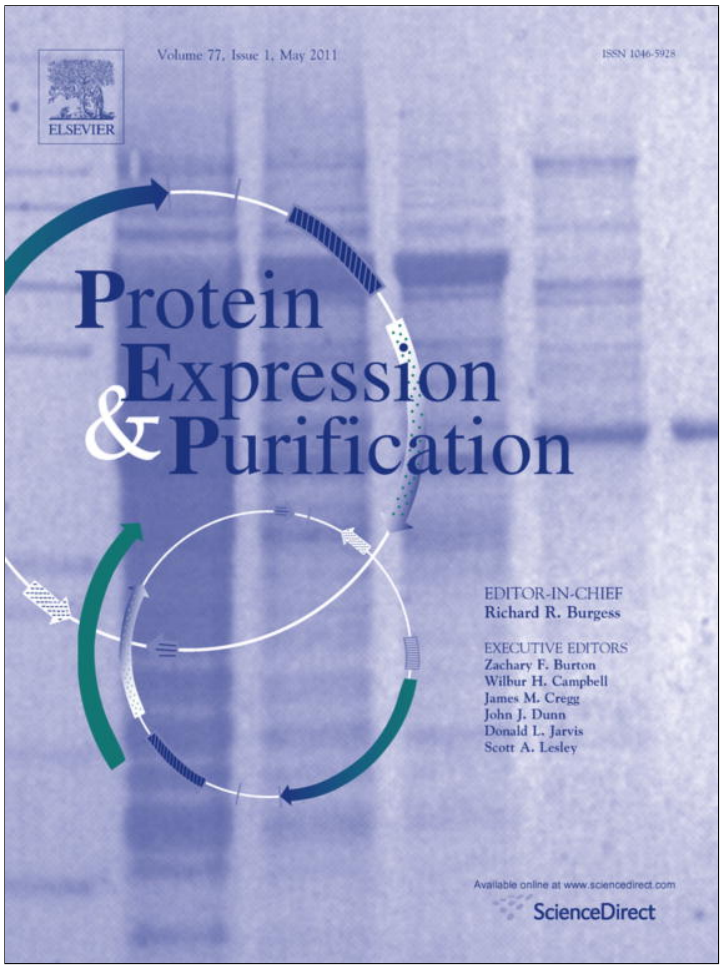

Protein Expression and Purification 77 (2011) 53–61

Contents lists available at ScienceDirect

Protein Expression and Purification

journal homepage: www.elsevier.com/locate/yprep

The substrate specificity of *Metarhizium anisopliae* and *Bos taurus* carboxypeptidases A: Insights into their use as tools for the removal of affinity tags

Brian P. Austin [a], József Tözsér [b], Péter Bagossi [b], Joseph E. Tropea [a], David S. Waugh [a,*]

[a] Macromolecular Crystallography Laboratory, Center for Cancer Research, National Cancer Institute at Frederick, P.O. Box B, Frederick, MD 21702-1201, USA
[b] Department of Biochemistry and Molecular Biology, Faculty of Medicine, University of Debrecen, Debrecen, Hungary

ARTICLE INFO

Article history:
Received 15 October 2010
and in revised form 5 November 2010
Available online 10 November 2010

Keywords:
Carboxypeptidase
Affinity tag
Polyhistidine tag
His-tag
Protease

ABSTRACT

Carboxypeptidases may serve as tools for removal of C-terminal affinity tags. In the present study, we describe the expression and purification of an A-type carboxypeptidase from the fungal pathogen *Metarhizium anisopliae* (MeCPA) that has been genetically engineered to facilitate the removal of polyhistidine tags from the C-termini of recombinant proteins. A complete, systematic analysis of the specificity of MeCPA in comparison with that of bovine carboxypeptidase A (BoCPA) was carried out. Our results indicate that the specificity of the two enzymes is similar but not identical. Histidine residues are removed more efficiently by MeCPA. The very inefficient digestion of peptides with C-terminal lysine or arginine residues, along with the complete inability of the enzyme to remove a C-terminal proline, suggests a strategy for designing C-terminal affinity tags that can be trimmed by MeCPA (or BoCPA) to produce a digestion product with a homogeneous endpoint.

Published by Elsevier Inc.

Introduction

Metallocarboxypeptidases are enzymes that have evolved to remove C-terminal amino acid residues of proteins or peptides with the aid of a Zn^{2+} ion. They are commonly regarded as being involved in the degradation of proteins and peptides, yet recent findings suggest a wide range of physiological roles for such enzymes [1]. Carboxypeptidases are classified on the basis of their specificity. For example, type A carboxypeptidases preferentially remove C-terminal amino acid residues having aromatic or branched aliphatic side chains whereas type B carboxypeptidases exhibit a strong preference for basic amino acids. The A-type carboxypeptidases are further sub classified into type A1 and type A2 isoforms in rodents and humans. Carboxypeptidase A1 preferentially catalyzes the removal of aliphatic residues from peptide substrates, while the A2 isoforms show higher specificity toward aromatic residues such as phenylalanine and tryptophan. Carboxypeptidase A2 is not present in the bovine pancreas. Instead the single bovine carboxypeptidase A (BoCPA)[1] has relatively broad substrate specificity [2].

The purification of a novel, type A carboxypeptidase (MeCPA) from the fungal entomopathogen *Metarhizium anisopliae* has been reported, and its specificity was compared to that of BoCPA using a limited set of substrates [3,4]. The catalytic mechanism of MeCPA was originally uncertain because its activity was inhibited by both di-isopropyl fluorophosphates (DIFP) and 1,10-phenanthroline [4], but later it was established to be a zinc carboxypeptidase [3]. Like animal type A carboxypeptidases and in contrast to bacterial carboxypeptidases, MeCPA lacks type B specificity.

Current strategies for recombinant protein expression frequently involve the use of affinity tags, often joined to the N-terminus of the protein being expressed. The hexahistidine tag (His-tag) is far and away the most commonly used affinity tag [5–7] and one of the few tags that is frequently fused to the C-termini of recombinant proteins. Endoproteolytic removal of C-terminal His-tags is complicated by the fact that the principal specificity determinants of endoproteolytic enzymes (*e.g.*, Factor Xa, thrombin, enteropeptidase/enterokinase, tobacco etch virus protease) are located on the N-terminal side of the scissile bond. Consequently, the removal of a C-terminal tag by any of them would leave behind a number of non-native residues (six in the case of tobacco etch virus protease). One could argue, therefore, that any gain achieved by endoproteolytic removal of a C-terminal hexahistidine tag would be offset by the presence of the residual protease recognition site.

A promising alternative is exoproteolytic removal of a C-terminal His-tag by a carboxypeptidase, as demonstrated previously with BoCPA (*e.g.*, [8,9]). To refine this method, we engineered a recombinant form of MeCPA with a C-terminal hexahistidine tag followed by two arginine residues. We chose MeCPA because it

* Corresponding author. Fax: +1 301 846 7148.
E-mail address: waughd@mail.nih.gov (D.S. Waugh).

[1] *Abbreviations used:* BoCPA, bovine carboxypeptidase A; MeCPA, *Metarhizium anisopliae* carboxypeptidase A; DIFP, di-isopropyl fluorophosphates; ORF, open reading frame; PCR, polymerase chain reaction.

1046-5928/$ - see front matter Published by Elsevier Inc.
doi:10.1016/j.pep.2010.11.005

can be produced in baculovirus infected insect cells and because its amino acid sequence suggested that it might have even broader specificity than BoCPA [3,4]. The arginine residues were intended to prevent the enzyme from digesting its own C-terminal His-tag. The polyhistidine tag facilitates the purification of recombinant pro-MeCPA, which is secreted from insect cells, and also assists in its separation from the products of a carboxypeptidase digest. Here, we describe the cloning, expression, and purification of the active enzyme using the baculovirus expression system. We also compare the specificity of recombinant MeCPA to that of BoCPA, the most commonly used carboxypeptidase for research and biotechnological purposes, using an oligopeptide-based HPLC assay. Finally, we show that recombinant MeCPA is readily able to remove polyhistidine tags from the C-termini of globular proteins.

Materials and methods

Molecular modeling of MeCPA

A molecular model of MeCPA was built by Modeller [8] based on the structure of BoCPA (PDB code: 3CPA) [9]. A sequence alignment of MeCPA and BoCPA, performed with the ClustalW program [10], is presented in Fig. 1. The alignment was verified by comparison of carboxypeptidases with deposited structural coordinates (data not shown). Structures were examined on a Silicon Graphics Fuel workstation using Sybyl (Tripos, St. Louis, MO, USA).

BoCPA

BoCPA (Type II-PMSF, C-9268) was purchased from Sigma–Aldrich (St. Louis, MO, USA).

Cloning of the MeCPA gene

M. anisopliae (Metschnikoff) Sorokin mycelium was a gift from Dr. Richard A. Humber of the USDA-ARS Collection of Entomopathogenic Fungal Cultures (ARSEF) in Ithaca, NY, USA. Genomic DNA was isolated from the mycelium using a kit from Invitrogen (Carlsbad, CA, USA).

Oligodeoxyribonucleotide primers complementary to the 5′ (5′-GGGG ACA ACT TTG TAC AAA AAA GTT GTG ATG AGA GTG GTT GCT TTC TTC GCC TG-3′) and 3′ (5′-GGGG ACA ACT TTG TAC AAG AAA GTT GCA CTC ATC TGC TGG AAG AGA TGC ATG G-3′) ends of the MeCPA cDNA sequence reported by Joshi and St Leger [4] with the addition of terminal attB1 and attB2 recombination sites were used to generate an amplicon by polymerase chain reaction (PCR) that was subsequently inserted by Gateway recombinational cloning into pDONR201 and sequenced in its entirety (Genebank accession code: EU919684). The genomic clone contained five exons, which were subsequently joined together by overlap extension PCR [11] to assemble the complete, uninterrupted pre-pro-MeCPA open reading frame (ORF) in the donor vector pDONR223 (Invitrogen). In the process, a DNA sequence encoding the residues His–His–His–His–His–His–Arg–Arg–STOP was added in-frame to the 3′ end of the final exon. The MeCPA ORF and the C-terminal His_6Arg_2 tag were subsequently inserted via the LR reaction into the insect cell transfer vector pDEST-8 (Invitrogen), which contains the baculovirus polyhedrin promoter, yielding baculovirus expression clone 1177-X1-8.

Construction of recombinant baculovirus and optimization of MeCPA production

The baculovirus expression clone 1177-X1-8 was introduced into DH10Bac cells (Invitrogen), and the transformants were selected on

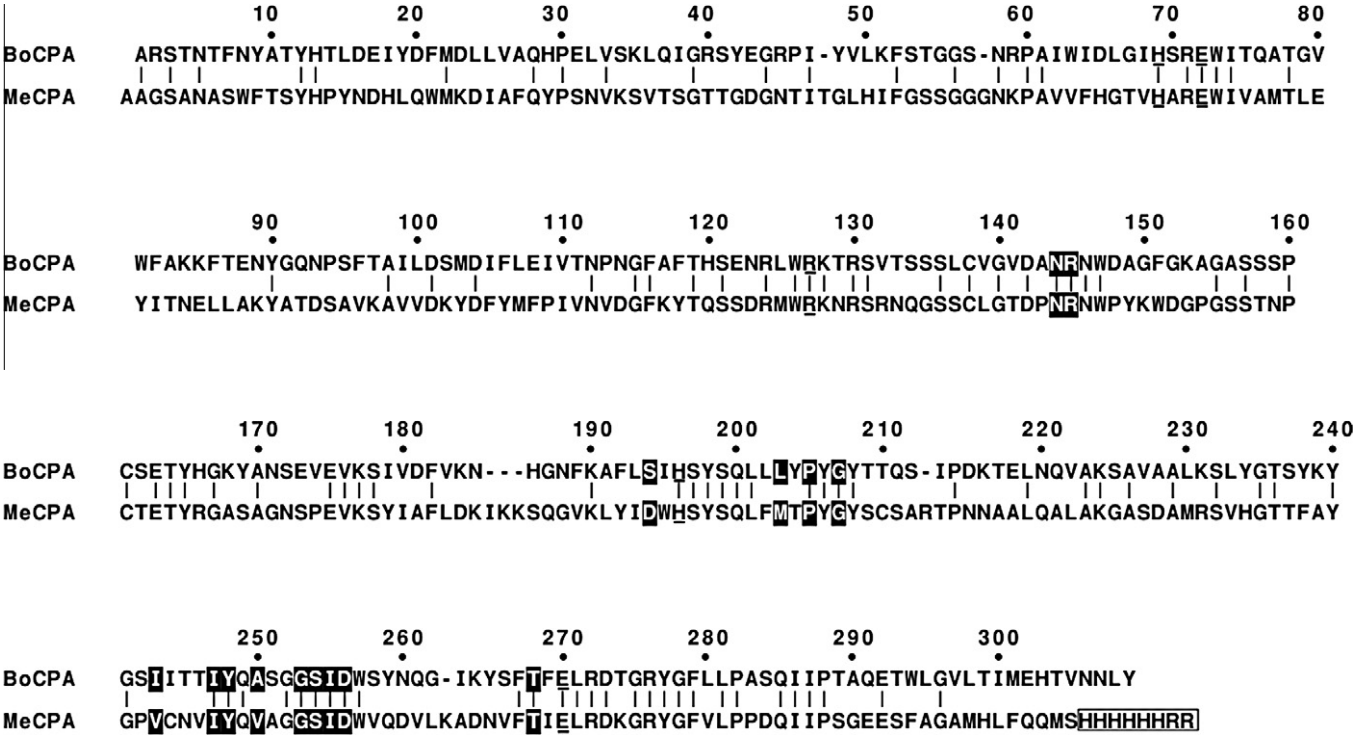

Fig. 1. Sequence alignment of bovine carboxypeptidase A (BoCPA) and *Metarhizium anisopliae* carboxypeptidase A (MeCPA). The sequence alignment is part of a multiple sequence alignment of carboxypeptidases made by ClustalW. Active site and Zn-coordinating residues are underlined. Residues forming the S1′ binding site are indicated in reverse-bold lettering. The recombinant form of MeCPA included an HHHHHHRR C-terminal sequence tag (boxed). The N-terminus of "mature" MeCPA shown here was generated by digestion of pro-MeCPA with thermolysin and verified by N-terminal amino acid sequencing (data not shown).

plates containing gentamycin, kanamycin and tetracycline with IPTG and X-gal. White colonies (signifying proper transposition of the expression clone into the baculovirus genome) were selected, grown in liquid media, and bacmid DNA was prepared by gentle alkaline lysis miniprep. Bacmid DNA was verified by two separate PCR reactions designed to amplify the flanking regions of the expression cassette. A recombinant baculovirus was generated from the bacmid DNA and a high-titer stock was prepared according to the instructions in the Bac-to-Bac product manual (Invitrogen). Sf9, Sf21 and Hi5 cells were cultivated in serum-free medium and infected with the high-titer stock at various multiplicities of infection. Samples of the conditioned medium were sampled after 0, 24, 48, 72, and 96 h and then analyzed by SDS-PAGE and Western blotting with an anti-His_6 antibody (Pierce Biotechnology, Woburn, MA, USA).

Purification and activation of MeCPA

All purification steps were carried out at 4 °C. Twenty liters of fresh Hi5 insect cell culture media containing secreted pro-MeCPA was diafiltered and concentrated to a volume of 2 L in 50 mM sodium phosphate pH 8.0, 150 mM NaCl, and 25 mM imidazole (buffer A). The concentrate was clarified by centrifugation, filtered through a 0.45 μm cellulose acetate membrane (Corning Incorporated, Corning, NY, USA) and applied to a 5 mL HisTrap column (GE Healthcare Bio-Sciences Corp., Piscataway, NJ, USA) equilibrated in buffer A. The column was washed to baseline and eluted with a 25–250 mM imidazole gradient in buffer A over five column volumes. The eluted pro-MeCPA was concentrated using an Amicon stirred cell and a YM-10 membrane (Millipore Corp., Bedford, MA, USA) to a volume of 5 mL, and applied to a HiPrep 26/60 Sephacryl S-100 gel filtration column (GE Healthcare Bio-Sciences) equilibrated with 25 mM MES pH 6.0, 150 mM NaCl (buffer B). Fractions containing pro-MeCPA were pooled and concentrated to 1 mg/mL. pro-MeCPA was activated using thermolysin at a 1:10 M ratio of enzyme to substrate for 2 h at room temperature. The digest was dialyzed against 20 mM MES at pH 6.0 and applied to a 5 mL HiTrap SP HP cation exchange column (GE Healthcare Bio-Sciences Corp.) equilibrated in the same buffer to remove the thermolysin enzyme. Activated MeCPA was eluted with a NaCl gradient from 0 to 2 M in buffer B over 20 column volumes. Peak fractions containing active MeCPA were pooled and concentrated to 1 mg/mL as above. The protein was reapplied to a Sephacryl S-100 sizing column equilibrated in buffer B and four peak fractions corresponding to MeCPA were pooled and concentrated as described above. Ultimately, about 0.2 mg of pure MeCPA was obtained from each liter of conditioned medium (Table 1). Half of the protein was stored at 4 °C and the remainder was flash frozen in liquid nitrogen and stored at −80 °C.

Determination of pH and salt optima

The optimum pH for MeCPA was determined by preparing three buffering systems over a pH range of 4.5–11.5 at 0.5 pH unit intervals. 2-(N-morpholino)ethanesulfonic acid (MES) buffer was used from pH 4.5 to 6.5, Tris (hydroxymethyl) aminomethane (Tris) was used from pH 7.0 to 8.5, and N-cyclohexyl-2-aminoethanesulfonic acid (CHES) buffer was used from pH 8.5 to 11.5. The commonly used carboxypeptidase substrate N-(3-(2-furyl)acryloyl)-l-phenylalanyl-phenylalanine (FAPP) was dissolved in each buffer at a concentration of 0.2 mM and used to monitor enzyme activity over time spectrophotometrically (Beckman DU600 spectrophotometer) at 340 nm after the addition of 50 ng of MeCPA to a 100 μL reaction volume (15 nM MeCPA) [12]. The decrease in absorbance over time was calculated over the linear portion of the curve.

To determine the optimum salt concentration, sodium chloride was added to 25 mM Tris–HCl pH 7.5, from 0 to 5 M and the enzyme activity was determined using the FAPP assay noted above.

Table 1
Purification of MeCPA.

Purification step	Total protein (mg/L)	Yield (%)	Enzyme activity (%)
Conditioned medium[a]	13.57	100	0
Affinity (NiNTA)	2.85	21	0
Thermolysin Activation	2.85	21	100
Cation exchange (SP)	1.29	9.5	100
Size exclusion (S100)	0.19	1.4	100

[a] 20 L.

Oligopeptide synthesis and characterization

Oligopeptides were obtained from SynPep (Dublin, CA, USA) with free carboxylate ends. Stock solutions were made in distilled water (or in 10 mM DTT in case of Cys- or Met-containing peptides) and the peptide concentrations were determined by amino acid analysis after peptide hydrolysis using a Beckman 6300 amino acid analyzer.

Enzyme kinetics

The carboxypeptidase assays were initiated by mixing 10 μl of carboxypeptidase solution that had been diluted at least 100-fold with 2× reaction buffer (50 mM Tris–HCl pH 8.2, 300 mM NaCl) with 10 μl of substrate solution (0.05–2.0 mM, the actual range was selected based on the approximate K_m values). The enzyme concentrations were determined by amino acid analysis. Measurements were performed at six different substrate concentrations. The reaction mixtures were incubated at 37 °C for 10 min and then stopped by the addition of 180 μl of 4.5 M guanidine–HCl containing 1% trifluoroacetic acid (TFA). An aliquot was injected onto a Nova-Pak C18 reversed-phase chromatography column (3.9 × 150 mm, Waters Corp., Milford, MA, USA) using an automatic injector. Substrates and cleavage products were separated using an increasing water–acetonitrile gradient (0–100%) in the presence of 0.05% TFA. To determine the correlation between peak areas of the cleavage products and their abundance, fractions were collected and analyzed by amino acid analysis. The k_{cat} values were calculated by assuming 100% activity for the enzymes. Kinetic parameters were determined by fitting the data obtained at less than 20% substrate hydrolysis to the Michaelis–Menten equation by using the Fig. P program (Fig. P Software Corp., Hamilton, ON, Canada). The standard deviations for the k_{cat}/K_m values were calculated as described [13]. If no saturation was obtained in the studied concentration range, the k_{cat}/K_m value was determined from the linear part of the rate versus the concentration profile.

Digestion of protein substrates

Protein substrates were diluted to 1 mg/mL in 25 mM Tris–HCl pH 7.5, 100 mM NaCl. MeCPA was added to approximate an enzyme to substrate molar ratio of 1:20. The final volume of each digest was 20 μL. The digests were incubated at room temperature for three hours and the reactions were quenched by the addition of 2 μL 0.5 M ethylenediaminetetraacetic acid (EDTA). Each substrate was digested three times on different days.

Following digestion, 20 μL of 2× mass spec buffer (20% methanol, 10% acetic acid) was added to each digest and analyzed by electrospray mass spectrometry. The protein sequence was input into the peptide tools function of the LC/MSD software package

from Agilent Technologies (Rev. A 09.01 [1206]) and each mass was checked against the protein sequence for fragments that corresponded to a substrate digest product. Masses that were within ±3 Da to a carboxyterminal digestion product were used to determine the extent of proteolysis. The percent relative abundance for all possible carboxypeptidase digest products was then normalized such that the sum of all products was 100%.

Results

Comparison of the molecular model of MeCPA with the structure of BoCPA

The predicted 3D structure of MeCPA is very similar to that of BoCPA (not shown). Sequence alignment of the two polypeptides (Fig. 1) indicates that the residues involved in catalysis and zinc binding are identical in both enzymes. Using the BoCPA numbering, these residues are His69, Glu72, Arg127, His196 and Glu270. Furthermore, comparison of the sequences and the 3D structures revealed that the substrate binding region is also similar in the two enzymes: there are 11 identical and only four different residues in the S1 binding site. Identical residues are Asn144, Arg145, Pro205, Gly207, Ile247, Tyr248, Gly253, Ser254, Ile255, Asp256 and Thr268. The four differing residues are Ser194/Asp, Leu203/Met, Ile243/Val and Ala250/Val in BoCPA and MeCPA, respectively (Fig. 2). BoCPA contains a single disulfide bond between Cys138 and Cys161, and both residues are also present in MeCPA. A second pair of cysteines in MeCPA (located at positions 210 and 244 relative to the BoCPA sequence) are absent in BoCPA but may very well form a second disulfide bond in MeCPA [3], since they are located in close proximity to one another in the molecular model of the latter enzyme. In this respect the structure of MeCPA resembles those of the type A2 carboxypeptidases [14], whereas BoCPA is more closely related to type A1 enzymes.

Characterization of MeCPA produced in baculovirus

The formation of two disulfide bonds had been previously predicted [3] and therefore the baculovirus expression system was selected for the production of recombinant MeCPA to facilitate proper folding of the enzyme. Maximum production of secreted pro-MeCPA occurred in Hi5 cells after 72 h (data not shown). After activation and purification, the enzyme was greater than 90% pure as judged by SDS–PAGE (data not shown). Samples stored at either 4 °C or −80 °C had equivalent enzymatic activity over a period of greater than 6 months. The final yield of pure MeCPA was approximately 200 μg/L of conditioned medium.

The N-terminal sequences of both pro-MeCPA and the active enzyme were determined by the Edman degradation technique (data not shown). The N-terminus of pro-MeCPA was heterogeneous. In contrast to what was predicted by Joshi and St Leger [3], the major species resulted from cleavage of the pre-pro-enzyme by signal peptidase between residues 16 and 17, with some cleavage also occurring between residues 17 and 18. Also at odds with the previous report [3], we observed little or no autoprocessing of the zymogen in the conditioned medium. Therefore, a series of proteases were tested to determine whether any of them would promote the conversion of pro-MeCPA into the active enzyme (data not shown). Among these, thermolysin was found to be very effective. Treatment of the zymogen with thermolysin gave rise to a stable digestion product comprised of 321 amino acids (residues 106–418 of MeCPA plus His_6Arg_2) that was highly resistant to further degradation.

Initially, MeCPA activity was monitored spectrophotometrically at 340 nM, using the substrate N-(3-(2-furyl)acryloyl)-l-phenylalanyl-phenylalanine (FAPP). The decrease in absorbance over time was calculated over the linear portion of the curve. The enzyme exhibited activity over a wide pH range of 6–9 (Fig. 3A) with an optimum pH of approximately 7.5. The enzyme also displayed a tolerance for high monovalent salt concentrations (up to at least 4 M), although its activity was found to be greatest in the absence of salt (Fig. 3B). Substantial enzymatic activity was observed at physiological salt concentration (ca. 150 mM).

Comparison of the specificity of MeCPA and BoCPA using oligopeptide substrates

To compare the specificity of the two carboxypeptidases, we developed an HPLC assay system using the oligopeptide VSQNPKX as a framework, wherein X was the variable amino acid residue. Considering that both Pro and Lys are very unfavorable for type A carboxypeptidases (<http://expasy.org/prosite/PDOC00123>), only residue X was expected to be removed from the substrates. As shown for VSQNPKA (Fig. 4), the product peak VSQNPK is well separated from the substrate peak on the reversed-phase HPLC chromatogram. By determining the amount of product corresponding to the product peak, a correlation with the integration value was established that could be used to calculate velocity data and Michaelis–Menten kinetic parameters, which are presented in Table 2.

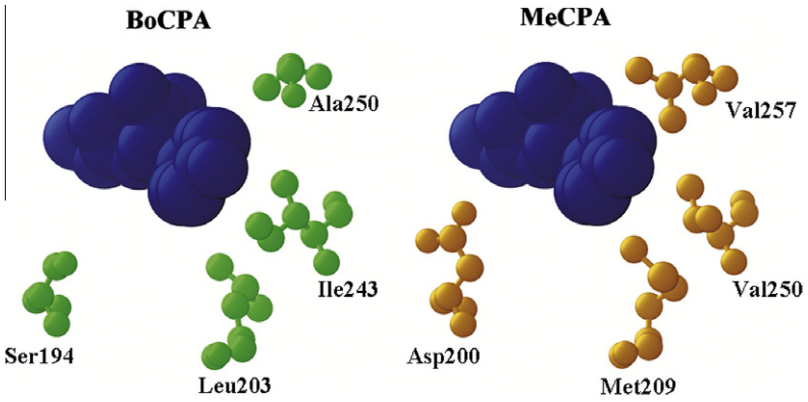

Fig. 2. Residues that differ in the substrate binding sites of BoCPA (left panel) and MeCPA (right panel). Residues of the substrate having Gly and Tyr at P1 and P1′, respectively, are shown in space-filling format (blue). Residues of BoCPA (green) and MeCPA (gold) are represented in ball-and-stick format. Structurally equivalent residues have different numbers due to gaps in the sequence alignment between BoCPA and MeCPA (Fig. 1).

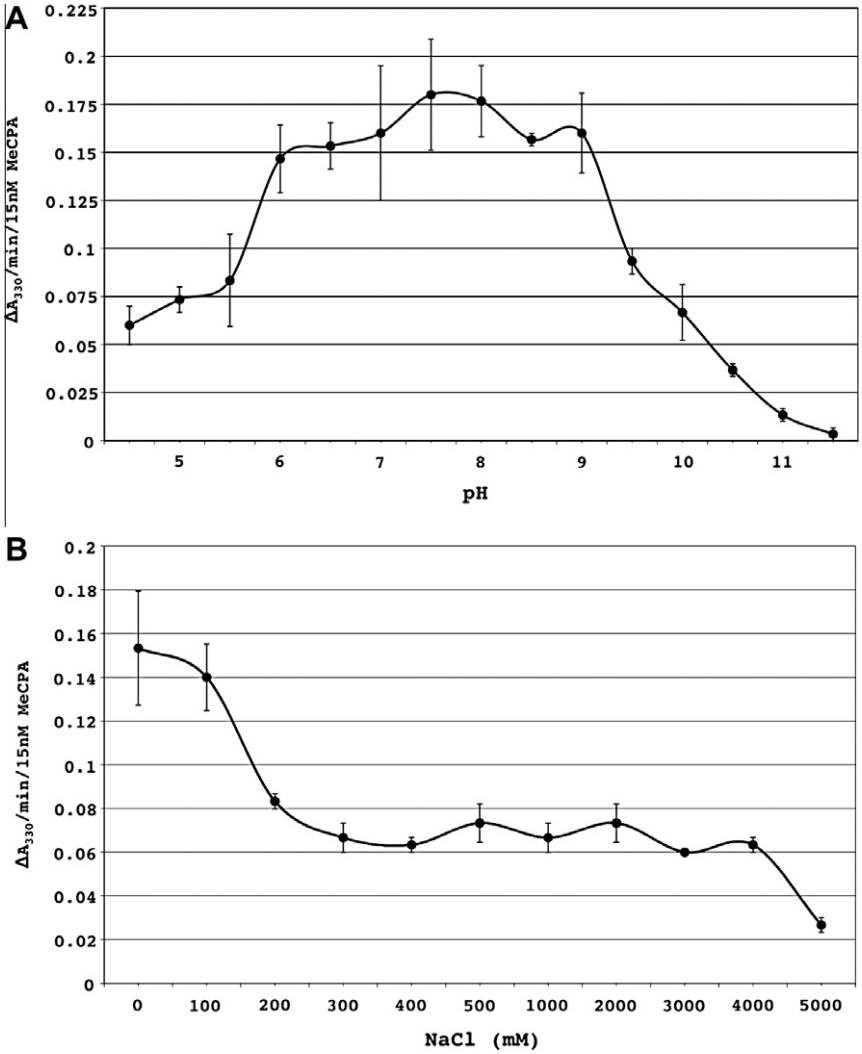

Fig. 3. Determination of optimum pH (A) and monovalent salt concentration (B) for MeCPA, measured by the decrease in absorbance at 340 nm using the FAPP assay ($N = 3$).

The substrate specificity of the two enzymes is very similar, with both of them exhibiting a preference for apolar residues, as expected for type A carboxypeptidases (Table 2). The four best substrates (having the highest k_{cat}/K_m values) and the three least susceptible peptides (being uncleavable under the conditions used) were the same for both enzymes; namely, peptides having F, Y, L, I or P, K, R residues in their P1′ sites, respectively. However, the rank order of the preferred residues is different (MeCPA: F > I > Y > L and BoCPA: F > Y > L > I). A somewhat different specificity profile was observed previously by Joshi and St Ledger using peptides containing only Phe, Tyr, Trp, Leu and Ala in the P1′ position [3]. In that study, while both enzymes appeared to prefer Phe at P1′, Trp was more readily removed by MeCPA than Tyr, and Leu was more readily removed by MeCPA than BoCPA [3]. These discrepancies may be due to the entirely different assay methods and conditions employed, including the pH (8.2 vs. 7.5), the length of the substrates, and the method of detection. A plot of $\ln(k_{cat}/K_m)$ values of MeCPA vs. $\ln(k_{cat}/K_m)$ values of BoCPA shows that the preference for apolar character is less pronounced for MeCPA than it is for BoCPA (Fig. 5). The values for bulky apolar residues are located close to the theoretical line for identical values, while polar residues form another line showing a higher reactivity with MeCPA. Intriguingly, the slope of the latter line was also very close to the theoretical value of 1. Histidine seems to be a moderately better substrate for MeCPA than BoCPA but is not removed as readily as most of the hydrophilic and polar residues.

Digestion of globular proteins with C-terminal polyhistidine tags

To investigate the plausibility of using MeCPA as a reagent for the removal of C-terminal polyhistidine tags, we next tested several globular proteins with C-terminal His_6 tags as substrates: *Yersinia pestis* SycE, SycH, and LcrV (Fig. 6). These proteins were selected because their crystal structures have been determined [15–17] and consequently each of them is known to have a partially disordered C-terminus. The carboxypeptidase digests were performed as described in Experimental Procedures and the product(s) were identified by electrospray mass spectrometry. When more than one digestion product was detected, their relative abundances were normalized to total 100 (%). Three independent experiments were performed under identical conditions on different days.

Digestion of SycE with BoCPA and MeCPA mainly produced homogeneous products lacking both the C-terminal His_6 tag and the last three residues of SycE (SFS), and terminating after the arginine closest to the C-terminus. However, both enzymes yielded a slight amount of underdigested substrate in one of the three experiments.

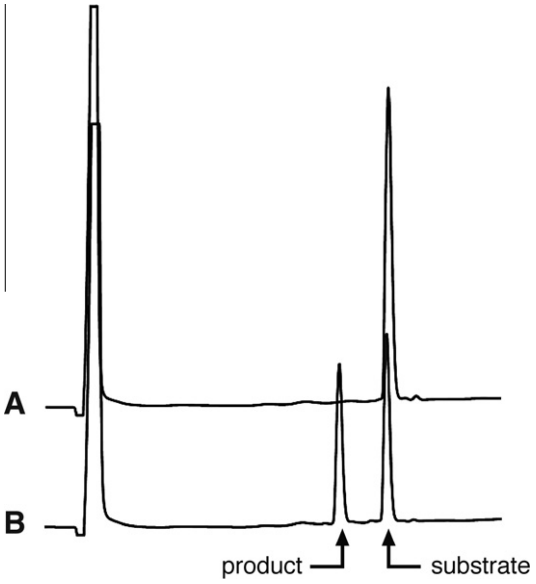

Fig. 4. Cleavage of the oligopeptide VSQNPKA by MeCPA. The substrate was incubated in the absence (A) or presence (B) of the enzyme as described in the Materials and Methods. After stopping the reaction, an aliquot was injected onto a reverse-phase HPLC column. The chromatograms produced by an acetonitrile gradient are shown.

In the case of SycH, all three experiments conducted with MeCPA produced homogeneous products in which the C-terminal His_6 tag and the last six residues of the native protein (SSHLLV) were removed. Interestingly, these products terminated with a serine residue even though the two consecutive serines immediately adjacent to this residue were removed. We interpret this observation as evidence that proline in the P1 position of the substrate is an impediment to digestion by MeCPA. The results obtained with BoCPA were less satisfactory, with all experiments retaining some undigested material and yielding a mixture of products. However, in no case did digestion proceed beyond the serine residue immediately adjacent to the proline, as was the case with MeCPA. Hence, proline in the P1 position also seems inhibitory to digestion by BoCPA. In general, SycH was underdigested by BoCPA in these experiments. Longer reaction times or higher enzyme concentrations would likely have given rise to more homogeneous products.

The opposite result was obtained when LcrV was digested with the two carboxypeptidases. In this case, all of the experiments with BoCPA produced a homogeneous product in which only the C-terminal His_6 tag was removed. This makes sense because the C-terminal residue of LcrV is a lysine, which is highly resistant to digestion by both enzymes (Table 1). In contrast, all the MeCPA digests yielded a mixture of products. Still, in each case the major product was the same as that generated by BoCPA. Here again, the problem seems to be underdigestion, although in experiment 2 there was evidence of overdigestion as well. We can think of no

Table 2
Comparison of the specificity of MeCPA and BoCPA. No cleavage was observed with the substrates having K, R or P at the C-terminus by either of the carboxypeptidases.

Peptide no.	Sequence	Enzyme	K_m (mM)	k_{cat} (s^{-1})	k_{cat}/K_m ($mM^{-1}\ s^{-1}$)
1.	VSQNPK**W**	MeCPA	0.255 ± 0.055	2.2 ± 0.07	8.7 ± 2.0
		BoCPA	0.494 ± 0.080	10.6 ± 0.7	21.4 ± 3.8
2.	VSQNPK**Y**	MeCPA	0.138 ± 0.023	23.6 ± 1.3	171.2 ± 30.0
		BoCPA	0.289 ± 0.055	26.1 ± 2.4	90.3 ± 19.0
3.	VSQNPK**F**	MeCPA	0.222 ± 0.013	98.0 ± 2.6	441.3 ± 28.4
		BoCPA	0.254 ± 0.034	41.8 ± 2.3	164.5 ± 23.8
4.	VSQNPK**L**	MeCPA	>1.0	N.D.	130.2 ± 12.1
		BoCPA	0.660 ± 0.079	51.8 ± 3.4	78.5 ± 10.8
5.	VSQNPK**I**	MeCPA	0.254 ± 0.042	66.6 ± 4.5	262.0 ± 46.9
		BoCPA	0.495 ± 0.057	13.5 ± 0.8	27.2 ± 3.5
6.	VSQNPK**M**	MeCPA	1.09 ± 0.12	32.7 ± 1.2	30.0 ± 3.5
		BoCPA	1.84 ± 0.40	6.4 ± 0.7	3.5 ± 0.8
7.	VSQNPK**V**	MeCPA	0.334 ± 0.045	34.7 ± 2.1	104.0 ± 15.3
		BoCPA	0.294 ± 0.031	7.2 ± 0.2	24.4 ± 2.7
8.	VSQNPK**C**	MeCPA	0.531 ± 0.058	18.3 ± 1.1	34.4 ± 4.3
		BoCPA	>2.0	N.D.	1.3 ± 0.1
9.	VSQNPK**A**	MeCPA	>1.0	N.D.	46.3 ± 5.8
		BoCPA	>1.0	N.D.	3.8 ± 0.2
10.	VSQNPK**T**	MeCPA	0.103 ± 0.010	0.79 ± 0.01	7.6 ± 0.8
		BoCPA	0.265 ± 0.035	0.11 ± 0.01	0.42 ± 0.06
11.	VSQNPK**S**	MeCPA	0.901 ± 0.157	3.08 ± 0.22	3.4 ± 0.6
		BoCPA	>2.0	N.D.	0.16 ± 0.01
12.	VSQNPK**H**	MeCPA	1.07 ± 0.17	4.7 ± 0.4	4.3 ± 0.4
		BoCPA	>1.0	N.D.	1.6 ± 0.2
13.	VSQNPK**Q**	MeCPA	1.57 ± 0.07	10.0 ± 0.21	6.4 ± 0.3
		BoCPA	>2.0	N.D.	0.61 ± 0.04
14.	VSQNPK**E**	MeCPA	>2.0	N.D.	0.19 ± 0.02
		BoCPA	>2.0	N.D.	0.013 ± 0.001
15.	VSQNPK**N**	MeCPA	0.181 ± 0.036	3.5 ± 0.3	19.5 ± 4.1
		BoCPA	0.444 ± 0.087	0.38 ± 0.04	0.86 ± 0.19
16.	VSQNPK**D**	MeCPA	>2.0	N.D.	0.27 ± 0.008
		BoCPA	>2.0	N.D.	0.009 ± 0.002
17.	VSQNPK**G**	MeCPA	1.39 ± 0.26	0.93 ± 0.08	0.67 ± 0.13
		BoCPA	>2.0	N.D.	0.054 ± 0.015

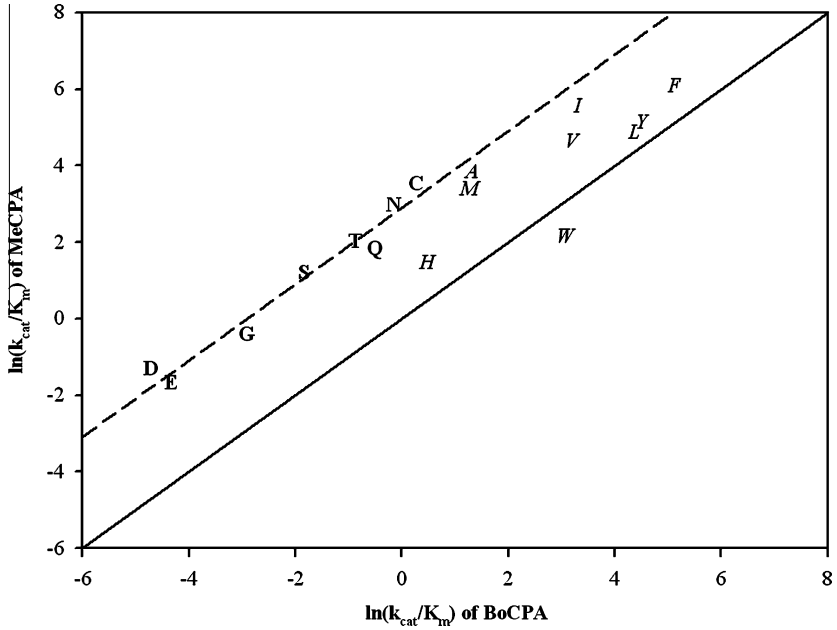

Fig. 5. Comparison of ln(k_{cat}/K_m) values for the digestion of oligopeptide substrates by MeCPA and BoCPA. The solid line indicates the theoretical position of the values if the specificity of the two enzymes was identical. The dashed line was obtained by fitting the values obtained for the substrates having P1′ amino acid residues indicated in bold.

***Y. pestis* SycE-His$_6$**

BoCPA	**1.**	LVQGAE/RLQTSSLISPPR$_{100}$SFSHHHHHH
	2.	LVQGAE/RLQTSSLISPPR$_{91}$SFSHHHHH$_9$H
	3.	LVQGAE/RLQTSSLISPPR$_{100}$SFSHHHHHH
MeCPA	**1.**	LVQGAE/RLQTSSLISPPR$_{100}$SFSHHHHHH
	2.	LVQGAE/RLQTSSLISPPR$_{100}$SFSHHHHHH
	3.	LVQGAE/RLQTSSLISPPR$_{93}$SFSHH$_8$HHHH

***Y. pestis* SycH-His$_6$**

BoCPA	**1.**	LS/CEPTMKKEEDDHRPS$_{42}$S$_{35}$S$_{18}$HLLVHHHHHH$_5$
	2.	LS/CEPTMKKEEDDHRPS$_{42}$S$_{35}$S$_{18}$HLLVHHHHHH$_5$
	3.	LS/CEPTMKKEEDDHRPS$_{81}$SSHLLVHHHHHH$_{19}$
MeCPA	**1.**	LS/CEPTMKKEEDDHRPS$_{100}$SSHLLVHHHHHH
	2.	LS/CEPTMKKEEDDHRPS$_{100}$SSHLLVHHHHHH
	3.	LS/CEPTMKKEEDDHRPS$_{100}$SSHLLVHHHHHH

***Y. pestis* LcrV-His$_6$**

BoCPA	**1.**	KYDSVMQRLLDD/TSGK$_{100}$HHHHHH
	2.	KYDSVMQRLLDD/TSGK$_{100}$HHHHHH
	3.	KYDSVMQRLLDD/TSGK$_{100}$HHHHHH
MeCPA	**1.**	KYDSVMQRLLDD/TSGK$_{37}$H$_9$H$_{54}$HHHH
	2.	KYD$_9$SV$_{11}$MQRLLDD/TSGK$_{67}$HH$_{13}$HHHH
	3.	KYDSVMQRLLDD/TSGK$_{95}$H$_5$HHHHH

Fig. 6. C-terminal sequences of globular protein substrates subjected to digestion by MeCPA and BoCPA. No electron density was observed in the crystal structures of these proteins (17–19) for residues following the forward slash mark, presumably because they are disordered. The numbers represent the relative abundances of the digestion products terminating at these locations normalized to total 100%.

straightforward explanation for the latter observation, which leads us to believe that it is an anomaly of some sort. Nevertheless, the presence of heterogeneous endpoints in some digests underscores the importance of monitoring their progress by mass spectrometry whenever possible.

Discussion

We have successfully cloned and produced the inactive zymogen of MeCPA using the baculovirus expression system, and we have shown that the zymogen can easily be activated by the endoprotease thermolysin. We have determined the pH and salt optima for MeCPA and developed an HPLC assay for specificity studies of MeCPA and BoCPA carboxypeptidases.

Based on the previously determined retention times of a large set of peptide products (VSQNX) obtained by cleavage of HIV-1 protease and accumulated data regarding how single amino acid changes alter the retention time of various peptides in a reversed-phase HPLC column [18], and assuming that Pro as well as Lys (or Arg) residues are not removed by A-type carboxypeptidases, we selected the template VSQNPKX for our study. In most cases the retention times of the substrates were substantially different than the retention time of the product VSQNPK, as exemplified in Fig. 4 for VSQNPKA. Even one amino acid change in the sequence VSQNRK resulted in the elution of the product with the injection peak hampering product detection (data not shown). The assay is similar to the one that we have used successfully to characterize the specificity of several retroviral proteases [18–21] as well as to compare the specificity of two potyvirus proteases [22]. An HPLC-based assay was also utilized to demonstrate the activity of a carboxypeptidase A from *Helicoverpa armigera* and recently an assay similar to ours was developed for the thrombin-activatable fibrinolysis inhibitor (TAFI), a B-type carboxypeptidase that plays an important role in the regulation of fibrinolysis [23]. An important advantage of our HPLC-based assay is its superior sensitivity: unlike the commonly used photometric assays [2,3,12], the HPLC-based assay was suitable to obtain kinetic parameters for various unfavorable P1′ residue-containing substrates. The data obtained for these residues are helpful in selecting P1′ amino acids that will result in homogeneous digestion products. The amino acid variations in the S1′ pocket of MeCPA as compared to BoCPA make its binding site slightly more polar and smaller than that of the latter enzyme (Fig. 2). Ser194/Asp is located at the periphery of the binding pocket. Therefore, these residues are not able to make favorable ion-pair formation with P1′ Lys or P1′ Arg residues, which may explain why peptides having such P1′ residues were resistant to the enzyme. The differently shaped, slightly smaller binding site of MeCPA may also explain the differences in the order of the best four peptides and the lower kinetic values for the substrate containing the bulky Trp residue.

There are sequence variations at residues that form the S1 specificity pocket of pancreatic carboxypeptidases and these variations have been correlated with the observed differences in specificity [2,14,24]. The B-type carboxypeptidase specificity toward positively charged residues is determined by the presence of Asp in position 253 or 255 [3]. MeCPA contains Gly and Ile at these positions, respectively, which may explain its lack of B-type specificity. Interestingly, residue 194 is Asp in MeCPA as compared to Ser in BoCPA (and human CPA1) and Thr in human CPA2. Although this variation decreases the hydrophobicity of S1, it does not allow the preferable binding of positively charged residues into this pocket. MeCPA resembles the A2 isoforms by having Gly and Ser residues at positions 253 and 254 (the same residues are also present in BoCPA). These residues provide a relatively larger binding site and facilitate the recognition of substrates with bulky aromatic side chains [25]. On the contrary, the presence of Val250 in MeCPA reduces the pocket size compared to the A2 form, and the presence of Thr at position 268 is rather a characteristic of the A1 and bovine CPA as compared to the presence of Ala in A2 [3].

An important aim of this project was to characterize MeCPA specificity and to compare it to that of BoCPA, which has been used as a reagent for the removal of polyhistidine tags from the C-termini of recombinant proteins (*e.g.*, [8,9]). Since it was assumed that MeCPA, lacking B-type activity, would prefer nonprotonated amino acid side chains in its S1 binding site, we increased the pH of the assay from the published 7.5 value to 8.2. This modification resulted in an approximately threefold increase in the efficiency with which histidine was removed from the oligopeptide substrate, while increases in the processing of other substrates were smaller or negligible (data not shown). MeCPA appears to be the more efficient catalyst overall (assuming that 100% of the enzymes in each preparation are active). The very inefficient digestion of peptides with C-terminal lysine or arginine residues, along with the complete inability of the enzyme to remove a C-terminal proline, suggests a strategy for designing C-terminal affinity tags that can be trimmed by MeCPA (or BoCPA) to produce a digestion product with a homogeneous endpoint.

Using a panel of three recombinant proteins with C-terminal His_6 tags, we were able to show that, like BoCPA, MeCPA is capable of removing the histidine residues as well as some disordered residues from the C-termini of the native protein sequences. A more thorough study with many diverse protein substrates is currently underway. However, analysis of the digestion products obtained from the three proteins tested in the current study has already provided useful information. First, as predicted, lysine residues provide an effective "roadblock" to digestion by MeCPA. Second, proline in the P1 position is a strong impediment to digestion of the P1′ residue. We were unable to confirm the "stopping power" of arginine residues in these experiments because in SycE, the only protein substrate in which the closest predicted inhibitory residue to the C-terminus is an arginine, a proline residue immediately precedes it (in the P1 position).

A question of considerable interest is whether or not the activity of MeCPA will be impeded when it encounters folded tertiary structure. If so, then this enzyme could be used not only to remove polyhistidine tags from the C-termini of proteins but also to trim some or all of the disordered residues from the C-termini of the proteins themselves. This would be of particular interest when crystallization of the target protein is the goal. We anticipate that a more comprehensive study using carefully chosen protein substrates will answer this question. In practice, however, many proteins are likely to contain one or more inhibitory residues (lysine, arginine, or proline) in the midst of a disordered C-terminus, as is true of the three proteins analyzed here. This problem might be overcome in one of two ways. First, a mixture of A and B-type carboxypeptidases should be capable of removing all disordered residues except proline (and the residue immediately after it). Second, it may be possible to relax the specificity of MeCPA by site-directed mutagenesis to accomplish the same result. Both avenues are currently being pursued.

The main advantage of the recombinant form of MeCPA described here is the presence of a C-terminal His_6 tag on the recombinant enzyme, which, in principle, should facilitate its removal from the products of a digest. However, it should be noted that as recently demonstrated for certain endoproteases [26,27], it may not be necessary to remove the enzyme from the substrate prior to setting up crystallization screens. This possibility remains to be explored for MeCPA.

Acknowledgments

This research was supported in part by the Intramural Research Program of the NIH, National Cancer Institute, Center for Cancer Research. Electrospray mass spectrometry experiments were conducted on the LC/ESMS instrument maintained by the Biophysics Resource in the Structural Biophysics Laboratory, Center for Cancer Research, National Cancer Institute at Frederick. We thank the staff of the SAIC-Frederick Protein Expression Laboratory for assistance with the construction of a recombinant baculovirus for the production of MeCPA.

References

[1] J.L. Arolas, J. Vendrell, F.X. Aviles, L.D. Fricker, Metallocarboxypeptidases: emerging drug targets in biomedicine, Curr. Pharm. Des. 13 (2007) 349–366.
[2] S.J. Gardell, C.S. Craik, E. Clauser, E.J. Goldsmith, C.B. Stewart, M. Graf, W.J. Rutter, A novel rat carboxypeptidase, CPA2: characterization, molecular cloning, and evolutionary implications on substrate specificity in the carboxypeptidase gene family, J. Biol. Chem. 263 (1988) 17828–17836.
[3] L. Joshi, R.J. St Leger, Cloning, expression, and substrate specificity of MeCPA, a zinc carboxypeptidase that is secreted into infected tissues by the fungal entomopathogen *Metarhizium anisopliae*, J. Biol. Chem. 274 (1999) 9803–9811.
[4] R.J. St Leger, M.J. Bidochka, D.W. Roberts, Characterization of a novel carboxypeptidase produced by the entomopathogenic fungus *Metarhizium anisopliae*, Arch. Biochem. Biophys. 314 (1994) 392–398.
[5] J. Arnau, C. Lauritzen, G.E. Petersen, J. Pedersen, Current strategies for the use of affinity tags and tag removal for the purification of recombinant proteins, Protein Expr. Purif. 48 (2006) 1–13.
[6] K. Terpe, Overview of tag protein fusions: from molecular and biochemical fundamentals to commercial systems, Appl. Microbiol. Biotechnol. 60 (2003) 523–533.
[7] D.S. Waugh, Making the most of affinity tags, Trends Biotechnol. 23 (2005) 316–320.
[8] A. Sali, T.L. Blundell, Comparative protein modelling by satisfaction of spatial restraints, J. Mol. Biol. 234 (1993) 779–815.
[9] D.W. Christianson, W.N. Lipscomb, X-ray crystallographic investigation of substrate binding to carboxypeptidase A at subzero temperature, Proc. Natl. Acad. Sci. USA 83 (1986) 7568–7572.
[10] J.D. Thompson, D.G. Higgins, T.J. Gibson, CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice, Nucleic Acids Res. 22 (1994) 4673–4680.
[11] R.M. Horton, H.D. Hunt, S.N. Ho, J.K. Pullen, L.R. Pease, Engineering hybrid genes without the use of restriction enzymes: gene splicing by overlap extension, Gene 77 (1989) 61–68.
[12] A. Bayes, A. Sonnenschein, X. Daura, J. Vendrell, F.X. Aviles, Procarboxypeptidase A from the insect pest *Helicoverpa armigera* and its derived enzyme. Two forms with new functional properties, Eur. J. Biochem. 270 (2003) 3026–3035.
[13] P. Boross, P. Bagossi, T.D. Copeland, S. Oroszlan, J.M. Louis, J. Tozser, Effect of substrate residues on the P2′ preference of retroviral proteinases, Eur. J. Biochem. 264 (1999) 921–929.
[14] I. Garcia-Saez, D. Reverter, J. Vendrell, F.X. Aviles, M. Coll, The three-dimensional structure of human procarboxypeptidase A2. Deciphering the basis of the inhibition, activation and intrinsic activity of the zymogen, EMBO J. 16 (1997) 6906–6913.
[15] U. Derewenda, A. Mateja, Y. Devedjiev, K.M. Routzahn, A.G. Evdokimov, Z.S. Derewenda, D.S. Waugh, The structure of *Yersinia pestis* V-antigen, an essential virulence factor and mediator of immunity against plague, Structure 12 (2004) 301–306.
[16] A.G. Evdokimov, J.E. Tropea, K.M. Routzahn, D.S. Waugh, Three-dimensional structure of the type III secretion chaperone SycE from *Yersinia pestis*, Acta Crystallogr. D Biol. Crystallogr. 58 (2002) 398–406.

[17] J. Phan, J.E. Tropea, D.S. Waugh, Structure of the *Yersinia pestis* type III secretion chaperone SycH in complex with a stable fragment of YscM2, Acta Crystallogr. D Biol. Crystallogr. 60 (2004) 1591–1599.
[18] J. Tozser, I.T. Weber, A. Gustchina, I. Blaha, T.D. Copeland, J.M. Louis, S. Oroszlan, Kinetic and modeling studies of S3–S3′ subsites of HIV proteinases, Biochemistry 31 (1992) 4793–4800.
[19] J. Kadas, I.T. Weber, P. Bagossi, G. Miklossy, P. Boross, S. Oroszlan, J. Tozser, Narrow substrate specificity and sensitivity toward ligand-binding site mutations of human T-cell Leukemia virus type 1 protease, J. Biol. Chem. 279 (2004) 27148–27157.
[20] J. Tozser, I. Blaha, T.D. Copeland, E.M. Wondrak, S. Oroszlan, Comparison of the HIV-1 and HIV-2 proteinases using oligopeptide substrates representing cleavage sites in Gag and Gag-Pol polyproteins, FEBS Lett. 281 (1991) 77–80.
[21] J. Tozser, D. Friedman, I.T. Weber, I. Blaha, S. Oroszlan, Studies on the substrate specificity of the proteinase of equine infectious anemia virus using oligopeptide substrates, Biochemistry 32 (1993) 3347–3353.
[22] J. Tozser, J.E. Tropea, S. Cherry, P. Bagossi, T.D. Copeland, A. Wlodawer, D.S. Waugh, Comparison of the substrate specificity of two potyvirus proteases, FEBS J. 272 (2005) 514–523.
[23] Z. Valnickova, I.B. Thogersen, J. Potempa, J.J. Enghild, Thrombin-activable fibrinolysis inhibitor (TAFI) zymogen is an active carboxypeptidase, J. Biol. Chem. 282 (2007) 3066–3076.
[24] A.L. Osterman, N.V. Grishin, S.V. Smulevitch, M.V. Matz, O.P. Zagnitko, L.P. Revina, V.M. Stepanov, Primary structure of carboxypeptidase T: delineation of functionally relevant features in Zn-carboxypeptidase family, J. Protein Chem. 11 (1992) 561–570.
[25] L. Catasus, J. Vendrell, F.X. Aviles, S. Carreira, A. Puigserver, M. Billeter, The sequence and conformation of human pancreatic procarboxypeptidase A2. cDNA cloning, sequence analysis, and three-dimensional model, J. Biol. Chem. 270 (1995) 6651–6657.
[26] A. Dong, X. Xu, A.M. Edwards, C. Chang, M. Chruszcz, M. Cuff, M. Cymborowski, R. Di Leo, O. Egorova, E. Evdokimova, E. Filippova, J. Gu, J. Guthrie, A. Ignatchenko, A. Joachimiak, N. Klostermann, Y. Kim, Y. Korniyenko, W. Minor, Q. Que, A. Savchenko, T. Skarina, K. Tan, A. Yakunin, A. Yee, V. Yim, R. Zhang, H. Zheng, M. Akutsu, C. Arrowsmith, G.V. Avvakumov, A. Bochkarev, L.-G. Dahlgren, S. Dhe-Paganon, S. Dimov, L. Dombrovski, P. Finerty, S. Flodin, A. Flores, S. Graslund, M. Hammerstrom, M.D. Herman, B.-S. Hong, R. Hui, I. Johansson, Y. Liu, M. Nilsson, L. Nedyalkova, P. Nordlund, T. Nyman, J. Min, H. Ouyang, H.-W. Park, C. Qi, W. Rabeh, L. Shen, Y. Shen, D. Sukumard, W. Tempel, Y. Tong, L. Tresagues, M. Vedadi, J.R. Walker, J. Weigelt, M. Welin, H. Wu, T. Xiao, H. Zeng, H. Zhu, *In situ* proteolysis for protein crystallization and structure determination, Nat. Methods 4 (2007) 1019–1021.
[27] A. Wernimont, A. Edwards, In situ proteolysis to generate crystals for structure determination: an update, PLoS ONE 4 (2009).