity was not observed among the plant-derived propeptides. Little is known about the structural features and binding modes of these propeptides to their target cysteine proteases. Clearly, a better understanding of these aspects could provide valuable information for the development of potent and highly selective inhibitors.

† NRCC Publication No. 39920. Supported by postdoctoral fellowships from the Conselho Nacional de Desenvolvimento Cientifico e Tecnologico, Brazil (E.C.), as well as NATO and the Ministère de la recherche et de l'espace, France (E.D.). The Biotechnology Research Institute and McGill University are members of the Protein Engineering Network of Centres of Excellence, sponsored by the government of Canada. J.S.M. is supported by the Shriners of North America.

* Author to whom correspondence should be addressed.

‡ Biotechnology Research Institute.

§ Visiting scientist from Instituto Butantan, C.P. 65, 05504, São Paulo, Brazil.

|| Visiting scientist from Institut National de la Recherche Agronomique, BP 1627, 44316 Nantes Cedex 03, France.

⊥ Joint Diseases Laboratory, Shriners Hospital for Crippled Children.

Department of Surgery, McGill University.

⊗ Abstract published in *Advance ACS Abstracts,* June 1, 1996.

[1] In the text, proregion refers to the peptide stretch located N-terminal to the mature enzyme in the proenzyme, while propeptide refers to a peptide corresponding to the proregion sequence but without the mature enzyme.

S0006-2960(95)02736-X CCC: $12.00

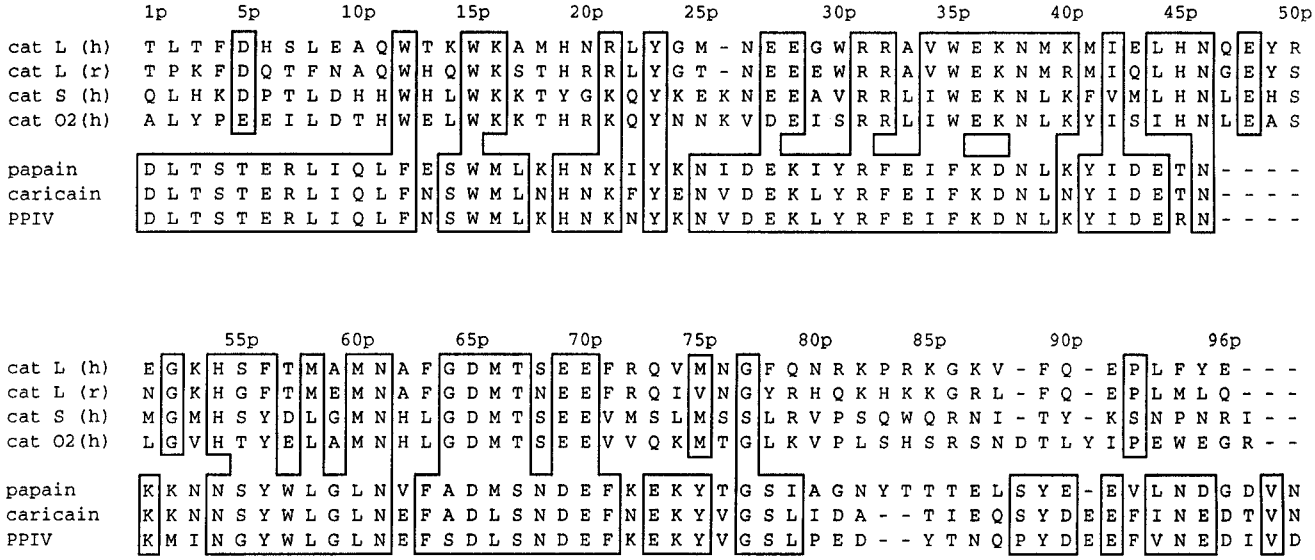


FIGURE 1: Sequence alignment of the proregions of cathepsin L from human (Gal & Gottesman, 1988) and rat (Ishidoh et al., 1987b), human cathepsin S (Wiederanders et al., 1992), human cathepsin O2, also designated as cathepsin O, and cathepsin K (Brömme & Okamoto, 1995; Inaoka et al., 1995; Shi et al., 1995), papain (Cohen et al., 1986), caricain (Revell et al., 1993), and PPIV (Taylor et al., 1995). An additional sequence has been published for cathepsin S containing a Thr at position 75p instead of Met (Shi et al., 1992). The papain, caricain, and PPIV proregions contain 10 additional N-terminal residues not shown in the figure. Sequence homologies are based on the mutation data matrix and reflect similarity of amino acid functions in their interactions with other amino acids in a protein (Schwartz & Dayhoff, 1979; George et al., 1990). The amino acids are arranged in the following groups: aromatic (Phe, Tyr, Trp), hydrophobic (Val, Leu, Met, Ile), basic (Lys, Arg, His), acid (Asp, Glu), amide (Asn, Gln), small (Gly, Ala, Pro, Thr, Ser), and Cys. Numbering is based on the cathepsin L proregion sequence.

The selectivity of propeptides for their corresponding mammalian cysteine proteases has not been investigated so far. The cathepsin B group of enzymes forms a distinct subclass of the papain family (Berti & Storer, 1995). This distinction between cysteine proteases can be extended to their proregions, which can be assigned to two groups. The first group contains the cathepsin L-like enzymes and includes cathepsin S, cathepsin H, and the papain-like plant enzymes. Proregions of enzymes in this group are typically greater than 90 amino acids in length. The second group is characterized by a much smaller proregion (62 amino acids) and comprises the cathepsins B from various sources. This difference might account for the observed selectivity of the cathepsin B propeptide for cathepsin B over papain. It remained to be seen, however, if the cathepsin L propeptide can display selectivity against enzymes of the cathepsin L group.

Within the cathepsin L group, the prosequences of a number of cysteine proteases are known. Those of the mammalian enzymes cathepsin L (human and rat) and cathepsin S (human) as well as the recently described sequence for cathepsin O2 (human) derived from cDNA cloning, in addition to the plant enzymes papain, caricain, and papaya proteinase IV, are shown in Figure 1. Even though the overall level of sequence identity is lower than that observed for the mature enzyme region, similarities can be observed between the sequences. In particular, two conserved motifs have been reported previously. The first motif is located between residues 59p and 65p[2] and consists of the sequence Gly-X_1-Asn-X_1-Phe-X_1-Asp (GNFD motif), where X_n represents a number (n) of variable amino acids (Ishidoh et al., 1987a; Vernet et al., 1995). A similar motif is also found in the cathepsin B group of propeptides (Gly-X_1-Asn-X_1-Tyr-X_2-Asp). The second motif identified in the cathepsin L group corresponds to Glu-X_3-Arg-X_3-Phe-X_2-Asn-X_3-Ile-X_3-Asn (ERFNIN motif), and is not found in the cathepsin B group (Karrer et al., 1993). In addition to these two motifs, various degrees of homology between the propeptides are also observed between positions 5p and 77p (Figure 1). The C-terminal region of the propeptides, however, varies greatly for the different species, for both the cathepsin B and cathepsin L propeptide groups. In addition, the N-terminal region of the cathepsin L proregion has been found to interact with a 43 kDa integral membrane protein and mediate procathepsin L binding to microsomal membranes at acidic pH (McIntyre & Erickson, 1993; McIntyre et al., 1994). The pH-dependency of this event suggests that membrane association could play a role in the intracellular transport of the proenzyme and/or in the processing of procathepsin L in acidified vesicles. Based on these findings, it appeared that the first 24 amino acids of the cathepsin L propeptide may be nonessential for inhibition of cathepsin L.

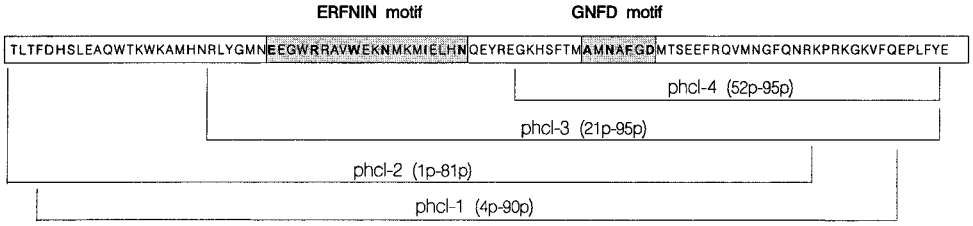
In the present study, the cathepsin L propeptide has been produced and tested against a number of cysteine proteases to investigate the selectivity of enzyme inhibition within the cathepsin L group of enzymes as well as with cathepsin B. Structural features of the propeptide, both free in solution and bound to cathepsin L, have been obtained through circular dichroism measurements. Truncated versions of the propeptide have also been characterized to evaluate the contribution of N- and C-terminal regions to inhibition.

MATERIALS AND METHODS

The substrate CBZ-Phe-Arg-MCA (carbobenzoxy-L-phenylalanyl-L-arginine-4-methylcoumarinyl-7-amide hydrochloride) and the inhibitor E64 [1-(L-*trans*-epoxysuccinyl-

[2] Residues in the propeptide are identified with the suffix p. Propeptide numbering is based on the procathepsin L sequence starting at the N-terminal amino acid (residue 1p). The C-terminal amino acid of the propeptide corresponds to position 96p and is followed by the N-terminal residue (position 1) of the mature enzyme.

Chart 1



L-leucylamino)-4-guanidinobutane] were purchased from IAF Biochem International Inc., Laval, Canada. Papain was obtained as the crystallized suspension in sodium acetate from Sigma Chemical Co. and was further purified and activated and the active site titrated as described previously (Ménard et al., 1990). Recombinant cathepsins B and S were prepared as described previously (Fox et al., 1992; Brömme et al., 1993). *Saccharomyces cerevisiae* strain BJ3501 was obtained from the Yeast Genetic Stock Center (University of California, Berkeley). The vector (pPIC9) and *Pichia pastoris* strain GS115 were purchased from Invitrogen Corp. (San Diego, CA).

Expression and Purification of Recombinant Human Cathepsin L. Human procathepsin L was expressed in *S. cerevisiae* as an α-factor fusion construct as used previously for the production of rat cathepsin B (Rowan et al., 1992). Primary cultures were grown for 2 days in selective medium (COMP-Ura) (Sherman et al., 1986) at 30 °C in a gyratory shaker (225 rpm), and secondary cultures were grown 3 additional days in a modification of the "production" medium developed by Ernst (1986) (SD + 4% casamino acids, buffered to pH 5.5 with 50 mM MES). Activity of the mature enzyme with the substrate CBZ-Phe-Arg-MCA was detected in the supernatant, indicating that processing of the proenzyme occurred during culture. The medium was centrifuged (10000*g*, 15 min at 4 °C), the supernatant (2 L) was concentrated on a spiral membrane (S1Y10, Amicon) and further concentrated, and the buffer was exchanged into 50 mM sodium acetate, pH 5.0, using a stirred ultrafiltration cell (YM-10 membrane, Amicon). The concentrate (50 mL) was applied to an S-Sepharose column (1 × 18 cm) equilibrated in 50 mM sodium acetate buffer, pH 5.0, using a Pharmacia FPLC system. The majority of the contaminating proteins were not retained by the column, and the enzyme was eluted with a NaCl gradient (0−0.5 M). SDS−PAGE of the active fractions demonstrated that the enzyme preparation was pure. In order to avoid autolysis, $HgCl_2$ (1 mM) was added to maintain the purified cathepsin L as a reversibly inhibited form.

Expression and Purification of Recombinant Human Cathepsin L Propeptide (phcl-2). By site-directed mutagenesis (Kunkel, 1985), the oligonucleotide 5′-CCTCTGTTTTACTAGGCCCCCAG-3′ was used to introduce a stop codon at position −1 (the C-terminal amino acid of the proregion) of the human procathepsin L/α-factor fusion construct. This oligonucleotide also introduced a *Bfa*I restriction site which was used for mutant screening. Culture supernatants prepared as above were concentrated by methanol precipitation (1:5), and the dried pellet was dissolved in water. The presence of the cathepsin L propeptide was detected by inhibition of cathepsin L activity. Control samples prepared from the culture media of yeast containing vector without the cathepsin L insert showed no inhibitory activity.

For propeptide purification, culture medium (4 L) was concentrated in an Amicon stirred cell to approximately 200 mL using a YM-10 membrane. The inhibitory activity against cathepsin L was found in the retentate which was precipitated by the addition of ammonium sulfate to 50% saturation and collected by centrifugation at 12000*g* for 20 min. The resulting pellet was dissolved in 10 mL of 50 mM phosphate buffer, pH 7, and loaded on a Vydac C4 column (1 × 25 cm). The propeptide was eluted at a flow rate of 3 mL/min with a two-step gradient (0−25% CH_3CN in 8 min, followed by 25−50% CH_3CN in 22 min, all in the presence of 0.1% TFA). Active fractions were purified to homogeneity by reverse-phase HPLC on an analytical Vydac C4 column (4.6 × 250 mm) using the gradient described above at a flow rate of 1 mL/min. Approximately 0.01 mg of the propeptide of human cathepsin L (phcl-2) was obtained from 4 L of culture medium.

Preparation and Purification of Cathepsin L Propeptide (phcl-1) from Recombinant Human Procathepsin L (Cys25Ser). By site-directed mutagenesis using the oligonucleotide GGGTCAGTGTGGCTCTTCTTGGGCTTTTAGTGC, the active site cysteine residue of the procathepsin L insert was converted to a serine (Cys25Ser) to prevent autoprocessing of the proenzyme. This oligonucleotide also introduced an *Ear*I site. In addition, the consensus sequence for N-linked oligosaccharide substitution (Asn108-Asp109-Thr110) was removed by introduction of the mutation Thr110Ala using the oligonucleotide TGCTAATGACGCCCGGCTTTGTGG which introduced an *Acy*I site. The mutated procathepsin L cDNA was inserted into the vector pPIC9 and used to transform *Pichia pastoris*.

Recombinant procathepsin L was produced using the culture conditions recommended by Invitrogen. A starter culture of yeast was cultured in 5 L of buffered minimal glycerol-complex medium (BMGY) at 30 °C with vigorous shaking for 2 days. The cells were collected under sterile conditions by centrifugation at 4 °C (4000*g*, 15 min) and used to inoculate 1 L of buffered minimal methanol-complex medium (0.5% methanol). Culture was continued for 3 days at 30 °C, again with vigorous shaking. After the first 2 days, additional methanol was added (half of the initial amount). The culture medium was clarified by centrifugation (4000*g*, 15 min) and the supernatant concentrated using an Amicon Spiral Concentrator with a S1Y10 membrane to about 150 mL and then to 40−45 mL using an Amicon PM10 membrane in a stirred cell. The concentrated solution was dialyzed overnight at 4 °C into 50 mM sodium acetate, pH 5.0, containing 0.02% sodium azide and 0.075 M NaCl (to prevent precipitation of the mutant protein). The solution was applied to a SP-Sepharose column (1.6 × 18 cm) equilibrated with the previous buffer at a flow rate of 1 mL/min. Elution was carried out using a 600 mL linear gradient to 0.5 M sodium chloride. Fractions were monitored for the mutant by SDS−PAGE. The yield of pure procathepsin L

Cys25Ser/Thr110Ala mutant was generally 10 mg/L of BMGY culture medium.

The purified proenzyme was processed to yield human cathepsin L propeptide (phcl-1) by addition of small amounts of active cathepsin L [procathepsin L (50 μM) incubated for 1 h with 70 nM cathepsin L, at 37 °C, pH 5.1]. The processing was stopped by addition of E-64 (0.7 μM). Processing yielded more than one form of the peptide (as detected by HPLC) resulting from cleavage at multiple positions. The major form was purified to homogeneity by reverse-phase HPLC on an analytical Vydac C4 column (4.6 × 250 mm) using a two-step gradient (0−30% CH_3CN in 10 min followed by 30−50% CH_3CN in 35 min, all in 0.1% TFA) at a flow rate of 1 mL/min.

Processing of procathepsin L Cys25Ser/Thr110Ala was also used to obtain the mature inactive cathepsin L (Cys25Ser/Thr110Ala). The mixture containing the processed proenzyme was dialyzed against 50 mM sodium acetate buffer, pH 5.0, containing 0.02% sodium azide and applied to a SP-Sepharose column (1 × 18 cm). Elution was carried out using a 250 mL linear gradient to 0.5 M sodium chloride. Mature cathepsin L Cys25Ser/Thr110Ala (single-chain form) was identified by SDS−PAGE, and N-terminal sequencing. This protein was used for circular dichroism measurements.

Synthesis and Purification of Truncated Cathepsin L Propeptides (phcl-3 and phcl-4). Two peptides corresponding to truncated versions of the human cathepsin L propeptide were synthesized using Fmoc chemistry (FastMoc) on an Applied Biosystems 431A solid-phase synthesizer. Peptide phcl-3 consisted of 75 amino acids corresponding to residues 21p−95p of the cathepsin L propeptide, while phcl-4 consisted of only 44 residues (52p−95p). The Glu residue at position 96p was not included since it was observed that active recombinant cathepsin L produced during autoprocessing of procathepsin L retained this residue. The crude peptides were purified to homogeneity by reverse-phase HPLC on a semipreparative Vydac C4 column (1 × 25 cm) using a linear elution gradient (10−50% acetonitrile in 40 min) at a flow rate of 3 mL/min.

Characterization of Cathepsin L Propeptides. Amino acid compositions of the synthetic peptides were verified by analysis on a Beckmann Model 6300 amino acid analyzer. The recombinant propeptides phcl-1 and phcl-2 were characterized by their N-terminal sequence coupled to molecular mass determination. N-Terminal sequences were determined on an Applied Biosystems 473A pulsed-liquid peptide sequencer. Mass spectrometric analysis was performed on a triple quadrupole mass spectrometer (API III LC/MS/MS system; Sciex, Thornhill, Canada) or a Kompact Maldi II analytical mass spectrometer (Kratos, Hamilton, Canada). Molecular masses were assigned to specific sequences using the program MSU-MASSMAP written by Liao et al. (1994). Propeptide phcl-1 was found to consist of 87 amino acids corresponding to residues 4p−90p and contained a Leu residue at position 78p (instead of Phe) apparently due to a point mutation that occurred after subcloning. The peptide phcl-2 corresponds to residues 1p−81p (81 amino acids) and was found to be oxidized, most probably at the methionine residues (the propeptide contains 8 such residues). Only a very small amount of oxidized peptide was detected with phcl-1. Since the peptides phcl-1 and phcl-2 were found to have a mainly disordered structure in aqueous solution and both possess four tryptophan and two tyrosine residues, the extinction coefficients of free tryptophan (ϵ = 5690 M^{-1} cm^{-1}) and free tyrosine (ϵ = 1280 M^{-1} cm^{-1}) at 280 nm (Gill & von Hippel, 1989) were used to estimate the extinction coefficients of the propeptides (25 320 M^{-1} cm^{-1}).

Kinetic Measurements. All kinetic measurements were performed at 25 °C using the substrate CBZ-Phe-Arg-MCA to monitor enzyme activity as described previously (Ménard et al., 1990). The assay conditions consisted of 50 mM sodium citrate buffer (pH 4.0−5.5), 0.2 M NaCl, 1 mM EDTA, 3 mM DTT, and 10% CH_3CN. The dissociation constants (K_i) for inhibition of cysteine proteases by the propeptides were determined by measuring the initial rate of substrate hydrolysis (v_s) in the presence of varying concentrations of inhibitor and at substrate concentrations kept well below K_M. Inhibitor and substrate concentrations were kept in excess over enzyme concentration. The K_i values at a given pH were obtained from a graph of $1/v_s$ vs [inhibitor] (Dixon, 1953). For inhibition of cathepsin L by peptides phcl-1, -2, and -3, nonlinearity in the initial portion of the progress curves indicated the presence of a "slow inhibition" process, and the rates were measured only after steady-state conditions had been reached. Due to the time needed to reach steady-state at very low inhibitor concentrations, it was not possible in those cases to use inhibitor concentrations lower or equal to the observed dissociation constant, so K_i was calculated from the relationship $v_0/v_s - 1 = [I]/K_i$ (Izquierdo-Martin & Stein, 1992), where v_0 represents the initial rate for substrate hydrolysis in the absence of inhibitor. The reported values of K_i are the averages of at least six measurements. For reactions exhibiting a slow-inhibition process, the pre-steady-state portion of the progress curves was also used to obtain the kinetic parameters k_1 and k_2 of Scheme 1, which represents the simplest model describing the inhibition of a cysteine protease by its propeptide based on the finding that k_{obs} is linear with inhibitor concentration up to the highest concentration used (Fox et al., 1992). This was achieved by fitting the curves at different inhibitor concentrations to eqs 1 and 2 (Cha, 1975; Williams & Morrison, 1979) where k_{obs} is the apparent first-order rate constant to reach steady-state. Due to the near-zero intercepts in the k_{obs} vs [inhibitor] plots, k_2 was calculated using the relationship $K_i = k_2/k_1$.

Scheme 1

$$E + I \underset{k_2}{\overset{k_1}{\rightleftharpoons}} EI$$

$$[P] = v_s t + \frac{(v_0 - v_s)[1 - \exp(-k_{obs} t)]}{k_{obs}} \quad (1)$$

$$k_{obs} = k_1[I] + k_2 \quad (2)$$

Circular Dichroism Measurements. CD spectra were measured on a Jobin Yvon CD6 dichrograph equipped with a thermostated cell holder, and data were recorded on-line using a personal computer. Spectra presented are averages of 8 accumulated scans with subtraction of the base line. The cells used had a path length of 0.02 cm (cathepsin L), 0.05 cm (cathepsin L−propeptide complex), or 0.1 cm (propeptide) for experiments in the far-UV spectral region (185−260 nm). For the near-UV spectral region (250−320 nm), cells having a path length of 0.5 cm (propeptide,

Table 1: Comparison of Pre-Steady-State Kinetic Parameters and Inhibitory Activity against Cysteine Proteases for Various Forms of the Cathepsin L Propeptide[a]

	cathepsin L			K_i (nM)		
propeptide	k_1 ($\times 10^6$ M^{-1} s^{-1})	k_2 ($\times 10^{-4}$ s^{-1})	K_i (nM)	cathepsin S	papain	cathepsin B
phcl-1 (4p−90p)	2.76 ± 0.82	2.43 ± 0.81	0.088 ± 0.013	44.6 ± 7.0	>1000[b]	>1000
phcl-2 (1p−81p)	1.38 ± 0.46	9.11 ± 3.84	0.66 ± 0.17			
phcl-3 (21p−95p)	0.20 ± 0.11	23.0 ± 16.8	11.5 ± 5.5	>1000	>1000	>1000
phcl-4 (52p−95p)			2900 ± 400			

[a] The experiments were done at pH 5.5 for cathepsin L and at pH 6.0 for cathepsins S and B and papain. [b] No inhibition was detected at 1 μM inhibitor concentration.

cathepsin L-propeptide complex) or 0.2 cm (cathepsin L) were used. All experiments were performed at 25 °C in 20 mM acetate buffer, pH 5.5, in the presence of 10% acetonitrile. Cathepsin L concentrations were 18.9 μM and 9.45 μM for the spectra in near- and far-UV regions, respectively. The concentration of the propeptide and of the cathepsin L−propeptide complex (1:1 ratio) was 25 μM in all experiments. The results are expressed in terms of molar ellipticity for the aromatic region and in terms of mean residue ellipticity for the peptide region.

RESULTS

Production of Propeptides in Yeast. Since the propeptide of cathepsin L is relatively long, expression in yeast was preferred over chemical synthesis to produce the peptide. Using *Saccharomyces cerevisiae*, the propeptide could be expressed successfully, but a number of problems were encountered. The yield after purification was extremely low, at approximately 0.01 mg from 4 L of culture, and the propeptide was found to be partly hydrolyzed, lacking the C-terminal 15 amino acid residues. In addition, oxidation of methionine residues was observed by mass spectrometry (up to eight additional oxygen atoms on the peptide). In an attempt to obtain the propeptide in higher yield, an inactive mutant of procathepsin L (Cys25Ser, also containing the Thr110Ala mutation to remove the glycosylation site) was produced in *Pichia pastoris*. This mutant was produced in high yield, and controlled processing of the proenzyme by adding small amounts of active cathepsin L followed by HPLC was used to obtain milligram amounts of pure cathepsin L propeptide (phcl-1). This peptide is also truncated but is larger than the one produced in *S. cerevisiae*, containing 87 of the 96 amino acids of the full propeptide (i.e., residues 4p−90p). In addition, oxidation was found to be negligible for this peptide. Attempts to produce the cathepsin L propeptide directly in *Pichia pastoris* were unsuccessful.

Characterization of Inhibitory Activity. The inhibitory activity of propeptides phcl-1 to phcl-4 against cathepsin L and other cysteine proteases of the papain family is reported in Table 1. Inhibitor phcl-1, which is closest to the full propeptide, is a potent inhibitor of cathepsin L at pH 5.5 with a K_i of 0.088 nM. Even though this peptide is shorter than the intact propeptide by nine residues (three and six residues at the N- and C-terminal regions, respectively), it can certainly be used to evaluate the affinity and selectivity of the enzyme−propeptide interaction. In fact, it could be argued that pchl-1 is a better peptide to use for evaluating the inhibitory activity since it does not contain two identified cleavage sites for cathepsin L. Indeed, the "full length" propeptide is also a substrate for cathepsin L since it can be

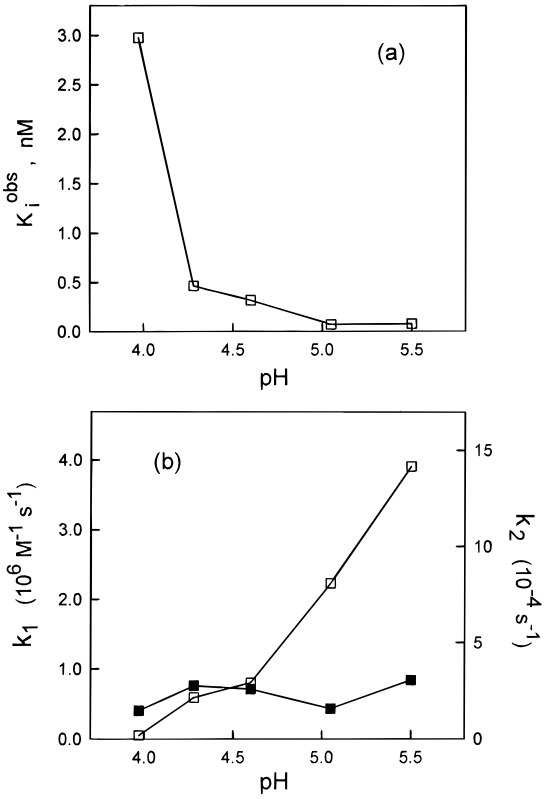


FIGURE 2: pH dependence of (a) K_i and (b) pre-steady-state parameters k_1 (□) and k_2 (■) for inhibition of cathepsin L by the peptide phcl-1.

hydrolyzed at positions 3p−4p and 90p−91p, as observed during the production of phcl-1 by processing of inactive procathepsin L by mature enzyme. It must be noted, however, that high enzyme concentration (relative to concentrations used in assays) and incubation for at least 1 h at 37 °C were necessary for processing to occur. HPLC experiments have been carried out with pchl-1 to confirm that hydrolysis of the peptide is not a problem under the assay conditions (data not shown). As observed with cathepsin B, the inhibition of cathepsin L by its propeptide is strongly pH-dependent: the affinity of propeptide phcl-1 for cathepsin L decreases significantly when the pH is lowered from 5.5 to 4.0 (Figure 2a). Despite the decrease in inhibitory activity at low pH, the phcl-1 propeptide is still a potent inhibitor of cathepsin L at pH 4, with a K_i of 3.0 nM. Inhibition was found to be selective for cathepsin L over other enzymes of the cathepsin L group. No inhibition of cathepsin B or papain could be observed up to 1 μM phcl-1. Inhibitory activity was detected against cathepsin S, but with a K_i of 44.6 nM, a value 510-fold lower than with cathepsin L. The peptide phcl-2 also displays high affinity for cathepsin L (K_i = 0.66 nM), but a significant increase in

K_i for cathepsin L inhibition is observed with the peptide phcl-3, where K_i = 11.5 nM, corresponding to a 130-fold decrease in inhibitory activity compared to phcl-1. It must be noted also that the preference for cathepsin L over cathepsin S is maintained with pchl-3. A smaller truncated version of the propeptide, containing residues 52p−95p (phcl-4), displays only weak inhibition of cathepsin L (Table 1).

For the most potent inhibitors (phcl-1, -2, and -3), time dependency of inhibition was observed, and pre-steady-state kinetic parameters (k_1 and k_2) could be measured and are reported in Table 1. The values obtained for the cathepsin L reaction with its propeptide are similar to those obtain with cathepsin B (Fox et al., 1992). The decrease in inhibitory activity observed for truncated peptides (phcl-2 and -3) can be attributed to the combined effect of a decrease in k_1 and an increase in k_2. For the pH-dependency of inhibition, the data reported in Figure 2b indicate that the variation in K_i with pH is mainly due to a variation in k_1, while k_2 remains virtually unchanged.

Circular Dichroism Measurements. The near-UV CD spectra of the isolated propeptide (phcl-1), cathepsin L (Cys25Ser/Thr110Ala), and the cathepsin L-propeptide complex (1:1 ratio) are presented in Figure 3a. In this region, aromatic residues contribute to CD signals when their side chains are submitted to asymmetric surroundings or involved in interactions with proteins. The fine dichroic bands between 280 and 300 nm indicate that tryptophan contributions are prominent in the CD spectra of cathepsin L and the propeptide−cathepsin L complex (Strickland, 1974). The more intense signal for the complex indicates that the contributions of aromatic residues of the propeptide or of aromatic residues at the surface of cathepsin L are stronger when the complex is formed. This suggests that hydrophobic interactions, within the propeptide and/or between the propeptide and surface residues on cathepsin L, are involved in the binding of the propeptide to the enzyme. In contrast, the spectrum of the propeptide alone is devoid of dichroic signal, indicating that aromatic amino acids are in symmetric surroundings and that the isolated propeptide is essentially devoid of compact tertiary structure.

The far-UV (peptide region) CD spectra of cathepsin L (Cys25Ser/Thr110Ala) and of the isolated propeptide pchl-1 are shown in Figure 3b. The spectrum of the isolated propeptide, with a minimum at 205 nm, indicates the presence of a large amount of random structure, although the intensities of the band at 191 nm and the shoulder at 220 nm suggest that ordered conformations (i.e., α-helix) contribute to the overall spectrum (Greenfield & Fasman, 1969). The addition of acetonitrile up to 40% (v/v) induces changes in the far-UV CD spectrum of isolated phcl-1 (increase of the band at 191 nm and decrease of the molar ellipticity at 220 nm) characteristic of an increase in α-helix content (data not shown). The CD spectrum of the complex is shown in Figure 3b. The shapes of the spectra for cathepsin L and the complex are similar, differing only in the intensities. Assuming no change in the cathepsin L structure occurs upon binding of the propeptide, the difference spectrum of the complex minus cathepsin L would correspond to the structure of the bound propeptide. Despite the noise in the 185−195 nm region, the subtracted CD spectrum (Figure 3b) and the isolated propeptide spectrum are relatively similar, suggesting that the propeptide bound to cathepsin L does not acquire significant ordered secondary structure.

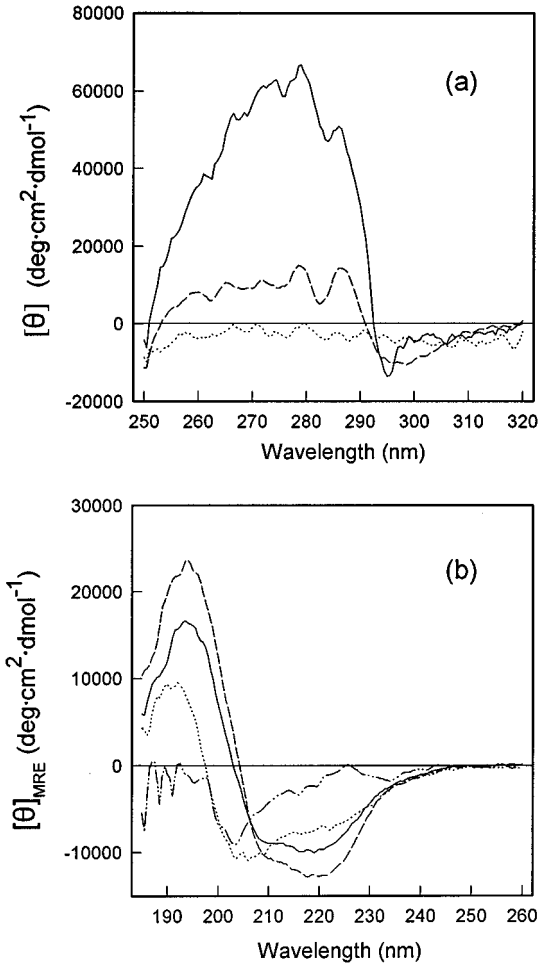


FIGURE 3: Circular dichroism spectra of the propeptide phcl-1 and of its complex with cathepsin L. (a) Near-UV CD spectra of cathepsin L Cys25Ser/Thr110Ala (− −), isolated phcl-1 (···), and the cathepsin L−phcl-1 complex (—). (b) Far-UV CD spectra of cathepsin L Cys25Ser/Thr110Ala (− −), isolated phcl-1 propeptide (···), and the cathepsin L−phcl-1 complex (—). The subtracted spectrum of the cathepsin L−phcl-1 complex minus cathepsin L is also shown (− ·· −). Conditions are as described under Materials and Methods.

DISCUSSION

The propeptide of human cathepsin L has been shown to be a potent inhibitor of the mature enzyme. This result, combined with the previous findings with rat cathepsin B as well as with the plant enzymes papain and PPIV, allows us to consider inhibition of cysteine proteases by their propeptides as a general feature of this family of enzymes. From the previous studies, however, the selectivity of the propeptides for their respective enzymes appeared unclear. The papain family of cysteine proteases can be divided into two subclasses: the cathepsin B-like enzymes and the others (including papain and cathepsin L). The main differences between the two subclasses are the insertions in cathepsin B, including the "occluding loop" (Musil et al., 1991), and the much shorter proregion in cathepsin B compared to other cysteine proteases. Because of these important differences, it was not surprising to find that the propeptide of cathepsin B does not significantly inhibit a member of the cathepsin L group of enzymes (i.e., papain) and is selective for

cathepsin B (Fox et al., 1992). Within the cathepsin L group of enzymes, however, recent results obtained for the propeptides of papain, caricain, and PPIV indicate a lack of selectivity against the plant enzymes (Taylor et al., 1995). The propeptide of papain has been shown to be a relatively good inhibitor of papain, with a K_i of 2 nM, but this peptide also inhibits caricain and chymopapain with similar values of K_i (8 nM and 12 nM, respectively). In addition, the propeptide of PPIV was also shown to be an inhibitor of the enzymes papain, chymopapain, and caricain with K_i values in the nanomolar range. The results obtained in the present study indicate that in addition to being a very potent inhibitor of cathepsin L ($K_i = 0.088$ nM), propeptide pchl-1 is also a selective inhibitor of cathepsin L over cathepsin S or papain. Selectivity can therefore be achieved within the cathepsin L group of enzymes.

The mammalian and plant proregions reported in Figure 1 overall display a 25% overall sequence homology. The selectivity of these propeptides is most probably not simply related to overall sequence homologies since possibly only a few residues could be crucial in determining selectivity. However, the lack of selectivity for the plant propeptides against plant enzymes reported by Taylor et al. (1985) may be explained by the high level of sequence homology (74% for the residues in Figure 1) between the propeptides of papain, caricain, and PPIV. In contrast, the four mammalian cathepsins in Figure 1 show 40% sequence homology (50% between human cathepins L and S), and the cathepsin L propeptide displays a 510-fold preference for cathepsin L over cathepsin S. In addition, the affinity of phcl-1 for cathepsin L is at least 11 000-fold higher than for papain or cathepsin B, indicating that selectivity within the mammalian enzymes (cathepsins L and S) is not as strong as between mammalian and plant enzymes, or between the cathepsin L and cathepsin B groups.

The sequences for mammalian propeptides represented in Figure 1 can be divided into three regions showing varying levels of homology. In the first 20 residues (1p−20p) and last 19 residues (78p−96p), the levels of homology within the mammalian group of propeptides are relatively low at 20% and 5%, respectively. The central region comprising residues 21p−77p displays a high level of homology (58%) and contains the previously reported ERFNIN and GNFD conserved motifs. The C-terminal region is highly variable for the propeptides of both the mammalian and plant cysteine proteases. Since selectivity is observed within the cathepsin L group (with the cathepsin L propeptide) but not with the plant propeptides, this region might not be important for inhibitor selectivity. The results obtained with peptide phcl-2, a relatively low value of K_i (0.66 nM) even though 15 residues are lacking at the C-terminal position of the propeptide, also suggest that this region contributes relatively little to overall inhibition. Contrary to the C-terminal region, the first 20 residues of the cathepsin L propeptide were found to be more important for inhibition of cathepsin L. The K_i for peptide phcl-3, which lacks the first 20 amino acids, is 11.5 nM, corresponding to a 130-fold decrease in affinity compared to peptide phcl-1 which lacks only 3 residues from the N-terminal. It must be noted that phcl-3 also contains five additional residues at the N-terminal position compared to phcl-1, but based on the result with phcl-2, we can assume that they do not significantly affect the comparison between phcl-1 and -3. Even though the decrease in affinity on going from phcl-1 to phcl-3 is significant, the first 20 residues cannot be considered as essential for the inhibitory activity of the propeptide. The N-terminal region of the cathepsin L proregion has been found to interact with a membrane protein (McIntyre & Erickson, 1993; McIntyre et al., 1994), and this portion of the proregion could therefore serve functions other than inhibition of the enzyme activity. It must be noted that contrary to our observations, cytotoxic T-lymphocytes antigen β (CTLA-2β), a peptide present in activated T-cells and mast cells and homologous to the proregion of mouse cathepsin L (Denizot et al., 1989), was shown to be a nonselective inhibitor of cathepsin L, papain, and cathepsin H with K_i values of 24, 25, and 67 nM, respectively (Delaria et al., 1994). Contrary to the cathepsin L propeptide, however, CTLA-2β possesses three free cysteine residues and can exist as a dimer or a disulfide-linked tetramer in solution (Delaria et al., 1994). CTLA-2β is therefore probably functionally distinct from the cathepsin L propeptide.

Inhibition of cathepsin L by its propeptide is pH-dependent. From pre-steady-state measurements, it was found that the variation in K_i can be attributed to an increase in k_1 when the pH is increased (Figure 2). The observation that only k_1 is pH-dependent is strongly indicative of a mechanism involving more than one step; i.e., the process does not correspond to the simple one-step mechanism described in Scheme 1. The pre-steady-state kinetic data can be fitted to the model of Scheme 1 because k_{obs} varies linearly with inhibitor concentration up to the highest concentrations used. However, in a multi-step process, the inhibitor might for example bind to the enzyme in an initial weak complex, followed by formation of a tighter enzyme−inhibitor complex. This initial complex would go undetected at the very low inhibitor concentrations used in the kinetic assays. In this model, the pH-dependency of k_1 could reflect the ionization of residues on the inhibitor prior to binding to the enzyme. From the shape of the pH-dependency curve, more than one residue would be involved, as previously noted with cathepsin B (Fox et al., 1992). It must be noted that with cathepsin B, k_2 and not k_1 was found to be pH-dependent. This again suggests that the mechanism is complex, or that the cathepsin B propeptide, which is distinct from that of cathepsin L, uses a different mechanism to mediate the pH regulation of inhibition. There are a number of ionizable groups on the propeptide and on the enzyme that could modulate the inhibition of cathepsin L by its propeptide. Positive identification of such residues will require more studies using enzyme and propeptide variants. Inhibition of cathepsin L by peptide phcl-3 was also pH-dependent (data not shown), indicating that the group(s) that participate(s) in the pH-dependency of inhibition is (are) not located in the first 20 residues of the propeptide. It must be remembered that processing of procathepsin L (and also procathepsin B) displays a pH-dependency very similar to that observed for the inhibition by propeptide (Smith & Gottesman, 1989; Rowan et al., 1992). In a previous study on processing of propapain, it was shown that alteration of the charge state in the GNFD motif of propapain could trigger processing and the GNFD motif was proposed to participate in the pH regulation of processing (Vernet et al., 1995). Residue Asp65p, which is conserved in both cathepsin L and cathepsin B groups of propeptides, might therefore play a role in the pH regulation of inhibition.

From the use of truncated propeptides, we have learned that the N-terminal region of the propeptide is more important for inhibition than the C-terminal region, which contributes very little. Structural information would be very beneficial to interpret these observations. In the absence of a crystal structure for the proenzyme or the propeptide, we have used circular dichroism measurements on the propeptide either free in solution or bound to cathepsin L. The data for the free propeptide indicate the presence of mainly random structure with some contribution from α-helix. It must be noted that the propeptide might exist in more than one conformation in equilibrium in solution (e.g., mainly random and mainly α-helix), and the CD spectrum would reflect the average contribution of these conformations. Upon binding to cathepsin L, the near-UV CD spectrum indicates that hydrophobic interactions, probably with Trp residues within the propeptide and/or between the propeptide and surface residues on cathepsin L, are involved in the binding of the propeptide to the enzyme. The human cathepsin L propeptide possesses four Trp residues, and three of these residues are conserved in the mammalian propeptides while the hydrophobic nature of this region is conserved in all sequences reported in Figure 1 (with the exception of Trp30p which is a Ser in human cathepsin O2). The far-UV (peptide region) CD spectrum of the complex between cathepsin L and its propeptide, however, suggests that the propeptide does not acquire significant ordered secondary structure on binding to cathepsin L. This result is surprising for many reasons. First, the presence of the ERFNIN motif, containing conserved residues separated by three amino acids, suggests that this motif is part of an α-helix as previously noted by Karrer et al. (1993). Secondary structure prediction programs also suggest the existence of α-helical components in the propeptide (data not shown), and addition of cosolvent (40% CH_3-CN) was shown to induce α-helix in the propeptide in solution. But the best argument against the absence of secondary structure comes from preliminary X-ray crystallography data on procathepsin L indicating that secondary structure is present in the proregion of procathepsin L (M. Cygler, personnal communication). It is not known at the moment if the propeptide bound to cathepsin L assumes the same structure as the proregion within procathepsin L. However, if that were the case, the difference CD spectrum between the cathepsin L−propeptide complex and cathepsin L in the far-UV region would be predicted to be quite different from the one represented in Figure 3b. It must be noted that aromatic side chains, particularly when present in clusters, can perturb the far-UV CD spectrum of a protein and make the estimation of secondary structure very difficult (Manning & Woody, 1989). The near-UV CD results indicate the presence of such clusters of aromatic residues in the cathepsin L−propeptide complex which might mask the presence of secondary structure elements in the far-UV CD spectrum. More work is needed to elucidate this aspect.

The proregion of cysteine proteases can regulate the activity of the zymogen and, when cleaved from the proenzyme, inhibit the mature enzyme. The observation of selective inhibition of cathepsin L over other members of the cathepsin L group of enzymes is an important finding with regard to future efforts in using cysteine protease propeptides as lead compounds for inhibitor development. It confirms that propeptides contain features ensuring that inhibition is highly selective for the proteases they originate from. By comparing the similarities and differences in the mechanism, activity, and structure between propeptides from various sources, we should be able to better understand the basis for their exquisite selectivity.

REFERENCES

Baker, D., Silen, J. L., & Agard, D. A. (1992) *Proteins: Struct., Funct., Genet. 12*, 339−344.

Berti, P. J., & Storer, A. C. (1995) *J. Mol. Biol. 246*, 273−283.

Brocklehurst, K., Willenbrock, F., & Salih, E. (1987) in *Hydrolytic Enzymes* (Neuberger, A., & Brocklehurst, K., Eds.) pp 39−158, Elsevier Biomedical Press, Amsterdam.

Brömme, D., & Okamoto, K. (1995) *Biol. Chem. Hoppe-Seyler 376*, 379−384.

Brömme, D., Bonneau, P. R., Lachance, P., Wiederanders, B., Kirschke, H., Peters, C., Thomas, D. Y., Storer, A. C., & Vernet, T. (1993) *J. Biol. Chem. 268*, 4832−4838.

Bryan, P., Wang, L., Hoskins, J., Ruvinov, S., Strausberg, S., Alexander, P., Almog, O., Gilliland, G., & Gallagher, T. (1995) *Biochemistry 34*, 10310−10318.

Cha, S. (1975) *Biochem. Pharmacol. 24*, 2177−2185.

Cohen, L. W., Coghlan, V. M., & Dihel, L. C. (1986) *Gene 48*, 219−227.

Delaria, K., Fiorentino, L., Wallace, L., Tamburini, P., Brownell, E., & Muller, D. (1994) *J. Biol. Chem. 269*, 25172−25177.

Demuth, H.-U. (1990) *J. Enzyme Inhib. 3*, 249−278.

Denizot, F., Brunet, J.-F., Roustan, P., Harper, K., Suzan, M., Luciani, M.-F., Mattei, M.-G., & Golstein, P. (1989) *Eur. J. Immunol. 19*, 631−635.

Dixon, M. (1953) *Biochem. J. 55*, 170−171.

Duffy, M. J. (1992) *Clin. Exp. Metastasis 10*, 145−155.

Ernst, J. F. (1986) *DNA 5*, 483−491.

Esser, R. E., Angelo, R. A., Murphey, M. D., Watts, L. M., Thornburg, L. P., Palmer, J. T., Talhouk, J. W., & Smith, R. E. (1994) *Arthritis Rheum. 37*, 236−247.

Fox, T., de Miguel, E., Mort, J. S., & Storer, A. C. (1992) *Biochemistry 31*, 12571−12576.

Fusek, M., Mares, M., Vagner, J., Voburka, Z., & Baudys, M. (1991) *FEBS Lett. 287*, 160−162.

Gal, S., & Gottesman, M. M. (1988) *Biochem. J. 253*, 303−306.

George, D. G., Barker, W. C., & Hunt, L. T. (1990) *Methods Enzymol. 183*, 333−351.

Gill, S. C., & von Hippel, P. H. (1989) *Anal. Biochem. 182*, 319−326.

Greenfield, N., & Fasman, G. D. (1969) *Biochemistry 8*, 4108−4116.

Inaoka, T., Bilbe, Ishibashi, O., Tezuka, K.-I., Kumegawa, M., & Kokubo, T. (1995) *Biochem. Biophys. Res. Commun. 206*, 89−96.

Ishidoh, K., Imajoh, S., Emori, Y., Ohno, S., Kawasaki, H., Minami, Y., Kominami, E., Katunuma, N., & Suzuki, K. (1987a) *FEBS Lett. 226*, 33−37.

Ishidoh, K., Towatari, T., Imajoh, S., Kawasaki, H., Kominami, E., Katunuma, N., & Suzuki, K. (1987b) *FEBS Lett. 223*, 69−73.

Izquierdo-Martin, M., & Stein, R. L. (1992) *J. Am. Chem. Soc. 114*, 1527−1528.

Karrer, K. M., Peiffer, S. L., & DiTomas, M. E. (1993) *Proc. Natl. Acad. Sci. U.S.A. 90*, 3063−3067.

Katunuma, N., & Kominami, E. (1987) *Rev. Physiol. Biochem. Pharmacol. 108*, 1−20.

Kunkel, T. A. (1985) *Proc. Natl. Acad. Sci. U.S.A. 82*, 488−492.

Liao, P.-C., Leykam, J., Andrews, P. C., Gage, D. A., & Allison, J. (1994) *Anal. Biochem. 219*, 9−20.

Manning, M. C., & Moody, R. W. (1989) *Biochemistry 28*, 8609−8613.

McIntyre, G. F., & Erickson, A. H. (1993) *Proc. Natl. Acad. Sci. U.S.A. 90*, 10588−10592.

McIntyre, G. F., Godbold, G. D., & Erickson, A. H. (1994) *J. Biol. Chem. 269*, 567−572.

Ménard, R., Khouri, H. E., Plouffe, C., Dupras, R., Rippoll, D., Vernet, T., Tessier, D. C., Laliberté, F., Thomas, D. Y., & Storer, A. C. (1990) *Biochemistry 29*, 6706−6713.

Musil, D., Zucic, D., Turk, D., Engh, R. A., Mayr, I., Huber, R., Popovic, T., Turk, V., Towatari, T., Katunuma, N., & Bode, W. (1991) *EMBO J. 10*, 2321−2330.

Revell, D. F., Cummings, N. J., Baker, K. C., Collins, M. E., Taylor, M. A. J., Sumner, I. G., Pickersgill, R. W., Connerton, I. F., & Goodenough, P. W. (1993) *Gene 127*, 221−225.

Rich, D. H. (1986) in *Proteinase Inhibitors* (Barrett, A. J., & Salvesen, G. S., Eds.) pp 153−178, Elsevier Science Publishers, B. V., Amsterdam.

Rowan, A. D., Mason, P., Mach, L., & Mort, J. S. (1992) *J. Biol. Chem. 267*, 15993−15999.

San Segundo, B., Martinez, M. C., Vilanova, M., Cuchillo, C. M., & Avilés, F. X. (1982) *Biochim. Biophys. Acta 707*, 74−80.

Shaw, E. (1990) *Adv. Enzymol. Relat. Areas Mol. Biol. 63*, 271−347.

Sheahan, K., Shuja, S., & Murnane, M. J. (1989) *Cancer Res. 49*, 3809−3814.

Sherman, F., Fink, G. R., & Hicks, J. B. (1986) in *Methods in Yeast Genetics*, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.

Shi, G.-P., Munger, J. S., Meara, J. P., Rich, D. H., & Chapman, H. A. (1992) *J. Biol. Chem. 267*, 7258−7262.

Shi, G.-P., Chapman, H. A., Bhairi, S. M., DeLeeuw, C., Reddy, V. Y., & Weiss, S. J. (1995) *FEBS Lett. 357*, 129−134.

Shwartz, R. M., & Dayhoff, M. O. (1979) in *Atlas of Protein Sequence and Structure* (Dayhoff, M. O., Ed.) Vol. 5, Suppl. 3, p 353, National Biomedical Research Foundation, Washington.

Smith, S. M., & Gottesman, M. M. (1989) *J. Biol. Chem. 264*, 20487−29495.

Strickland, E. H. (1974) *Crit. Rev. Biochem*. 113−175.

Taylor, M. A. J., Baker, K. C., Briggs, G. S., Connerton, I. F., Cummings, N. J., Pratt, K. A., Revell, D. F., Freedman, R. B., & Goodenough, P. W. (1995) *Protein Eng. 8*, 59−62.

Trabandt, A., Gay, R. E., Fassbender, H. G., & Gay, S. (1991) *Arthritis Rheum. 34*, 1444−1451.

Vernet, T., Berti, P. J., de Montigny, C., Musil, R., Tessier, D. C., Ménard, R., Magny, M.-C., Storer, A. C., & Thomas, D. Y. (1995) *J. Biol. Chem. 270*, 10838−10846.

Wiederanders, B., Brömme, D., Kirschke, H., vonFigura, K., Schmidt, B., & Peters, C. (1992) *J. Biol. Chem. 267*, 13708−13713.

Williams, J. W., & Morrison, J. F. (1979) *Methods Enyzmol. 63*, 437−467.

BI952736S