Characterization of the substrate specificity of the major cysteine protease (cruzipain) from *Trypanosoma cruzi* using a portion-mixing combinatorial library and fluorogenic peptides

Elaine DEL NERY*, Maria A. JULIANO*, Morten MELDAL†, Ib SVENDSEN†, Julio SCHARFSTEIN‡, Adrian WALMSLEY§ and Luiz JULIANO*||

*Department of Biophysics, Escola Paulista de Medicina, Rua Tres de Maio 100, 04044-020 São Paulo, Brazil, †Department of Chemistry, Carlsberg Laboratory, Gamle Carlsberg Vej 10, DK-2500 Valby-Copenhagen, Denmark, ‡Molecular Immunology Laboratory, Instituto de Biofísica Carlos Chagas Filho, C. C. S. Universidade Federal do Rio de Janeiro, Rio de Janeiro 21949, Brazil, and §Department of Molecular Biology and Biotechnology, University of Sheffield, P.O. Box 594, First Court Western Bank, Sheffield S10 2UH, U.K.

The substrate specificity of the major cysteinyl proteinase of the parasitic protozoan *Trypanosoma cruzi* (cruzipain) was investigated, by combinatorial replacement of amino acid residues at positions P_5 – P'_5 , using a fluorescent quenched solid-phase library assay. Positively charged residues appear to be a general preference in the P_5 – P_3 and the P'_5 – P'_3 positions, while a hydrophobic residue was always required at the P_2 position. A broad range of amino acids could be accepted at the P'_1 position. A clear difference in terms of specificity between cruzipain and human cathepsin L was observed for the accommodation of Pro at the P_2 position. The P_1 specificity was investigated by a more detailed enzyme kinetic analysis using peptidyl-MCA (where

MCA is methylcoumarin amide) and Abz-peptidyl-EDDnp [where Abz is *o*-aminobenzoic acid and EDDnp is *N*-(2,4-dinitrophenyl)ethylenediamine] as substrates, and the results were compared with those obtained using human cathepsin L. Cruzipain showed a clear preference for benzyl-Cys or Arg at the P_1 position. Human cathepsin L presented similar behaviour to that of cruzipain for the hydrolysis of the ϵ -NH₂-Cap-Leu-Xaa-MCA (where Cap is ϵ -aminocaproyl) and Abz-Lys-Leu-Xaa-Phe-Ser-Lys-Gln-EDDnp series, whereas the mammalian enzyme was able to tolerate large P_1 residues, such as phenylalanine, better than cruzipain in the latter series.

INTRODUCTION

Cysteine proteinases show a broad substrate specificity, which is an obstacle in the design of specific inhibitors to selectively target individual proteinases. Their critical role in the life cycle of parasitic organisms also makes them attractive drug-design targets for the parasitic diseases [1]. Cruzipain (or GP57/51) is the major cysteine proteinase in *Trypanosoma cruzi* [2,3], the aetiological agent of Chagas' disease, and it is the immunodominant target of the host immune response [4,5]. The amino acid sequence data, coupled with enzymic characterization, classified this proteinase as a member of the papain superfamily [4,6]. The X-ray crystal structure of recombinant cruzipain, lacking the 130-amino-acid C-terminal extension, revealed a 215-residue polypeptide composed of two well defined domains [7]. The catalytic residues Cys²⁵, His¹⁵⁹ and Asn¹⁷⁵ and the extended substrate binding site are found in a cleft between the two domains [7]. The analysis of sequence alignment of each domain with mammalian proteins reveals that the central domain of cruzipain is closely related to cathepsin L (56.8% identity), whereas no obvious similarities are found for the peculiar C-terminal extension [8]. The proteinase activity seems to be essential for parasite growth and transformation, because low concentrations of irreversible membrane-permeable cysteine proteinase inhibitors, which also are very effective inhibitors of cruzipain, block the intracellular multiplication of the amasti-

gotes [9]. It was recently demonstrated that biotinyl-Phe-Ala-fluoromethyl ketone and polyclonal antibodies to cruzipain selectively label the intracellular amastigote form of *Trypanosoma cruzi* in an *in vitro* model of mammalian cell infection [7]. The enzyme appears near the surface of the intracellular parasite, and biotinyl-Phe-Ala-fluoromethyl ketone did not label the host cell proteinases. These putative functions, all at important stages in the life cycle of the parasite, make cruzipain an excellent target for specific inhibitors that may be used to develop a new chemotherapy.

The substrate specificity of cruzipain was recently investigated using intramolecularly quenched fluorescent substrates derived from cystatin peptides [10], which include either the substrate-like N-terminal segment of cystatins (Leu-Val-Gly) or the papain-sensitive sequence Gln-Val-Val-Ala-Gly. In order to obtain a more comprehensive picture of the importance of the amino acids at positions P_5 – P'_5 of substrates for cruzipain, a portion mixing combinatorial library [11] of internally quenched peptide substrates 3-nitrotyrosine [Tyr(NO₂)]/Lys(Abz) [12] was assayed (where Abz is *o*-aminobenzoic acid). From the information obtained using the library, a series of intramolecularly quenched fluorogenic peptides with the general structure Abz-peptidyl-Tyr(NO₂)-Asp was designed and assayed as cruzipain substrates. In addition, peptidyl-MCA [13] (where MCA is methylcoumarin amide) and Abz-peptidyl-Gln-EDDnp [10,14–16] [where EDDnp is *N*-(2,4-dinitrophenyl)ethylenediamine] were synthesized with

Abbreviations used: Abz, *o*-aminobenzoic acid; Bzl, benzyl; Cap, ϵ -aminocaproyl (6-aminoheptanoic); EDDnp, *N*-(2,4-dinitrophenyl)ethylenediamine; Boc, *t*-butoxycarbonyl; Dhbt, 3,4-dihydro-4-oxo-1,2,3-benzotriazin-3-yl; DMF, *N,N*-dimethylformamide; DTT, dithiothreitol; E-64, *trans*-epoxysuccinyl-L-leucylamido-(4-guanidino)butane; Fmoc, fluorenylmethoxycarbonyl; MCA, methylcoumarin amide; PEGA, poly(ethylene glycol) acrylamide copolymer; Pfp, pentafluorophenyl; *t*Bu, *t*-butyl; TFA, trifluoroacetic acid; Tyr(NO₂) or Y(NO₂), 3-nitrotyrosine.

|| To whom correspondence should be addressed.

systematic variation at the P₁ position and studied as substrates for cruzipain and for human cathepsin L for comparison.

EXPERIMENTAL

Materials

All solvents were purchased from Labscan Ltd. (Dublin, Ireland). Fluoren-9-ylmethoxycarbonyl (Fmoc) amino acids and their pentafluorophenyl (Pfp) ester derivatives were purchased from Bachem and NovaBiochem. Boc-anthranilic acid-ODhbt ester (where Boc is t-butoxycarbonyl and Dhbt is 3,4-dihydro-4-oxo-1,2,3-benzotriazin-3-yl) and Fmoc-nitrotyrosine were prepared as previously described [12]. Poly(ethylene glycol) acrylamide copolymer (PEGA) resin [17] was synthesized by inversion suspension polymerization of partially acryloylated bis-2-aminopropyl poly(ethylene glycol) (M_r 1900; Fluka) in hexane/tetrachloromethane as previously described [18].

Optical rotations were recorded on a Perkin–Elmer 241 polarimeter and are given in units of 10^{-1} degrees $\cdot$ cm³ $\cdot$ g⁻¹. Matrix-assisted laser desorption/ionization (MALDI)-MS was performed on a Finnigan MAT 2000 instrument with a matrix of α -cyano-4-hydroxycinnamic acid. Electrospray-MS was performed on a VG-Quattro instrument from Fisons. HPLC was performed on a Waters HPLC system with Delta pack C18 columns (200 mm $\times$ 25 mm; 15 μ m; flow rate 10 ml/min for preparative separations) with solvent A [0.1% trifluoroacetic acid (TFA) in water] and solvent B (0.1% TFA in 90% aq. acetonitrile) unless otherwise indicated. Amino acid analyses were carried out in a Beckman 6300 amino acid analyser following hydrolysis in 6 M HCl with 5% (v/v) phenol at 110 °C for 48 h. Amino acid sequencing was performed on an Applied Biosystems Sequencer model 477A equipped with an on-line phenylthiohydantoin-amino acid analyser (model 120A), according to the protocol of the manufacturer.

Enzymes

Purified cruzipain (GP57/51) was obtained from crude aqueous extracts of *T. cruzi* Dm28c strain epimastigotes as described by Lima et al. [19]. Cathepsin L were purchased from Calbiochem/Novabiochem (La Jolla, CA, U.S.A.). The molar concentrations of the enzyme solutions were determined by active-site titration with E-64 [*trans*-epoxysuccinyl-L-leucylamido-(4-guanidino) butane] (Sigma) as previously described [20].

Synthesis of a substrate library for cruzipain

A library with the structure Y(NO₂)X¹X²X³X⁴(K/R)X⁵X⁶K-(Abz)-PEGA₁₉₀₀ was prepared in a 20-column library synthesizer with a mixing chamber. The residues X¹–X⁶ designate all the encoded amino acids except for Cys and Trp, which were substituted with Lys and Arg; thus the amounts of Lys and Arg incorporated were twice that of any other amino acid in the X¹–X⁶ positions. Cys and Trp were avoided, since they were not found to be essential in the preliminary experiments and they may cause problems in synthesis and analysis. At the position 5 only Lys and Arg were added, to 10 columns each.

PEGA₁₉₀₀ resin (2 g; 0.18 mmol of NH₂/g) was prepared by the partial acryloylation (0.77 equiv.) procedure [18] using acrylamide as an additive that was swollen with *N,N*-dimethylformamide (DMF) in a syringe synthesizer. It was washed with 20% piperidine in DMF and then with 6 vol. of DMF. The resin was coupled with Fmoc-Lys(Boc)-OPfp (684 mg; 1.08 mmol) and Dhbt-OH (57 mg) for 6 h, and then washed with 6 vol. of DMF, 4 vol. of dichloromethane and 1 vol. of 50% TFA in

dichloromethane. It was then treated for 20 min with 50% TFA in dichloromethane and washed with 6 vol. of dichloromethane and 4 vol. of DMF. Excess solvent was removed by suction, and Boc-Abz-ODhbt (430 mg; 1.08 mmol) and 4-ethylmorpholine (46 ml; 0.36 mmol) were added in DMF (15 ml). The resin was washed with 5 vol. of DMF and distributed uniformly between the columns in the library synthesizer and washed with 2 vol. of DMF from the first parallel liquid distribution head of the synthesizer. The Fmoc group was cleaved with 20% piperidine in DMF (1 vol. for 1 min and 1 vol. for 15 min) from the second parallel liquid distribution head of the synthesizer, and the resin was washed with 6 vol. of DMF. Each of the 18 Fmoc-amino acid-OPfp esters (3 equiv.) and Dhbt-OH (1 equiv.) were coupled for 18 h in respective columns. The amino acid side-chain protection groups used were trityl (His, Gln and Asn), *t*-butyl (Ser, Thr, Glu, Asp and Tyr), Boc (Lys) and pentamethyl chromatinylsulphonyl (Arg), and stock solutions were prepared and kept at –20 °C. The resin was washed with 5 vol. of DMF, and the synthesizer was filled with DMF (54 ml) and turned upside down. It was shaken vigorously for 30 min and turned upright to give an even distribution between the columns. The Fmoc group was cleaved as above and the cycle was repeated. Then followed a cycle where only Fmoc-Arg(pentamethyl chromatinylsulphonyl)-OPfp and Fmoc-Lys(Boc)-OPfp were coupled in 10 columns each and the resin was mixed. Four more cycles with the 18 Fmoc-amino acid-OPfp esters and mixing were performed. When a column occasionally did not completely lose the yellow colour of the ion pair Dhbt-O⁻/H₃N⁺-resin, indicating incomplete coupling, the procedure was repeated in that column. After final Fmoc deprotection, the resin was washed with DMF and dichloromethane and dried. It was then treated with 95% aq. TFA for a period of 2.5 h and washed with 95% aq. acetic acid and dichloromethane. The resin was finally treated with 5% di-isopropyl ethylamine in DMF, washed with DMF and dichloromethane, dried and subjected to enzyme hydrolysis.

Substrates

Peptidyl-MCA substrates were synthesized by the standard solution-phase peptide synthesis procedure starting from protected aminoacyl-MCA; this procedure was significantly improved by employing the phosphorous oxychloride anhydride method as previously described [13]. The details of syntheses of the intramolecularly quenched fluorogenic peptides having Abz at the N-terminus and EDDnp attached to glutamine, a requirement for the solid-phase peptide synthesis strategy employed, were as reported elsewhere [21]. Multiple-column peptide synthesis of Abz-peptidyl-Tyr(NO₂)-Asp-OH substrates were carried out as previously described [12,22].

The final deprotected peptides were purified by semi-preparative HPLC using an Econosil C-18 column (10 μ m; 22.5 mm $\times$ 250 mm) and two solvent systems: A, TFA/water (1:1000, v/v); B, TFA/acetonitrile/water (1:900:100, by vol.). The column was eluted at a flow rate of 5 ml/min with a 10% (or 30%)–50% (or 60%) gradient of solvent B over 30 or 45 min. Analytical HPLC was performed using a C-R7A Shimadzu UV/visible detector and a Shimadzu RF-535 fluorescence detector, coupled to an Ultrasphere C-18 column (5 μ m; 4.6 mm $\times$ 150 mm) which was eluted with solvent systems A and B above at a flow rate of 1.7 ml/min and a 10–80% gradient of B over 15 min. The HPLC column eluates were monitored by their absorbance at 220 nm and by fluorescence emission at 420 nm following excitation at 320 nm for intramolecularly quenched fluorogenic peptides, and by fluorescence emission at 395 nm following excitation at 325 nm for MCA substrates.

Hydrolysis of the peptide library

The PEGA-resin-bound peptide library (1 g; 20 ml) was washed with reaction buffer [25 mM sodium acetate, 2 mM EDTA, 0.2 M NaCl, 2 mM dithiothreitol (DTT), pH 5.0]. The resin was suspended to a final volume that contained 20% of additional liquid over the top of the resin and 100 ml (35 mg) of cruzipain (preactivated with 10 mM DTT on ice for 20 min). The reaction was carried out for 5 h at 30 °C with agitation. The resin was washed three times with water in order to obtain a clear supernatant, and then re-equilibrated with reaction buffer and treated with another portion of cruzipain (35 mg). After 3 days of incubation, it was washed successively with 2% TFA, water, 2% sodium hydrogen carbonate solution and finally water. The resin was lyophilized for storage. The resin from the above treatment was swollen in water and investigated under the fluorescence microscope in 30 small portions on a glass plate in an excess of water at pH 7.5. From each portion the most fluorescent bead, for which a ring of illumination was observed, was collected by transfer to the dry periphery of the plate and further transfer with a closed glass capillary to a sequencing filter.

Fluorimetric enzyme assay

Hydrolysis of the internally quenched fluorescent peptides at 37 °C in 100 mM NaH₂PO₄, 10 mM EDTA, 5 mM DTT, 400 mM NaCl and 2 mM Cys, pH 6.3, by cruzipain was followed by measuring the fluorescence at $\lambda_{\text{ex}} = 320$ nm and $\lambda_{\text{em}} = 420$ nm in a Hitachi F-2000 spectrofluorimeter. The hydrolysis of peptidyl-MCA substrates was performed under the same conditions ($\lambda_{\text{ex}} = 380$ nm; $\lambda_{\text{em}} = 460$ nm). The enzyme was pre-activated in 200 μ l of the described buffer at 37 °C for 10 min prior to the hydrolysis of substrates. The 1 cm path-length cuvette containing 1.8 ml of the substrate solution was placed in a thermostatted cell compartment. After the temperature was equilibrated, the enzyme solution was added and the increase in fluorescence with time was recorded continuously for 10 min. The slope was converted into mol of substrate hydrolysed per min, based on the fluorescence curves for standard peptide solutions before and after total enzymic hydrolysis. A solution of Abz-Phe-Arg-OH was used as a standard for the fluorescence measurements; this was prepared from the tryptic hydrolysis of Abz-Phe-Arg-4-nitroanilide and its concentration was determined from the absorbance at 405 nm, assuming $\epsilon_{405} = 8900 \text{ M}^{-1} \cdot \text{cm}^{-1}$ for 4-nitroanilide. Provided that HPLC analysis of the different fluorogenic substrates did not show any significant contamination, their concentrations in solution could be determined from the fluorescence obtained following total tryptic hydrolysis. The second-order rate constant $k_{\text{cat}}/K_{\text{m}}$ was measured under pseudo-first-order conditions using a substrate concentration far below the K_{m} . The assays with human cathepsin L were performed under similar conditions at pH 5.8.

Cleavage site identification

The products from the hydrolysis of Abz-peptidyl-Y(NO₂)D-OH were identified by HPLC on a Novapak C-18 column (5 μ m; 3.2 mm $\times$ 150 mm) equilibrated with 0.1% H₃PO₄ (solvent C). The column was eluted at a flow rate of 1.7 ml/min with 0% solvent D (90% acetonitrile/0.1% H₃PO₄) over 3 min followed by a 0–80% (v/v) gradient of the same solvent over 20 min. The elution profile was determined at 283 nm and by fluorescence at 420 nm after excitation at 320 nm. All peptides were incubated with activated cruzipain (4 nM) or cathepsin L in the buffers described above. Standard solutions of the substrates and of

authentic synthetic peptides corresponding to the products of substrate hydrolysis were chromatographed under the same conditions.

RESULTS AND DISCUSSION

Hydrolysis of peptides generated from the peptide library

A preliminary experiment of hydrolysis with cruzipain was carried out with a library, Y(NO₂)X¹X²X³X⁴X⁵X⁶K(Abz)-PEGA_{1900/300}, previously prepared for subtilisin [22]. Subjecting this low-swelling library to cruzipain afforded only hydrolysis on the surface of some beads, which were collected by their fluorescence and sequenced. However, the proteinase did not enter into the resin and the cleavage point could not be determined. The effective substrates were RGPKRY, AGPKLR, RLKKRL, RNKLYK, QHKMYK, K-PKVF and KPRKYK, all of which had a surprisingly high content of basic residues. Based on the results with a non-biased library, a second library with the structure Y(NO₂)X¹X²X³X⁴(K/R)X⁵X⁶K(Abz) was synthesized on PEGA₁₉₀₀, which swelled as much as 17-fold in water. Lys and Arg were incorporated twice at the randomized positions X¹–X⁶, and were also used instead of Trp and Cys, which are prone to side reactions. This strategy was applied in order to obtain a larger number of peptides with Lys/Arg residues at the randomized positions that could result in a larger number of peptides that are susceptible to hydrolysis by cruzipain.

The library was subjected to hydrolysis by the enzyme, and the most fluorescent beads were collected and submitted to sequence analysis. Cruzipain presented high specificity for basic residues in positions P₃ and P₁ and a hydrophobic, preferentially non-aromatic, residue was always required in position P₂. When basic residues were not present in position P₃ or P₂, compensation by basic residues in positions P₄, P₁ or P₃ was often observed. The (K/R)X(K/R) and (K/R)XX(K/R) patterns for P₁–P₃ were recognized by cruzipain and could be correlated with the presence of a negatively charged Asp¹⁸ residue in the S₂/S₃ region of the active site of the enzyme. No immediate enzyme preference was observed in the solid-phase assay for any specific amino acid at position P₁; however, the substitution of Ala with Leu increased the activity in solution (results not shown). The solid-phase assay is supposed to have a cumulative tendency to segregate the best substrates; however, our results suggest that information derived from such studies cannot be completely transferred to solution studies in order to provide substrates with the greatest susceptibility. This behaviour is probably attributable to non-additivity of the interactions between each single enzyme subsite

Table 1 Kinetic parameters for the hydrolysis by cruzipain of internally quenched fluorogenic peptides containing the general sequence Abz-peptidyl-Y(NO₂)D-OH with substitutions at the P₂ position

Cruzipain was assayed at 37 °C in 100 mM NaH₂PO₄, 10 mM EDTA, 5 mM DTT, 400 mM NaCl and 2 mM Cys, pH 6.3. $k_{\text{cat}}/K_{\text{m}}$ values were determined under pseudo-first-order conditions, as described in the Experimental section. The cleavage site is indicated by ↓.

Peptide no.	Substrate							$k_{\text{cat}}/K_{\text{m}}$ (mM·s) ^{−1}
	P ₄	P ₃	P ₂	P ₁	↓ P ₁	P ₂	P ₃	
1	K	K	W	K	A	K	K	344
2	K	K	F	K	A	K	K	1758
3	K	K	Y	K	A	K	K	2895
4	K	K	L	K	A	K	K	3421
5	K	K	V	K	A	K	K	5710

with a specific amino acid. This is consistent with the subsite compensation observed in the studies reported herein [23]. In addition, there are differences in the environment, since in the library assays the peptides were bound to gel-like solid phases, and the high incidence of basic residues in the peptides generated by the solid-phase peptide library could be, in part, a requirement for peptides to become hydrated and thus accessible to cruzipain.

Preference of cruzipain for a hydrophobic non-aromatic residue at position P_2

The cruzipain preference for substrates containing hydrophobic non-aromatic residues at the P_2 position was observed in the library assay and confirmed by the specificity constant (k_{cat}/K_m) values obtained for the hydrolysis in solution of substrates in the series Abz-KKXXKAKKY(NO₂)D-OH (Table 1). Although the presence of Leu, Val or even Tyr in the X position did not induce a dramatic change in the specificity constant values for cruzipain, the S_2 pocket did not seem to accommodate the large Trp side chain efficiently.

Scan of Ala and Pro on Abz-KKLKAKKRY(NO₂)D-OH

The substrate Abz-KKLKAKKRY(NO₂)D-OH (peptide 6 in Table 2) was designed using the more frequent residues that were depicted in the library. It was taken as the lead substrate, and Ala and Pro were scanned in all positions. Leu instead of Val was maintained at P_2 because we have systematically used Leu in both the peptidyl-MCA and Abz-peptidyl-Q-EDDnp substrates (see Tables 3 and 4), which were initially designed at a non-prime site similar to that in E-64; it is known that the side chain of the Leu residue of this inhibitor occupies the S_2 subsite on binding to papain. On the other hand, Ala and Pro were chosen for scanning the lead substrate because the former has the smallest side chain among the amino acids and the latter introduced some degree of conformational restriction.

The presence of Ala in positions P_3 – P_1 (peptides 8–10; Table 2) resulted in a significant decrease in the k_{cat}/K_m value, whereas its presence at positions P'_2 – P'_4 (peptides 11–13) increased the k_{cat}/K_m value. Interestingly, as also observed with papain [24], the presence of Ala at the P'_2 position had a very favourable effect on the susceptibility of the substrate to cruzipain, which resulted in one of the best substrates in this series. However, the S'_2 – P'_2 interaction seems to depend on other enzyme–substrate subsite interactions, because no significant effect was detected on the k_{cat}/K_m value for the hydrolysis of Abz-LVGGAQ-EDDnp in comparison with Abz-LVGGRQ-EDDnp [10]. The lower k_{cat}/K_m value with Ala at P_3 (peptide 8) and the higher value for the substrate with Ala at P'_3 (peptide 12) in comparison with the lead substrate are consistent with the data from the library assay, since basic amino acids were detected at P_3 and P'_3 with frequencies of 60% and 15% respectively in the peptides analysed. In the peptide 9, the absence of a bulky hydrophobic side chain at position P_2 decreased significantly the specificity constant, confirming the observations shown in Table 1 and discussed above. Ala at position P_1 (peptide 10) did not induce a large decrease in the specificity constant.

A Pro residue at position P_2 or P'_3 resulted in substrates (peptides 15 and 19 respectively; Table 2) that were hydrolysed with significantly greater k_{cat}/K_m values than the lead substrate (peptide 6). The substrate with Pro at P_3 (peptide 14) was hydrolysed with lower k_{cat}/K_m value, and the presence of Pro at P'_2 (peptide 18) had no significant effect in comparison with lead peptide 6. The high susceptibility of the Lys–Ala bond in Abz-KKPKAKKRY(NO₂)D-OH (peptide 15) with Pro

at position P_2 is a particularly interesting observation for cruzipain specificity, because of the general requirement of thiol proteases for substrates containing hydrophobic amino acids with aromatic or aliphatic branched side chains at the P_2 position [13,25]. Pro should not be expected to be accepted by the enzyme due to its imino group, which creates conformational restrictions and impairs hydrogen bonding. In addition, we observed that several Xaa-Pro-Arg-4-nitroanilide peptides are resistant to hydrolysis by papain (results not shown), and peptide 15 was resistant to cathepsin L. Specific hydrolysis by cruzipain of the peptide dansyl-VPGGP-NH₂, with Pro at position P_2 , was observed by Serveau [26]. The inability to hydrolyse peptidyl-4-nitroanilide substrates containing a Pro residue in the P_2 position has also been demonstrated for cathepsins B, L and S [27]. In addition, four out of seven peptides identified as substrates for cruzipain in the subtilisin library described in the Experimental section (RGPKRY, AGPKLR, RLKKRL, RNKLYK, QHKMYK, K-PKVF and KPRKYK) contain Pro-Lys/Arg amino acid pairs, suggesting that Pro in these peptides could occupy the P_2 position. Finally, Pro at positions P_1 and P'_1 (peptides 16 and 17) resulted in peptides that are cleaved by cruzipain at Lys–Leu bonds with significantly lower k_{cat}/K_m values.

P_1 specificity analysis

The S_1 subsite of cysteine proteinases has a less defined specificity than the S_2 subsite, and most of their common substrates contain Arg as the P_1 residue. We studied the susceptibility to cruzipain of peptides generated from the lead substrate with systematic variation at the P_1 position (peptides 20–29; Table 2). Of this series, cruzipain showed a marked preference for Phe, Cys(tBu), Leu and Tyr residues (where tBu is t-butyl), whereas the substrates containing amino acids with more hindered β -branched side chains, such as Ile and Val, were more resistant to hydrolysis.

The substrate with His at position P_1 showed significant susceptibility to cruzipain, which is consistent with the observation of an autocatalytic cleavage at a His residue in the propeptide region of recombinant cruzipain [6]. This favourable effect of His could also be related to the protonated or unprotonated forms of the imidazole side chain, since all the hydrolyses of substrates were carried out at pH 6.3, which is close to the pK of imidazole in peptides [28].

Hydrolysis of peptidyl-MCA and Abz-peptidyl-Q-EDDnp substrates by cruzipain and human cathepsin L

Benzyloxycarbonyl-Phe-Arg-MCA and -Arg-Arg-MCA are commercially available fluorescent peptides that are commonly used as substrates for thiol proteinases; however, they are also very susceptible to hydrolysis by trypsin-like enzymes. On the other hand, the S_1 subsite of thiol proteases is not strictly selective [29], as discussed above for cruzipain. In accordance with the X-ray analysis of enzyme–inhibitor complexes, this seems to be a consequence of the position of the P_1 side chain, which points upwards, away from the enzyme surface [30]. However, analysis of the crystal structure of the recombinant rat cathepsin B–benzyloxycarbonyl-Arg-Ser(O-Bzl)-chloromethyl ketone complex (where Bzl is benzyl) shows that the benzyl group of the P_1 residue of the inhibitor occupies a large hydrophobic pocket, which was suggested to be the S'_1 subsite [31]. Based on these considerations and on the structure of E-64, we have previously synthesized the ϵ -NH₂-Cap-Leu-Xaa-MCA series (where Cap is ϵ -aminocaproyl), in which Xaa was the benzyl derivative of

Table 2 Kinetic parameters for the hydrolysis by cruzipain of substrates generated from the lead sequence Abz-KKLKAKKRY(NO₂)D-OH

k_{cat}/K_m values were determined under pseudo-first-order conditions, as described in the Experimental section. The cleavage site is indicated by ↓. A trace mark (–) indicates resistance to cleavage at this point.

Peptide no.	Substrate									k_{cat}/K_m (mM·s) ^{–1}
	P ₄	P ₃	P ₂	P ₁	↓	P' ₁	P' ₂	P' ₃	P' ₄	
6	K	K	L	K		A	K	K	R	3800
Alanine scan										
7	<u>A</u>	K	L	K		A	K	K	R	4737
8	K	<u>A</u>	L	K		A	K	K	R	1421
9	K	K	<u>A</u>	K		A	K	K	R	231
10	K	K	L	<u>A</u>		A	K	K	R	2842
11	K	K	L	K		A	<u>A</u>	K	R	9421
12	K	K	L	K		A	K	<u>A</u>	R	5316
13	K	K	L	K		A	K	R	<u>A</u>	4632
Proline scan										
14	K	<u>P</u>	L	K		A	K	K	R	1947
15	K	K	<u>P</u>	K		A	K	K	R	5894
16	K	K	↓L	<u>P</u>		–	A	K	R	579
17	K	K	↓L	K		–	<u>P</u>	K	R	842
18	K	K	L	K		A	<u>P</u>	K	R	3421
19	K	K	L	K		A	K	<u>P</u>	R	9053
P ₁ substitutions										
20	K	K	L	<u>F</u>		A	K	K	R	9583
21	K	K	L	<u>C</u> *		A	K	K	R	8368
22	K	K	L	<u>H</u>		A	K	K	R	7684
23	K	K	L	<u>L</u>		A	K	K	R	6500
24	K	K	L	<u>Y</u>		A	K	K	R	6167
25	K	K	L	<u>M</u>		A	K	K	R	4210
26	K	K	L	<u>W</u>		A	K	K	R	3833
27	K	K	L	<u>R</u>		A	K	K	R	2421
28	K	K	L	<u>V</u>		A	K	K	R	1583
29	K	K	L	<u>I</u>		A	K	K	R	917

* Cys(tBu).

cysteine, threonine or serine side chains, and demonstrated that ϵ -NH₂-Cap-Leu-Cys(Bzl)-MCA was a highly susceptible and specific substrate for papain in comparison with chymotrypsin, and was resistant to trypsin [13].

Table 3 shows the kinetic parameters for the hydrolysis by cruzipain and cathepsin L of the series ϵ -NH₂-Cap-Leu-Xaa-

MCA. Cruzipain hydrolysed the substrates of this series more efficiently than did cathepsin L, and both enzymes accepted Bzl groups well, depending on the amino acid to which this group was attached. The substrate containing Thr(Bzl) (peptide 32) was hydrolysed with the lowest K_m values in this series by both cruzipain and cathepsin L. The presence of Cys(Bzl) increased the susceptibility of the substrate to cleavage, particularly with cruzipain, which hydrolysed ϵ -NH₂-Cap-Leu-Cys(Bzl)-MCA (peptide 30) with the highest k_{cat} value. It is interesting to note the difference of one order of magnitude in the k_{cat} and K_m values for hydrolysis by cruzipain of peptides 30 and 31, for which the structural difference corresponds to substitution of oxygen with sulphur. For comparison, ϵ -NH₂-Cap-Leu-Xaa-MCA was extended for Xaa = Phe, Tyr(Me), Arg and Gly. Cruzipain and cathepsin L accepted this substrate containing Phe (peptide 33) well; however, the methoxide group in the *para* position (peptide 34) increased the K_m value for cruzipain and decreased the k_{cat} value for cathepsin L. The substrate containing Arg (peptide 35) was poorer than those containing Cys(Bzl), Thr(Bzl) or Phe with both enzymes, and that containing Gly (peptide 36) was the poorest substrate in the series for both enzymes. These data contrast with those obtained on hydrolysis of these peptides with papain, which hydrolyses ϵ -NH₂-Cap-Leu-Arg-MCA more efficiently than any other peptide in this series [13].

The series Abz-KLXFSKQ-EDDnp was synthesized, using for X the same amino acids employed in the series ϵ -NH₂-Cap-Leu-Xaa-MCA. Phe was employed at the P'₁ position, because a hydrophobic amino acid at this position improved hydrolysis by cruzipain (results not shown); in addition, a series of substrates with Phe at the P'₁ position was observed to be more susceptible to papain and cathepsin B [32]. Furthermore, we intended to check whether the benzyl group of Phe at the P'₁ position would interfere with those attached to the side chains of Cys(Bzl), Thr(Bzl) or Ser(Bzl), which were claimed occupy the S'₁ site in the study of the cathepsin B–benzyloxycarbonyl-Arg-Ser(O-Bzl)-chloromethyl ketone complex mentioned above [31]. On the other hand, Ser was used at P'₂ because of the very favourable effect of its small side chain, as for Ala in the hydrolysis of peptide 11 (Table 2) by cruzipain, as well as by papain [24]. Furthermore, Ser was chosen instead of Ala in order to guarantee greater solubility for the peptides of this series.

Table 4 shows the kinetic parameters for the hydrolysis of this series of internally quenched fluorogenic peptides. Cruzipain and cathepsin L hydrolysed these peptides with specificity constant (k_{cat}/K_m) values one or two orders of magnitude higher than

Table 3 Kinetic parameters for the hydrolysis by cruzipain and cathepsin L of fluorogenic peptides containing the general sequence ϵ -NH₂-Cap-Leu-Xaa-MCA with substitutions at the P₁ position

Cruzipain was assayed at 37 °C in 100 mM NaH₂PO₄, 10 mM EDTA, 5 mM DTT, 400 mM NaCl and 2 mM Cys, pH 6.3. Cathepsin L was assayed in 50 mM NaH₂PO₄, 5 mM EDTA, 5 mM DTT, 200 mM NaCl and 2 mM Cys, pH 5.8.

Peptide no.	Xaa	Cruzipain			Cathepsin L		
		K_m (μM)	k_{cat} (s ^{–1})	k_{cat}/K_m (mM·s) ^{–1}	K_m (μM)	k_{cat} (s ^{–1})	k_{cat}/K_m (mM·s) ^{–1}
30	Cys(Bzl)	1.4	7.6	5429	2.1	0.5	238
31	Ser(Bzl)	14.0	0.6	43	5.4	0.05	9
32	Thr(Bzl)	0.3	0.8	2667	0.3	0.15	500
33	Phe	0.9	0.6	667	2.2	0.5	227
34	Tyr(Me)	16.2	2.2	136	3.0	0.07	23
35	Arg	3.4	1.2	353	5.1	0.4	78
36	Gly	11.0	0.5	45	26.9	0.1	4

Table 4 Kinetic parameters for the hydrolysis by cruzipain and cathepsin L of fluorogenic peptides containing the general sequence Abz-KLXFSKQ-EDDnp with substitutions at the P₁ position

Cruzipain was assayed at 37 °C in 100 mM NaH₂PO₄, 10 mM EDTA, 5 mM DTT, 400 mM NaCl and 2 mM Cys, pH 6.3. Cathepsin L was assayed in 50 mM NaH₂PO₄, 5 mM EDTA, 5 mM DTT, 200 mM NaCl and 2 mM Cys, pH 5.8.

Peptide no.	Xaa	Cruzipain			Cathepsin L		
		K _m (μM)	k _{cat} (s ⁻¹)	k _{cat} /K _m (mM·s) ⁻¹	K _m (μM)	k _{cat} (s ⁻¹)	k _{cat} /K _m (mM·s) ⁻¹
37	Cys(Bzl)	0.4	20.1	50 250	1.8	12.5	6 944
38	Ser(Bzl)	0.4	0.8	2 000	0.5	0.3	600
39	Thr(Bzl)	0.2	1.6	8 000	1.3	0.3	231
40	Phe	1.6	9.8	6 125	0.3	8.1	27 000
41	Arg	0.4	18.1	45 250	0.8	11.5	14 375
42	Gly	0.6	1.8	3 000	0.7	0.5	714

those for peptidyl-MCA substrates. More specifically, noteworthy 184- and 128-fold increases in the k_{cat}/K_m value were observed for the hydrolysis by cathepsin L and cruzipain respectively of Abz-KLRFSKQ-EDDnp (peptide 41) in comparison with ϵ -NH₂-Cap-Leu-Arg-MCA (peptide 35). A second significant increase in the k_{cat}/K_m value was observed when comparing the substrates containing Gly (peptides 36 and 42). The only exception was the hydrolysis by cathepsin L of the peptide containing Thr(Bzl) (peptide 39), for which the k_{cat}/K_m value was lower than for the corresponding peptidyl-MCA substrates (peptide 32). These data show that the amino acid residues at positions P'₁-P'₄, and EDDnp at C-terminal carboxyl site, have striking effects on the specificity and on the rate of hydrolysis by both enzymes, in comparison with MCA-containing substrates. Comparatively, the effects of Thr(Bzl), Ser(Bzl) and Cys(Bzl) on the susceptibility to cruzipain of the internally quenched fluorescent substrates are similar to those observed with the MCA-containing peptides. Abz-KLC(Bzl)FSKQ-EDDnp (peptide 37) and Abz-KLFFSKQ-EDDnp (peptide 40) are the best substrates for cruzipain and cathepsin L respectively, followed by the peptide that contains Arg (peptide 41). Whereas cathepsin L was found to prefer Phe at position P₁ of the internally quenched fluorescent peptides, cruzipain preferred Arg of the encoded amino acids. Therefore, with the Abz-peptidyl-EDDnp series, cruzipain behaves in a manner similar to that in the solid-phase assay, in which basic residues were also preferred by the parasite enzyme. On the other hand, these data contrast with those obtained with the series generated from the lead peptide Abz-KKLKAKKRY(NO₂)D-OH, where peptides containing Phe, Leu or Tyr at position P₁ are preferentially hydrolysed (Table 2). These differences are probably due to the non-additivity of interactions between each single enzyme subsite with a specific amino acid, as is also indicated by the subsite compensation observed [23].

In conclusion, some good substrates for cruzipain have been synthesized, and the subsite specificity has been analysed both in solution and in solid-phase assays. There seems to be a dual specificity in S₁ of the isolated cruzipain, and the subsite specificity is highly interdependent. It is therefore relevant to use combinatorial methods, where the substrates must compete for the enzyme, in order to assess the activity of the proteinase under conditions resembling those experienced by the parasite during invasion. We have observed a clear difference in terms of specificity between cruzipain and human cathepsin L with regard to the accommodation of Pro at the P₂ position. The results of

the present study are being used as a starting point in the design of inhibitor libraries for cruzipain.

We thank Kirsten Lilja for her excellent technical assistance in the library and multiple-column synthesis. This work was supported by the Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP), Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) and Financiadora de Estudos e Projetos (FINEP) through the PADCT programme.

REFERENCES

- McKerrow, J. H. (1989) Exp. Parasitol. **68**, 111–115
- Bontempi, E., Franke de Cazzulo, B. M., Ruiz, A. M. and Cazzulo, J. J. (1984) Comp. Biochem. Physiol. **77B**, 599–604
- Bontempi, E., Martinez, J. and Cazzulo, J. J. (1989) Mol. Biochem. Parasitol. **33**, 43–48
- Murta, A. C. M., Persechini, P. M., Padron, T. S., Souza, V., Guimarães, J. and Scharfstein, J. (1990) Mol. Biochem. Parasitol. **43**, 27–38
- Scharfstein, J., Schechter, M., Senna, M., Peralta, J. M., Medonça-Previato, L. and Milles, M. A. (1986) J. Immunol. **137**, 1336–1341
- Eakin, A. E., Mills, A. A., Harth, G., McKerrow, J. H. and Craik, C. S. (1992) J. Biol. Chem. **267**, 7411–7420
- McGrath, M. E., Eakin, A. E., Engel, J. C., McKerrow, J. H., Craik, C. S. and Fletterick, R. J. (1995) J. Mol. Biol. **247**, 251–259
- Aslund, L., Henriksson, J., Campetella, O., Frasch, A. C. C., Petersson, U. and Cazzulo, J. J. (1991) Mol. Biochem. Parasitol. **45**, 345–347
- Meirelles, M. N., Juliano, L., Carmona, E., Silva, S. G., Costa, E. M., Murta, A. C. M. and Scharfstein, J. (1992) Mol. Biochem. Parasitol. **52**, 175–184
- Serveau, C., Lalmanach, G., Juliano, M. A., Scharfstein, J., Juliano, L. and Gauthier, F. (1996) Biochem. J. **313**, 951–956
- Lam, K. S., Salmon, S. E., Hersh, E. M., Hruby, V. J., Kazmierski, W. M. and Knapp, R. J. (1991) Nature (London) **354**, 82–84
- Meldal, M. and Breddam, K. (1991) Anal. Biochem. **195**, 141–147
- Alves, L. C., Almeida, P. C., Franzoni, L., Juliano, L. and Juliano, M. A. (1996) Peptide Res. **9**, 92–96
- Chagas, J. R., Juliano, L. and Prado, E. S. (1990) Anal. Biochem. **192**, 419–425
- Chagas, J. R., Portaro, F. C. V., Hirata, I. Y., Almeida, P. C., Juliano, M. A., Juliano, L. and Prado, E. S. (1995) Biochem. J. **306**, 63–69
- Del Nery, E., Chagas, J. R., Juliano, M. A., Prado, E. S. and Juliano, L. (1995) Biochem. J. **312**, 233–238
- Meldal, M. (1992) Tetrahedron Lett. **33**, 3077–3080
- Auzanneau, F. I., Meldal, M. and Bock, K. (1995) J. Peptide Sci. **1**, 31–44
- Lima, A. P. C. A., Scharfstein, J., Storer, A. C. and Ménard, R. (1992) Mol. Biochem. Parasitol. **56**, 335–338
- Barret, A. J., Kempthorne, A. A., Brown, M. A., Kirschke, H., Knight, C. G., Masaharu, T. and Hanada, K. (1982) Biochem. J. **201**, 189–198
- Hirata, Y., Cezari, M. H. S., Nakaie, C. R., Boschov, P., Ito, A. S., Juliano, M. A. and Juliano, L. (1994) Lett. Peptide Sci. **1**, 299–308

-
- 22 Meldal, M., Svendsen, I., Breddam, K. and Auzanneau, F. I. (1994) *Proc. Natl. Acad. Sci. U.S.A.* **91**, 3314–3318
- 23 Grøn, H. and Breddam, K. (1992) *Biochemistry* **31**, 8967–8971
- 24 García-Echeverría, C. and Rich, D. H. (1992) *Biochem. Biophys. Res. Commun.* **187**, 615–619
- 25 Schechter, I. and Berger, A. (1967) *Biochem. Biophys. Res. Commun.* **27**, 157–162
- 26 Serveau, C. (1995) Ph.D. Thesis, Université François Rebelais, Tours, France
- 27 Brömme, D., Steinert, A., Friebe, S., Fittkau, S., Wiederanders, D. and Kirschke, H. (1989) *Biochem. J.* **264**, 475–481
- 28 Juliano, L. and Paiva, A. C. M. (1974) *Biochemistry* **13**, 2445–2450
- 29 Gray, C. F., Boukouvelas, J., Szawelski, R. J. and Wharton, C. W. (1984) *Biochem. J.* **219**, 325–328
- 30 Turk, D., Podobnik, M., Popovic, T., Katunuma, N., Bode, W., Huber, R. and Turk, V. (1995) *Biochemistry* **34**, 4791–4797
- 31 Jia, Z., Hasnain, S., Hiramata, T., Lee, X., Mort, S. J., To, R. and Huber, C. P. (1995) *J. Biol. Chem.* **270**, 5527–5533
- 32 Ménard, R., Carmona, E., Plouffe, C., Brömme, D., Konishi, Y., Lesebvre, J. and Storer, A. C. (1993) *FEBS Lett.* **328**, 107–110
-

Received 10 July 1996/9 December 1996; accepted 12 December 1996