Available online at www.sciencedirect.com
SCIENCE DIRECT®
Analytical Biochemistry 335 (2004) 244–252
ANALYTICAL BIOCHEMISTRY
www.elsevier.com/locate/yabio

Positional-scanning combinatorial libraries of fluorescence resonance energy transfer peptides to define substrate specificity of carboxydipeptidases: assays with human cathepsin B

Simone Silva Cotrin, Luciano Puzer, Wagner Alves de Souza Judice, Luiz Juliano*, Adriana K. Carmona, Maria Aparecida Juliano*

Department of Biophysics, Escola Paulista de Medicina, UNIFESP, Rua Três de Maio, 100, São Paulo 04044-020, Brazil

Received 10 July 2004

Abstract

We have developed positional scanning synthetic combinatorial libraries to define the substrate specificity of carboxydipeptidases. The library Abz-GXXZXK(Dnp)-OH, where Abz is *ortho*-aminobenzoic acid, K(Dnp) is N^{ε}-2,4-dinitrophenyl-lysine with free carboxyl group, the Z position was successively occupied with 1 of 19 amino acids (cysteine was omitted), and X represents randomly incorporated residues, was assayed initially with human cathepsin B, and arginine was defined as one of the best residues at the P_1 position. To examine the selectivity of S_1', S_2, and S_3 subsites, the sublibraries Abz-GXXRZK(Dnp)-OH, Abz-GXZRXK(Dnp)-OH, and Abz-GZXRXK(Dnp)-OH were then synthesized. The peptide Abz-GIVRAK(Dnp)-OH, which contains the most favorable residues in the P_3–P_1' positions identified by screening of the libraries with cathepsin B, was hydrolyzed by this enzyme with $k_{cat}/K_m = 7288\ mM^{-1}\ s^{-1}$. This peptide is the most efficient substrate described for cathepsin B to this point, and it is highly selective for the enzyme among the lysosomal cysteine proteases.

Keywords: Cysteine peptidase; Lysosome; Amino acids; Exopeptidase

Combinatorial peptide library approaches have been used as an alternative to the systematic assays of individual peptides for specificity determination of endoproteases. All of the combinatorial peptide library methods involve their generation followed by screening to identify the optimal substrate sequences. Several methodologies have been described and can be obtained by biological or synthetic procedures. The biological method involves the display of the peptide libraries on filamentous phage, the identification of the best substrates by molecular biology tools, and the cleavage site determination in each substrate [1–4]. Synthetic peptide libraries can contain millions of compounds, but the identification of the substrates and the cleavage site require robust analytical assays such as Edman degradation [5–7], mass spectrometry [8,9], and chromatography [10]. Support-bound fluorescence resonance energy transfer (FRET)[1] peptide libraries have been prepared

* Corresponding authors. Fax: +55 11 5575 9617.
E-mail address: juliano.biof@epm.br (M.A. Juliano).

[1] *Abbreviations used:* FRET, fluorescence resonance energy transfer; PS–SCL, positional scanning synthetic combinatorial library; AMC, [7-amino-4-methyl]coumarin; ACC, [7-amino-4-carbamoylmethyl]coumarin; Dnp, 2,4-dinitrophenyl; Abz, *ortho*-aminobenzoic acid; K(Dnp), N^{ε}-2,4-dinitrophenyl-lysine; Pmc, 2,2,5,7,8-pentamethyl-chlromam-6-sulfonyl; TFA, trifluoroacetic acid; DMF, dimethylformamide; DCM, dichloromethane; TBTU, 2-(1H-bezotriazole-1-yl)-1,1,3,3-tetramethyluronium tetrafluoroborate; HOBT, *N*-hydroxybenzotriazole; EDT, 1,2-ethanodithiol; HPLC, high-performance liquid chromatography; ACN, acetonitrile; MALDI–TOF, matrix-assisted laser desorption/ionization–time of flight; DTE, dithioerythritol; DMSO, dimethyl sulfoxide; EDTA, ethylenediaminetetraacetic acid.

0003-2697/$ - see front matter © 2004 Elsevier Inc. All rights reserved.
doi:10.1016/j.ab.2004.09.012

by the process of split synthesis, resulting in a single peptide sequence on each of the resin beads [11]. Partial proteolysis of the substrates bound to the solid support and sequence determination by Edman degradation of the substrates on the most fluorescent beads provides the sequences and the cleavage sites of the optimal substrates [12,13].

The use of positional scanning synthetic combinatorial libraries (PS–SCLs) of fluorogenic peptide substrates is a strategy where each position in the peptide sequence is occupied in turn by a single amino acid residue. Mixtures of amino acids must be coupled to mixtures of resin-bound amino acids for the synthesis of the libraries. The earlier described PS–SCLs were formed by [7-amino-4-methyl]coumarin (AMC)- or [7-amino-4-carbamoylmethyl]coumarin (ACC)-containing fluorogenic peptides having the C-terminal carboxyl group occupied by the fluorogenic AMC or ACC group. Besides some intrinsic limitations of the synthetic process due to nonequimolar or incomplete libraries, the reported results using this approach seem to be very convincing [14–17].

All of the above peptide libraries for substrates for proteases were designed to investigate their endopeptidase specificities. There are several peptidases of biological interest that are predominantly carboxydipeptidase such as angiotensin I-converting enzyme and cathepsin B at acid pH, which are present in mammals and in other lower living organisms. In the current study, we synthesized a PS–SCL specially designed for screening peptidases with carboxydipeptidase activity that contain peptides with a free C-terminal carboxyl group. Initially, we prepared a library with the general structure Abz-GXXZXK(Dnp [2,4-dinitrophenyl])-OH, where the Z position was successively occupied with 1 of 19 amino acids (cysteine was omitted to avoid dimerization) and X represents randomly incorporated residues. To ensure equal coupling of the randomized residues, a balanced isokinetic mixture of 19 amino acids was used in the synthesis procedure as described previously [18]. Assaying this library, we could define the enzyme preference for the P_1 position (according to the nomenclature of Schechter and Berger [19]) of the substrate because the carboxydipeptidase activity will remove the dipeptide XK(Dnp)-OH that contains the fluorescence quencher group (Dnp), resulting in fluorescence increasing. This library was assayed with human cathepsin B, and arginine was defined as one of the best residues in the P_1 position of the substrates. Then, to examine the selectivity of S'_1, S_2, and S_3 subsites of this enzyme, the sublibraries Abz-GXXRZK(Dnp)-OH, Abz-GXZRXK(Dnp)-OH, and Abz-GZXRXK(Dnp)-OH were synthesized, with arginine being fixed at the P_1 position. The presence of arginine in all sublibraries renders them more susceptible to cathepsin B and provides more solubility in water of the generated peptides.

Cathepsin B is significantly larger than papain (245 vs. 212 residues) due to insertions distributed throughout the molecule [20]. Particularly characteristic is the presence of the occluding loop in the catalytic cleft of cathepsin B, which contains His110 and His111 and forms its S'_2 subsite. His110 interacts with Asp22 through ion pair stabilizing electrostatic interaction, and His111 is the major determinant of the exopeptidase activity of cathepsin B by virtue of its electrostatic interaction with the free carboxylate group of the substrates [21–27]. The occluding loop forms a wall impairing the normal threading of substrate through the active site cleft [20], resulting in greater carboxydipeptidase activity than endopeptidase activity of cathepsin B, particularly on synthetic peptides. The substrate specificity for these two activities of cathepsin B was described in detail previously using a series of isolated FRET peptides modified systematically in each position [22,27,28]. Given all of this information available in the literature, cathepsin B is a very convenient protease to assay a PS–SCL of fluorogenic peptide substrates designed for carboxydipeptidases. In addition, more information will be obtained on this important lysosomal cysteine protease that is known to be involved in a variety of physiological and pathological conditions [29,30], particularly in tumor invasion and metastasis [31–34], apoptosis [35–37], and angiogenesis [38].

Materials and methods

Chemicals were obtained from commercial suppliers and used without further purification. The Fmoc amino acids, Fmoc-Ala-OH, Fmoc-Arg(Pmc [2,2,5,7,8-pentamethyl-chlromam-6-sulfonyl])-OH, Fmoc-Asn(Trt)-OH, Fmoc-Asp(O-t-Bu)-OH, Fmoc-Glu(O-t-Bu)-OH, Fmoc-Gln(Trt)-OH, Fmoc-Gly-OH, Fmoc-His(Trt)-OH, Fmoc-Ile-OH, Fmoc-Leu-OH, Fmoc-Lys(Boc)-OH, Fmoc-Phe-OH, Fmoc-Pro-OH, Fmoc-Ser(O-t-Bu)-OH, Fmoc-Thr(O-t-Bu)-OH, Fmoc-Trp(Boc)-OH, Fmoc-Tyr(O-t-Bu)-OH, Fmoc-Val-OH, Fmoc-Met, Fmoc-Lys(Dnp)-OH, and NovaSyn-TGA resin, which contains the trifluoroacetic acid (TFA) labile 4-hydroxymethylphenoxyacetic acid linker, were purchased from Novabiochem (San Diego, CA, USA). Dimethylformamide (DMF) with low amine content was used as a solvent for all syntheses.

Enzymes

Human recombinant cathepsin B was obtained as described previously [21]. Cruzain, truncated in the C-terminal extension, was obtained from *Escherichia coli* (strain DH5*a* containing the expression plasmid) kindly supplied by McKerrow (University of California, San Francisco) following the procedure reported previously

[39]. Recombinant human cathepsins K, V, and L were expressed in the yeast *Pichia pastoris* as described previously [21]. Human recombinant cathepsin X was expressed in *P. pastoris* and was activated using cathepsin L as described elsewhere [40]. In the final purification step, cathepsin X was separated from cathepsin L by ion exchange chromatography on SP Sepharose Fast Flow (Amersham–Pharmacia Biotech, Uppsala, Sweden) at pH 4.4. The molar concentrations of all active cathepsins were determined by titration using the irreversible inhibitor E-64 [41]. CPB 2.8 ΔCTE, a recombinant cysteine protease of *Leishmania mexicana*, was obtained as described earlier [42].

Synthesis of positional scanning libraries of FRET peptides

Synthesis of Fmoc-Lys(Dnp)-NovaSyn-TGA resin

Fmoc-Lys(Dnp)-OH (9 mmol), 1-(mesitylene-2-sulfonyl)-3-nitro-1,2,4-triazole (8 mmol), and 1-methyl-imidazole (5 mmol) dissolved in 5 ml anhydrous dichloromethane (DCM) were added to 13.5 g NovaSyn-TGA resin (0.25 mmol/g) in restricted anhydrous conditions. The mixture was kept at room temperature with gentle shaking for 12 h, and then the resin was washed with DCM and DMF and the Fmoc-Lys(Dnp)-OH loading was determined by quantification of Fmoc released by treatment of 1 mg of the resin with 20% piperidine in DMF followed by absorbance determination at $\lambda = 305$ nm (extinction coefficient 6410 M^{-1} cm^{-1}).

Synthesis of library Abz-GXXZXK(Dnp)-OH

The first randomized coupling (X) was done in the free α-amino of Lys(Dnp)-NovaSyn-TGA resin using a 10-fold excess of an isokinetic mixture of Fmoc amino acids [43] with the same excess of 2-(1H-bezotriazole-1-yl)-1,1,3,3-tetramethyluronium tetrafluoroborate (TBTU) and *N*-hydroxybenzotriazole (HOBT) as coupling reagents. A 20-fold excess of *N*-methylmorphorline was added as the organic base for this reaction. The coupling of the fixed amino acid (Z) was done by distributing 0.1 mmol of Fmoc-XK(Dnp)-resin in 19 reaction vessels, and the synthesis was continued by coupling in each vessel 1 of the 19 Fmoc amino acids using the same excess of reagents (cysteine was excluded). The next couplings of randomized isokinetic mixtures of amino acids, as well as glycine (which was a common amino acid in all mixtures), were done in each of 19 separate vessels. The Abz was added as Boc-Abz, and the mixtures of each sublibrary were removed from the resin in each vessel by treatment with TFA–1,2-ethanodithiol (EDT)–Anisol–H_2O (92.0:0.2:0.8:7.0) for 8 h. The resin was filtered and washed with 95% acetic acid, and the resulting liquid mixtures were lyophilized three times, redissolving in H_2O between lyophilization steps.

Synthesis of the libraries Abz-GXZRXK(Dnp)-OH, Abz-GZXRXK(Dnp)-OH, and Abz-GXXRZK(Dnp)-OH

The libraries Abz-GXZRXK(Dnp)-OH and Abz-GZXRXK(Dnp)-OH were essentially synthesized as described above except that arginine and glycine were present in all mixtures of sublibraries. The Abz-GXXRZK(Dnp)-OH library was synthesized in 19 separate vessels due to the position of the fixed amino acid (Z) that is just before Lys(Dnp).

Synthesis and purification of single fluorescent Abz peptides and FRET peptides

The peptides were synthesized by solid phase using Fmoc methodology. All syntheses were done in an automated bench-top simultaneous multiple solid phase peptide synthesizer (PSSM 8 system from Shimadzu, Tokyo, Japan). The final unprotected peptides were purified by semipreparative high-performance liquid chromatography (HPLC) using an Econosil C-18 column (10 μm, 22.5 × 250 mm) and a two-solvent system: TFA/H_2O (1:1000, v/v) and TFA/acetonitrile [ACN]/H_2O (1:900:100, v/v/v). The column was eluted at a flow rate of 5 ml min^{-1} with a 10, 30–50, or 60% gradient of solvent B over 30 or 45 min. Analytical HPLC was performed using a binary HPLC system from Shimadzu with an SPD-10AV Shimadzu UV/vis detector and a Shimadzu RF-535 fluorescence detector coupled to an Ultrasphere C-18 column (5 μm, 4.6 × 150 mm), which was eluted with solvent systems A1 (H_3PO_4/H_2O, 1:1000, v/v) and B1 (ACN/H_2O/H_3PO_4, 900:100:1, v/v/v) at a flow rate of 1.7 ml min^{-1} and a 10–80% gradient of B1 over 15 min. The HPLC column-eluted materials were monitored by their absorbance at 220 nm and by fluorescence emission at 420 nm following excitation at 320 nm for the peptides containing Abz–Dnp pair. The molecular mass and purity of synthesized peptides were checked by matrix-assisted laser desorption/ionization–time of flight (MALDI–TOF) mass spectrometry (TofSpec-E, Micromass) and/or peptide sequencing using a protein sequencer PPSQ-23 (Shimadzu).

Peptide library screen

The enzymatic hydrolysis by cathepsin B of the FRET peptides was done in 100 mM sodium acetate containing 200 mM NaCl, pH 4.5, at 37 °C in a final volume of 1 ml. The enzyme (1–10 nM) was preactivated in the presence of 2.5 mM dithioerythritol (DTE) for 5 min before the addition of the substrates. Substrates from the PS–SCLs were dissolved in dimethyl sulfoxide (DMSO), and the concentrations of the library solutions were determined by colorimetric determination of the 2,4-dinitrophenyl group, where $\varepsilon_{365} = 17{,}300$ M^{-1} cm^{-1}. The concentrations of library solutions were calibrated so as to perform all of the assays at 100 nM. The reac-

tion was initiated by the addition of the library solution, and the hydrolysis of the FRET peptides was quantified using a Hitachi F-2500 spectrofluorometer by measuring the fluorescence of Abz at 420 nm following excitation at 320 nm. The reactions were followed over 300 s, and the initial linear increment of fluorescence with time was taken as the velocity of hydrolysis. In this condition, the hydrolysis is proportional to the specificity constant k_{cat}/K_m [17] because similar preferences for hydrolysis were observed varying the library concentration (50 and 200 nM).

Determination of the cleaved peptide bonds

The peptide Abz-GIVRAK(Dnp)-OH was tested as substrate for cathepsins B, K, L, V, and X and cruzain. The cleaved bonds were identified by isolation of the digestion fragments by reverse-phase HPLC on a C18 column equilibrated in 10% solvent B (90% ACN, 10% TFA, v/v). The column was eluted at a flow rate of 1 ml min^{-1} with a 10–80% gradient of solvent B over 28 min. The elution profile was determined at 220 nm and by fluorescence at 420 nm after excitation at 320 nm. The Abz-containing fragments were compared with authentic synthetic sequences and/or by amino acid sequencing, and molecular mass determination was made by MALDI–TOF mass spectrometry.

pH dependence of the cathepsin B activity toward the substrate Abz-GIVRAK(Dnp)-OH

The pH profile of cathepsin B activity was obtained by measuring the k_{cat}/K_m parameters of hydrolysis of the peptide Abz-GIVRAK(Dnp)-OH under pseudo first-order conditions. The determinations were carried out in a standard buffer containing 25 mM acetic acid, 25 mM Mes, 75 mM Tris, 25 mM glycine, and 2 mM ethylenediaminetetraacetic acid (EDTA). The pH ranges were 3.5–10.0 and were adjusted with 1 M NaOH and HCl. The enzyme was preactivated with 2.5 mM DTE at 37 °C for 5 min.

Results

Screening of Abz-GXXZXK(Dnp)-OH library by human cathepsin B

Fig. 1 shows the activity of human cathepsin B on the Abz-GXXZXK(Dnp)-OH library. Arg was the preferred amino acid at the P_1 position, but the enzyme also well accepted the hydrophobic residue Leu and amino acids with short side chain, such as Ser and Ala, or even without side chain, such as Gly. This profile of preferences is very similar to those found previously with a

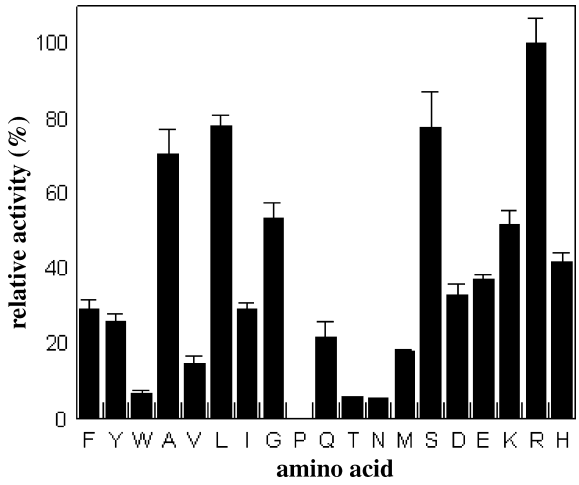

Fig. 1. P_1 substrate specificity of human cathepsin B. The positional scanning fluorimetric combinatorial library with general sequence Abz-GXXZXK(Dnp)-OH screened with human cathepsin B is shown. First, 100 nM of mixture with fixed amino acid (Z) was incubated with cathepsin B (5–10 nM) in 100 mM sodium acetate containing 200 mM NaCl, pH 4.5, at 37 °C. The enzyme was preactivated for 5 min with 2.5 mM DTE. The assays were determined measuring the initial velocity of each run. The *y* axis represents the relative activity obtained as described in Materials and methods. The *x* axis shows the specific amino acids represented by the one-letter codes. The errors are shown at the tops of the bars, each of which represents the average of three determinations.

series of peptides derived from Dnp-GFXFW-OH, where X is natural amino acids [28]. A limited support-bound FRET peptide library composed of nine amino acids (Gly, Ala, Ser, Thr, Asn, Gln, Val, Phe, and Tyr) also showed that Gly-containing peptides were the more frequent peptides selected for the endopeptidase activity of cathepsin B [44].

Screening of Abz-GXZRXK(Dnp)-OH, Abz-GZXRXK(Dnp)-OH, and Abz-GXXRZK(Dnp)-OH sublibraries by human cathepsin B

The Abz-GXZRXK(Dnp)-OH sublibrary screening provides the preferences of cathepsin B for the amino acids at the P_2 position of the substrates. The middle panel of Fig. 2 (P_2) shows the activity of cathepsin B on this sublibrary. This screening is particularly relevant because the primary specificity of papain-like proteases is largely determined by the S_2–P_2 interaction [45]. It is noteworthy that Val is the best choice of the enzyme among the amino acids with aliphatic side chains and that the susceptibility of the peptides with Val was similar to that with Phe, which is also well accepted by cathepsin B, as shown in the middle panel of Fig. 2 and in previous observations [28,46]. This library also demonstrated that cathepsin B accommodated well the basic amino acids Arg, Lys, and His (in this order of preference). The presence of the negatively charged amino acids Asp and Glu resulted in poor hydrolysis, and

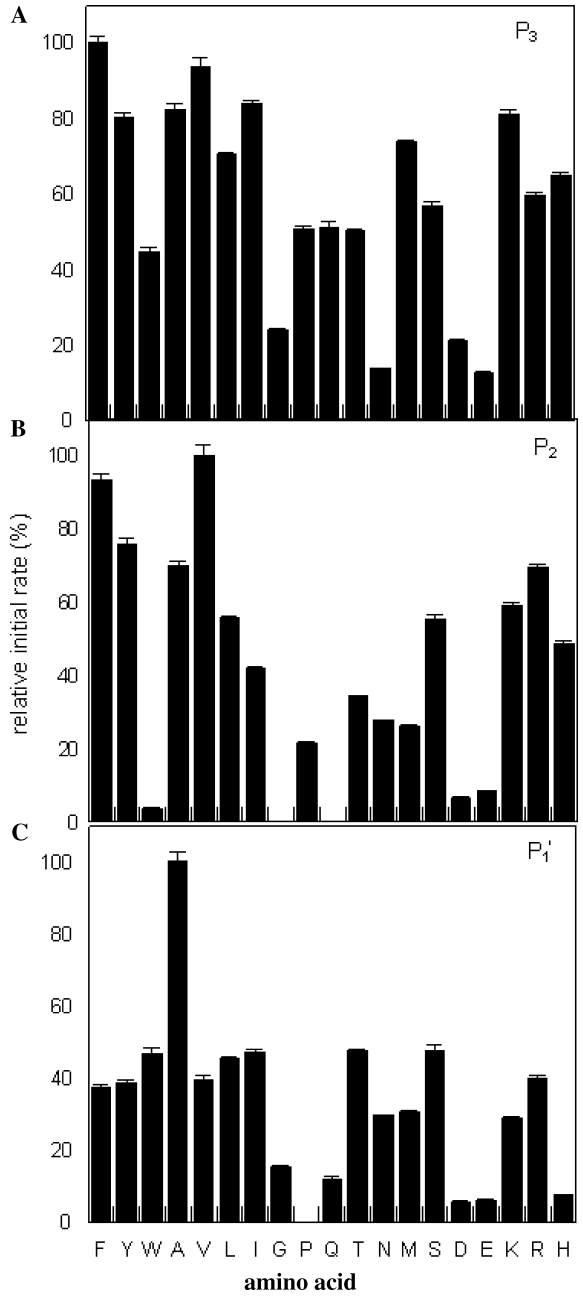

Fig. 2. P_2, P_3, and P'_1 substrate specificity of human cathepsin B. The positional scanning fluorimetric combinatorial libraries with the general sequences Abz-GXZRXK(Dnp)-OH (middle panel, P_2), Abz-GZXRXK(Dnp)-OH (top panel, P_3), and Abz-GXXRZK(Dnp)-OH (bottom panel, P'_1) are shown. The conditions are similar to those described in Fig. 1.

the presence of Gly and Gln resulted in no hydrolysis. Similar results were observed previously with substrates based on Dnp-GXRFW-OH containing these amino acids [28].

Screening of the Abz-GZXRXK(Dnp)-OH sublibrary provided the preferences of cathepsin B for the amino acids at the P_3 position of the substrates. The top panel of Fig. 2 (P_3) shows the activity of cathepsin B on this sublibrary. Ile was the preferred amino acid at the P_3 position, followed by Lys and Phe. The preference of the S_3 subsite of cathepsin B for aromatic residues was reported previously [47]. The worst amino acids at the P_3 position are Asn and the negatively charged residues Glu and Asp. All of these results suggest that the S_3 subsite has, in addition to a hydrophobic pocket, a positive electrostatic environment.

Abz-GXXRZK(Dnp)-OH sublibrary screening provides the preferences of cathepsin B for the amino acids at the P'_1 position of the substrates. The bottom panel of Fig. 2 (P'_1) shows the activity of cathepsin B on this sublibrary. The greater preference of the S'_1 subsite of cathepsin B for hydrophobic residues, with either aliphatic or aromatic structure, is very clear. These data are in accordance with earlier observations with substrates that permit endopeptidase activity [48] or carboxydipeptidase activity [28]. The complete resistance of the peptides containing Pro in this sublibrary is noteworthy due to the imide nature of its peptide bond with the P_1 amino acid. This result gives significant support to the initial assumption that the libraries described in this work would be cleaved only after the antipenultimate amino acid.

Cathepsin B carboxydipeptidase activity in all libraries

To verify whether cathepsin B could present carboxymonopeptidase activity on the above-described PS–SCLs, removing only Lys(Dnp)-OH from the peptides, all of the sublibraries were incubated for 24 h with 20 nM of cathepsin B, and Lys(Dnp)-OH was not detected in examining the reaction mixtures by HPLC. In addition, we verified the potentiality of cathepsin B carboxymonopeptidase activity, looking for its activity on the peptides Abz-FRFW-OH, Abz-FFFR-OH, and Abz-FFRW-OH as well as on the same peptides with their C-terminal carboxyl group amidated. These peptides were designed to put Phe and Arg, the best amino acids for the P_2 position for cathepsin B activity, at the penultimate site of these peptides; however, cathepsin B removed the C-terminal dipeptides from all peptides except Abz-FFRW-NH_2. This peptide was cleaved at the R-W bond releasing Trp-NH_2. Therefore, the free C-terminal carboxyl group directs the carboxydipeptidase activity of cathepsin B.

Kinetics of hydrolysis and its pH profile of a peptide designed from library data

The peptide Abz-GIVRAK(Dnp)-OH was synthesized based on the residue at each position, from P_3 to P'_1, that resulted in the highest hydrolytic activity of each library or sublibrary screened by cathepsin B. Abz-GIVRAK(Dnp)-OH was hydrolyzed only at the Arg-Ala peptide bond with the highest catalytic efficiency (k_{cat}/

Table 1
Kinetic constants for hydrolysis by cathepsin B and cruzain of substrates Abz-peptidil(Dnp)-OH designed from screening of combinatorial libraries

	Abz-GIVRAK(Dnp)-OH				Abz-GIVRPK(Dnp)-OH			
	K_m (μM)	k_{cat} (s^{-1})	k_{cat}/K_m ($mM^{-1}\ s^{-1}$)	Cleavage site(s)	K_m (μM)	k_{cat} (s^{-1})	k_{cat}/K_m ($mM^{-1}\ s^{-1}$)	Cleavage site
Cathepsin B	5.9	43	7288	R ↓ A 100%	Resistant			
Cruzain	0.8	0.06	75	V ↓ R 39% R ↓ A 61%	0.06	0.02	333	V ↓ R 100%

Note. The assays were performed in 100 mM acetate buffer containing 200 mM NaCl, pH 4.5, at 37 °C. The enzymes were activated for 5 min with 2.5 mM DTE. Measurements were done as described in Materials and methods. Cleavage sites are indicated as arrows (↓). The standard deviations of the kinetic constants were less than 5%. The assays were performed under pseudo first-order conditions.

K_m) described for cathepsin B to this point. The catalytic constant (k_{cat}) made the main contribution to the obtained k_{cat}/K_m value (Table 1). The analog peptide Abz-GIVRPK(Dnp)-OH, with Pro at the P_1' position, was also synthesized and was completely resistant to hydrolysis by cathepsin B, even at a concentration 10 times higher than that used to hydrolyze Abz-GIVRAK(Dnp)-OH. Cruzain is a recombinant form of the major cysteine protease of *Trypanosoma cruzi* and accepts Arg at the P_2 position [49], and it has carboxydipeptidase activity similar to that of cathepsin B [50]. Cruzain was also assayed with Abz-GIVRAK(Dnp)-OH and Abz-GIVRPK(Dnp)-OH, and it hydrolyzed Abz-GIVRPK(Dnp)-OH at the Val-Arg bond, working on this substrate as an endopeptidase, and poorly hydrolyzed the peptide Abz-GIVRAK(Dnp)-OH at the Val-Arg and Arg-Ala bonds (Table 1). In addition, the mixture of peptides of sublibrary Abz-GXXRZK(Dnp)-OH (where Z is Pro), which was resistant to cathepsin B, was significantly hydrolyzed by cruzain. These results confirm the exclusive carboxydipeptidase activity of cathepsin B on the peptide library described in this article.

To investigate the specificity of cathepsin B toward the peptide Abz-GIVRAK(Dnp)-OH, we assayed this substrate with the other cysteine proteases such as human cathepsins K, L, V, and X and with CPB 2.8 ΔCTE, which is a recombinant form of the most highly expressed cysteine protease of *L. mexicana*. Table 2 shows that the k_{cat}/K_m values for these enzymes are one or two orders of magnitude lower than those for cathepsin B and that their cleavage site differs among the enzymes, with some of them being cleaved at two peptide bonds. Therefore, the peptide Abz-GIVRAK(Dnp)-OH has a high specificity for cathepsin B compared with other cysteine proteases.

The pH profile of Abz-GIVRAK(Dnp)-OH hydrolysis by cathepsin B is shown in Fig. 3. The values $pK_1 = 3.45$ and $pK_2 = 5.50$ could be adjusted throughout the points with pH 4.5 for the maximum activity. This result indicates that the carboxydipeptidase activity of cathepsin B is more efficient at acid pH than is its endopeptidase activity or hydrolysis of peptidyl-AMC substrates without a free carboxyl group, where the

Table 2
Kinetic constants for hydrolysis by cathepsins B, K, L, V, and X and *L. mexicana* CPB of Abz-GIVRAK(Dnp)-OH

Peptidase	Abz-GIVRAK(Dnp)-OH	
	k_{cat}/K_m ($mM^{-1}\ s^{-1}$)	Cleavage site(s)
Cathepsin B	7288	R ↓ A (100%)
Cathepsin K	133.3	V ↓ R (100%)
Cathepsin L	100	V ↓ R (55%) R ↓ A (45%)
Cathepsin V	32	R ↓ A (100%)
Cathepsin X	17	A ↓ K(Dnp)-OH (100%)
CPB 2.8 Δ CTE	50	R ↓ A (100%)
Cruzain	75	V ↓ R (39%) R ↓ A (61%)

Note. The enzymes were activated for 5 min with 2.5 mM DTE. The following buffers were used: 100 mM acetate buffer containing 200 mM NaCl, pH 4.5, for cathepsins B, X, and cruzain; 100 mM acetate buffer containing 200 mM NaCl, pH 5.5, for cathepsins K and L and CPB 2.8 ΔCTE; and 100 mM sodium phosphate with 2.5 mM of EDTA, pH 6.5, for cathepsin V. Measurements were done as described in Materials and methods. Cleavage sites are indicated as arrows (↓). The standard deviations of the kinetic constants were less than 5%. The assays were performed at 37 °C.

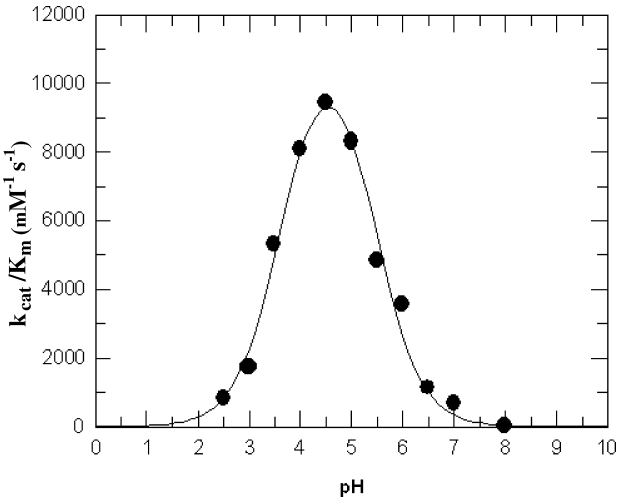

Fig. 3. pH dependence of Abz-GIVRAK(Dnp)-OH hydrolysis by cathepsin B. The pH activity profile for the hydrolysis of Abz-GIVRAK(Dnp)-OH by cathepsin B under pseudo first-order conditions is shown. The determinations were done in a standard buffer containing 25 mM acetic acid, 25 mM Mes, 75 mM Tris, 25 mM glycine, and 2 mM EDTA. The pH ranges were 3.5–10.0 and were adjusted with 1 M NaOH and HCl. The enzyme was preactivated with 2.5 mM DTE at 37 °C for 5 min.

higher hydrolytic activity is shifted to a pH higher than 6 [45,50].

Discussion

The PS–SCL of FRET peptides containing a free C-terminal carboxyl group provides a library of peptides in solution suitable to evaluate the carboxydipeptidase activity of proteases that possess this activity as their major peptidase function. The PS–SCL of FRET peptides we describe here used Lys(Dnp) to carry both the fluorescence quencher group (Dnp) and the free carboxyl group, which is a necessary feature of the substrates for proteases that are predominantly carboxydipeptidases.

The obtained results are essentially in accordance with the already known substrate specificity of cathepsin B. The S_2 specificity is between that of cathepsin L, which prefers Phe, and cathepsin S, which prefers Leu [51], and this library also demonstrated that cathepsin B accommodated the basic amino acids Arg, Lys, and His (in this order of preference) at the S_2 subsite. This is an expected result given that cathepsin B has the residue Glu245 at the bottom of its S_2 subsite. It is curious that substrates containing Pro, but with a free carboxyl group, are reasonably hydrolyzed by cathepsin B, whereas the peptide Abz-APRSAAQ-EDDnp was very poorly hydrolyzed by papain, cathepsin L, and cathepsin B [22]. It seems that substrates with Pro at the P_2 position can be accepted only by cathepsin B executing its carboxydipeptidase activity and that, according to this view, the hydrolysis of Dnp-GPRFW-OH presented $k_{cat}/K_m = 309\ mM^{-1}\ s^{-1}$ [28].

The crystal structure of recombinant rat cathepsin B [52] shows that the benzyl ring of the Z group from Z-Arg-Ser(OBzl), covalently linked to the active site of the protease, makes an aromatic–aromatic interaction with Tyr75 at the S_3 binding site. However, the preference for aliphatic residues such as Ile could result from a best fitting of the 2-butyl group at the S_3 subsite. Because Lys is also well accepted, it is possible that only the hydrophobic and aliphatic portion of the four methylenes of the Lys side chain interacts with S_3 and that the ammonium group could stick out to the solvent.

The prime site cathepsin B substrate specificity observed with the PS–SCL of FRET peptides containing a free C-terminal carboxyl group confirmed the preference for hydrophobic amino acids [28,46,53]. The free C-terminal carboxyl group at the P'_2 position of the substrates directs the carboxydipeptidase activity of cathepsin B, and this is related to its occluding loop, which is a unique structure [27]. The carboxydipeptidase activity of cathepsin B is more efficient at acid pH than is its endopeptidase activity or hydrolysis of peptidyl-AMC substrates without a free carboxyl group, where the higher hydrolytic activity is shifted to a pH higher than 6 [45,50]. This result suggests that the carboxydipeptidase activity of cathepsin B could occur preferentially inside lysosomes and that the endopeptidase activities could occur out of the cells.

Table 3
Earlier reported kinetic parameters for the hydrolysis by human cathepsin B of the more susceptible substrates

Peptide	K_m	k_{cat}	k_{cat}/K_m	Reference
Z-FR-MCA	23.0	76.0	3260	[48]
Abz-FRF-EDDnp	3.5	2.9	828	[48]
Abz-FRL-EDDnp	3.1	2.5	806	[48]
Dnp-GFRFW-OH	1.2	2.1	1725	[28]
Abz-FRAK(Dnp)-OH	17.9	34.8	1944	[28]
Abz-FFDap(Dnp)W-OH	2.1	5.4	2571	[28]

Note. Dap(Dnp), 3-[2,4-dinitrophenyl]-(2,3-diamino propionic acid).

The peptide Abz-GIVRAK(Dnp)-OH obtained from the PS–SCL of FRET peptides containing a free C-terminal carboxyl is the best substrate described for cathepsin B to this point. Table 3 shows, for comparison, the kinetic parameters for the hydrolysis of the best substrates reported previously for cathepsin B with and without the free C-terminal carboxyl group. In addition, looking for possible natural substrates for cathepsin B containing the amino acid sequence closer to that of Abz-GIVRAK(Dnp)-OH, we found the very similar sequences VIRAKF and VIRAKA in human TIMP-1 and TIMP-2, respectively, which correspond to residues 18–23 of them. TIMP-1 and TIMP-2 are tissue inhibitors of metalloproteases, as well as of angiogenesis, and (curiously) were also efficiently hydrolyzed by cathepsin B [38]. We synthesized the peptide acetyl-VIRAKF-NH_2, and the hydrolysis by cathepsin B at the R-A bond was monitored by HPLC. This result indicates that Abz and Lys(Dnp) are not determinants for the hydrolysis of the PS–SCL of FRET peptides.

In conclusion, we have described a PS–SCL specially designed for screening peptidases with carboxydipeptidase activity. Based on the analysis of these libraries, we obtained a very efficient and specific substrate for human cathepsin B.

Acknowledgments

This work was supported by Fundação de Amparo Pesquisa do Estado de São Paulo (FAPESP), Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), and Human Frontiers for Science Progress (RG 00043/2000-M). We also thank Michael J. Blackman for making constructive comments and proofreading the manuscript.

References

[1] D.J. Matthews, J.A. Wells, Substrate phage: selection of protease substrates by monovalent phage display, Science 260 (1993) 1113–1117.

[2] L. Ding, G.S. Coombs, L. Strandberg, M. Navre, D.R. Corey, E.L. Madison, Origins of the specificity of tissue-type plasminogen activator, Proc. Natl. Acad. Sci. USA 92 (1995) 7627–7631.

[3] S.J. Deng, D.M. Bickett, J.L. Mitchell, M.H. Lambert, R.K. Blackburn, H.L. Carter, J. Neugebauer, G. Pahel, M.P. Weiner, M.L. Moss, Substrate specificity of human collagenase 3 assessed using a phage-displayed peptide library, J. Biol. Chem. 275 (2000) 31422–31427.

[4] S.M. Cloutier, J.R. Chagas, J.P. Mach, C.M. Gygi, H.J. Leisinger, D. Deperthes, Substrate specificity of human kallikrein 2 (hK2) as determined by phage display technology, Eur. J. Biochem. 269 (2002) 2747–2754.

[5] J.R. Petithory, F.R. Masiarz, J.F. Kirsch, D.V. Santi, B.A. Malcolm, A rapid method for determination of endoproteinase substrate specificity: specificity of the 3C proteinase from hepatitis A virus, Proc. Natl. Acad. Sci. USA 88 (1991) 11510–11514.

[6] A.J. Birkett, D.F. Soler, R.L. Wolz, J.S. Bond, J. Wiseman, J. Berman, R.B. Harris, Determination of enzyme specificity in a complex mixture of peptide substrates by N-terminal sequence analysis, Anal. Biochem. 196 (1991) 137–143.

[7] B.E. Turk, L.C. Cantley, Using peptide libraries to identify optimal cleavage motifs for proteolytic enzymes, Methods 32 (2004) 398–405.

[8] J. Berman, M. Green, E. Sugg, R. Anderegg, D.S. Millington, D.L. Norwood, J. McGeehan, J. Wiseman, Rapid optimization of enzyme substrates using defined substrate mixtures, J. Biol. Chem. 267 (1992) 1434–1437.

[9] G.M. McGeehan, D.M. Bickett, M. Green, D. Kassel, J.S. Wiseman, J. Berman, Characterization of the peptide substrate specificities of interstitial collagenase and 92-kDa gelatinase: implications for substrate optimization, J. Biol. Chem. 269 (1994) 32814–32820.

[10] V. Schellenberger, C.W. Turck, L. Hedstrom, W.J. Rutter, Mapping the S′ subsites of serine proteases using acyl transfer to mixtures of peptide nucleophiles, Biochemistry 32 (1993) 4349–4353.

[11] K.S. Lam, M. Lebl, Synthesis of a one-bead one-compound combinatorial peptide library, Methods Mol. Biol. 87 (1998) 1–6.

[12] M. Meldal, I. Svendsen, K. Breddam, F.I. Auzanneau, Portion-mixing peptide libraries of quenched fluorogenic substrates for complete subsite mapping of endoprotease specificity, Proc. Natl. Acad. Sci. USA 91 (1994) 3314–3318.

[13] P.M.St. Hilaire, M. Willert, M.A. Juliano, L. Juliano, M. Meldal, Fluorescence-quenched solid phase combinatorial libraries in the characterization of cysteine protease substrate specificity, J. Comb. Chem. 1 (1999) 509–523.

[14] T.A. Rano, T. Timkey, E.P. Peterson, J. Rotonda, D.W. Nicholson, J.W. Becker, K.T. Chapman, N.A. Thornberry, A combinatorial approach for determining protease specificities: application to interleukin-1β converting enzyme (ICE), Chem. Biol. 4 (1997) 149–155.

[15] N.A. Thornberry, T.A. Rano, E.P. Peterson, D.M. Rasper, T. Timkey, M. Garcia-Calvo, V.M. Houtzager, P.A. Nordstrom, S. Roy, J.P. Vaillancourt, K.T. Chapman, D.W. Nicholson, A combinatorial approach defines specificities of members of the caspase family and granzyme B: functional relationships established for key mediators of apoptosis, J. Biol. Chem. 272 (1997) 17907–17911.

[16] J.L. Harris, B.J. Backes, F. Leonetti, S. Mahrus, J.A. Ellman, C.S. Craik, Rapid and general profiling of protease specificity by using combinatorial fluorogenic substrate libraries, Proc. Natl. Acad. Sci. USA 97 (2000) 7754–7759.

[17] B.J. Backes, J.L. Harris, F. Leonetti, C.S. Craik, J.A. Ellman, Synthesis of positional-scanning libraries of fluorogenic peptide substrates to define the extended substrate specificity of plasmin and thrombin, Nat. Biotechnol. 18 (2000) 187–193.

[18] L.W. Herman, G. Tarr, S.A. Kates, Optimization of the synthesis of peptide combinatorial libraries using a one-pot method, Mol. Diversity 2 (1996) 147–155.

[19] I. Schechter, A. Berger, On the active site of proteases: III. Mapping the active site of papain—specific peptide inhibitors of papain, Biochem. Biophys. Res. Commun. 32 (1968) 898–902.

[20] M.E. McGrath, The lysosomal cysteine proteases, Annu. Rev. Biophys. Biomol. Struct. 28 (1999) 181–204.

[21] D.K. Nägler, A.C. Storer, F.C.V. Portaro, E. Carmona, L. Juliano, R. Ménard, Major increase in endopeptidase activity of human cathepsin B upon removal of occluding loop contacts, Biochemistry 36 (1997) 12608–12615.

[22] F.C.V. Portaro, A.B.F. Santos, M.H.S. Cezari, M.A. Juliano, L. Juliano, L. Carmona, Probing the specificity of cysteine proteinases at subsites remote from the active site: analysis of P_4, P_3, P'_2, and P'_3 variations in extended substrates, Biochem. J. 347 (2000) 123–129.

[23] D. Musil, D. Zucic, R.A. Engh, I. Mayr, R. Huber, T. Popovic, V. Turk, T. Towatari, N. Katunuma, W. Bode, The refined 2.15 Å X-ray crystal structure of human liver cathepsin B: the structural basis for its specificity, EMBO J. 10 (1991) 2321–2330.

[24] M. Cygler, J. Sivaraman, P. Grochulski, R. Coulombe, A.C. Storer, J.S. Mort, Structure of rat procathepsin B: model for inhibition of cysteine protease activity by the proregion, Structure 4 (1996) 405–416.

[25] D. Turk, M. Podobnick, T. Popovic, M. Katunuma, W. Bode, R. Huber, V. Turk, Crystal structure of cathepsin B inhibited by CA030 at 2.0 Å resolution: a basis for the design of specific epoxysuccinyl inhibitors, Biochemistry 34 (1995) 4791–4797.

[26] C. Illy, O. Quraishi, J. Wang, E. Purisima, T. Vernet, J.S. Mort, Role of occluding loop in cathepsin B activity, J. Biol. Chem. 272 (1997) 1197–1202.

[27] J.C. Krupa, S. Hasnain, D.K. Nägler, R. Ménard, J.S. Mort, S'_2 substrate specificity and the role of His^{110} and His^{111} in the exopeptidase activity of human cathepsin B, Biochem. J. 361 (2002) 613–619.

[28] M.H.S. Cezari, L. Puzer, M.A. Juliano, A.K. Carmona, L. Juliano, Cathepsin B carboxydipeptidase specificity analysis using internally quenched fluorescent peptides, Biochem. J. 368 (2002) 365–369.

[29] W. Halangk, M.M. Lerch, B. Brandt-Nedelev, W. Roth, M. Ruthenbuerger, T. Reinheckel, W. Domschke, H. Lippert, C. Peters, J. Deusing, Role of cathepsin B in intracellular trypsinogen activation and the onset of acute pancreatitis, J. Clin. Invest. 106 (2000) 773–781.

[30] J.S. Mort, D.J. Buttle, Molecules in focus: cathepsin B, Int. J. Biochem. Cell Biol. 29 (1997) 715–720.

[31] S. Yan, M. Sameni, B.F. Sloane, Cathepsin B and human tumor progression, Biol. Chem. 379 (1998) 113–123.

[32] B.A. Frosch, I. Berquin, M.R. Emmert-Buck, K. Moin, B.F. Sloane, Molecular regulation, membrane association, and secretion of tumour cathepsin B, APMIS 107 (1999) 28–37.

[33] A.M. Szpaderska, A. Frankfater, An intracellular form of cathepsin B contributes to invasiveness in cancer, Cancer Res. 61 (2001) 3493–3500.

[34] A.A. Sinha, M.P. Jamuar, M.J. Wilson, J. Rozhin, B.F. Sloane, Plasma membrane association of cathepsin B in human prostate cancer: biochemical and immunological electron microscopy analysis, Prostate 49 (2001) 172–184.

[35] M.E. Guicciardi, J. Deussing, H. Miyoshi, S.F. Bronk, P.A. Svingeon, C. Peters, S.H. Kaufmann, G.J. Gores, Cathepsin B contributes to TNF-α-mediated hepatocyte apoptosis by promot-

ing mitochondrial release of cytochrome *c*, J. Clin. Invest. 106 (2000) 1127–1137.

[36] V. Stoka, B. Turk, S.L. Schendel, T.-H. Kim, T. Cirman, S.J. Snipas, L.M. Ellerby, D. Bredesen, H. Freeze, M. Abrahamson, D. Brömme, S. Krajewski, J.C. Reed, X.-M. Yin, V. Turk, G.S. Salvensen, Lysosomal protease pathways to apoptosis: cleavage of Bid, not pro-caspases, is the most likely route, J. Biol. Chem. 276 (2001) 3149–3157.

[37] T. Cirman, K. Oresic, G.D. Mazovec, V. Turk, J.C. Reed, R.M. Myers, G.S. Salvesen, B. Turk, Selective disruption of lysosomes in HeLa cells triggers apoptosis mediated by cleavage of Bid by multiple papain-like lysosomal cathepsins, J. Biol. Chem. 279 (2004) 3578–3587.

[38] G. Kostoulas, A. Lang, H. Nagase, A. Baici, Stimulation of angiogenesis through cathepsin B inactivation of the tissue inhibitors of matrix metalloproteinases, FEBS Lett. 455 (1999) 286–290.

[39] A.E. Eakin, A.A. Mills, G. Harth, J.H. McKerrow, C.S. Craik, The sequence, organization, and expression of the major cysteine protease (cruzain) from *Trypanosoma cruzi*, J. Biol. Chem. 267 (1992) 7411–7420.

[40] D.K. Nägler, R. Zhang, W. Tam, T. Sulea, E.O. Purisima, R. Ménard, Human cathepsin X: a cysteine protease with unique carboxypeptidase activity, Biochemistry 38 (1999) 12648–12654.

[41] A.J. Barrett, H. Kirshke, Cathepsin B, cathepsin H, and cathepsin L, Methods Enzymol. 80 (1981) 771–778.

[42] S.J. Sanderson, K.G. Pollock, J.D. Hilley, M. Meldal, P.S. Hilaire, M.A. Juliano, L. Juliano, J.C. Mottram, G.H. Coombs, Expression and characterization of a recombinant cysteine proteinase of *Leishmania mexicana*, Biochem. J. 347 (2000) 383–388.

[43] J.M. Ostresh, J.H. Winkle, V.T. Hamashin, R.A. Houghten, Peptide libraries: determination of relative reaction rates of protected amino acids in competitive couplings, Biopolymers 34 (1994) 1681–1689.

[44] J.J. Peterson, C.F. Meares, Cathepsin substrates as cleavable peptide linkers in bioconjugates selected from a fluorescence quench combinatorial library, Bioconjug. Chem. 9 (1998) 618–626.

[45] D.K. Nagler, W. Tam, A.C. Storer, J.C. Krupa, J.S. Mort, R. Menard, Interdependency of sequence and positional specificities for cysteine proteases of the papain family, Biochemistry 38 (1999) 4868–4874.

[46] R.L. Melo, R.C. Barbosa Pozzo, L.C. Alves, E. Perissutti, G. Caliendo, V. Santagada, L. Juliano, M.A. Juliano, Synthesis and hydrolysis by cathepsin B of fluorogenic substrates with the general structure benzoyl-X-ARG-MCA containing non-natural basic amino acids at position X, Biochim. Biophys. Acta 1547 (2001) 82–94.

[47] A. Taralp, H. Kaplan, I.I. Sytwu, I. Vlattas, R. Bohacek, A.K. Knap, T. Hirama, C.P. Huber, S. Hasnain, Characterization of the S_3 subsite specificity of cathepsin B, J. Biol. Chem. 270 (1995) 18036–18043.

[48] R.L. Melo, L.C. Alves, E. Del Nery, L. Juliano, M.A. Juliano, Synthesis and hydrolysis by cysteine and serine proteases of short internally quenched fluorogenic peptides, Anal. Biochem. 293 (2001) 71–77.

[49] S.A. Gillmor, C.S. Craik, R.J. Fletterick, Structural determinants of specificity in the cysteine protease cruzain, Protein Sci. 6 (1997) 1603–1611.

[50] W.A. Judice, L. Puzer, S.S. Cotrin, A.K. Carmona, G.H. Coombs, L. Juliano, M.A. Juliano, Carboxydipeptidase activities of recombinant cysteine peptidases: cruzain of *Trypanosoma cruzi* and CPB of *Leishmania mexicana*, Eur. J. Biochem. 271 (2004) 1046–1053.

[51] D. Bromme, P.R. Bonneau, P. Lachance, A.C. Storer, Engineering the S_2 subsite specificity of human cathepsin S to a cathepsin L- and cathepsin B-like specificity, J. Biol. Chem. 269 (1994) 30238–30242.

[52] Z. Jia, S. Hasnain, T. Hirama, X. Lee, J.S. Mort, R. To, C.P. Huber, Crystal structures of recombinant rat cathepsin B and a cathepsin B–inhibitor complex: implications for structure-based inhibitor design, J. Biol. Chem. 270 (1995) 5527–5533.

[53] R. Ménard, E. Carmona, C. Plouffe, D. Brömme, Y. Konishi, J. Lefebvre, A.C. Storer, The specificity of the S'_1 subsite of cysteine proteases, FEBS Lett. 328 (1993) 107–110.