Characterization of a multimeric, eukaryotic prolyl aminopeptidase: an inducible and highly specific intracellular peptidase from the non-pathogenic fungus *Talaromyces emersonii*

Cathal S. Mahon,^{1,2†} Anthony J. O'Donoghue,^{1,2†} David H. Goetz,² Patrick G. Murray,¹ Charles S. Craik² and Maria G. Tuohy¹

Correspondence

Maria G. Tuohy
maria.tuohy@nuigalway.ie

¹Department of Biochemistry, National University of Ireland, University Road, Galway, Ireland

²Department of Pharmaceutical Chemistry, University of California, San Francisco, CA 94158, USA

Fungi are capable of degrading proteins in their environment by secreting peptidases. However, the link between extracellular digestion and intracellular proteolysis has scarcely been investigated. Mycelial lysates of the filamentous fungus *Talaromyces emersonii* were screened for intracellular peptidase production. Five distinct proteolytic activities with specificity for the *p*-nitroanilide (pNA) peptides Suc-AAPF-pNA, Suc-AAA-pNA, K-pNA, F-pNA and P-pNA were identified. The native enzyme responsible for the removal of N-terminal proline residues was purified to homogeneity by ammonium sulfate fractionation followed by five successive chromatographic steps. The enzyme, termed *Talaromyces emersonii* prolyl aminopeptidase (TePAP), displayed a 50-fold specificity for cleaving N-terminal Pro-X ($k_{\text{cat}}/K_m = 2.1 \times 10^6 \text{ M}^{-1} \text{ s}^{-1}$) compared with Ala-X or Val-X bonds. This intracellular aminopeptidase was optimally active at pH 7.4 and 50 °C. Peptide sequencing facilitated the design of degenerate oligonucleotides from homologous sequences encoding putative fungal proline aminopeptidases, enabling subsequent cloning of the gene. TePAP was shown to be relatively uninhibited by classical serine peptidase inhibitors and to be sensitive to selected cysteine- and histidine-modifying reagents, yet gene sequence analysis identified the protein as a serine peptidase with an α/β hydrolase fold. Northern analysis indicated that *Tepap* mRNA levels were regulated by the composition of the growth medium. Highest *Tepap* transcript levels were observed when the fungus was grown in medium containing glucose and the protein hydrolysate casitone. Interestingly, both the induction profile and substrate preference of this enzyme suggest potential co-operativity between extracellular and intracellular proteolysis in this organism. Gel filtration chromatography suggested that the enzyme exists as a 270 kDa homo-hexamer, whereas most bacterial prolyl aminopeptidases (PAPs) are monomers. Phylogenetic analysis of known PAPs revealed two diverse subfamilies that are distinguishable on the basis of primary and secondary structure and appear to correlate with the subunit composition of the native enzymes. Sequence comparisons revealed that PAPs with key conserved topological features are widespread in bacterial and fungal kingdoms, and this study identified many putative PAP candidates within sequenced genomes. This work represents, to our knowledge, the first detailed biochemical and molecular analysis of an inducible PAP from a eukaryote and the first intracellular peptidase isolated from the thermophilic fungus *T. emersonii*.

Received 15 May 2009
Revised 14 June 2009
Accepted 17 June 2009

†These authors contributed equally to this work.

Abbreviations: AEBSF, 4-(2-aminoethyl) benzenesulfonyl fluoride hydrochloride; DEPC, diethylpyrocarbonate; E-64, *trans*-epoxy-succinyl-L-leucylamido-(4-guanidino)-butane; NEM, *N*-ethylmaleimide; PAP, prolyl aminopeptidase; PCMB, *p*-chloromercuribenzoate; pNA, *p*-nitroanilide; Pro-AMC, L-proline-7-amino-4-methylcoumarin; TLCK, *N*-tosyl-L-lysine chloromethyl ketone; TPCK, *N*-tosyl-L-phenylalanine chloromethyl ketone.

The GenBank/EMBL/DDBJ accession number for the prolyl aminopeptidase sequence of *Talaromyces emersonii* is AF439997.

Four supplementary tables, presenting the effects of culture medium composition on intracellular peptidase activity in *T. emersonii*, the purification of PAP from *T. emersonii*, biochemically characterized PAP enzymes, and the identification of PAPs in fungi, with supplementary references, are available with the online version of this paper.

INTRODUCTION

As chemoheterotrophs, fungi acquire all their nutrients from the environment and as a result have evolved a battery of hydrolases capable of degrading most biopolymers that they encounter in nature. The genes encoding these enzymes are regulated by the presence or absence of specific inducers, with examples including cellulase induction by cellulose (Murray *et al.*, 2003), amylase induction by (iso)maltose (Kato *et al.*, 2002), and peptidase induction by peptides and proteins (Polaina & MacCabe, 2007). The secreted xylanolytic (Polizeli *et al.*, 2005), cellulolytic (Stricker *et al.*, 2008) and proteolytic systems (Kredics *et al.*, 2005) in fungi have been well characterized due to their relative ease of purification and high production yields. From studies to date it is clear that in addition to endo-acting secreted peptidases, fungi possess numerous intracellular peptidases with a range of substrate specificities that complete protein degradation.

Talaromyces emersonii is a thermophilic filamentous fungus that inhabits soil, decaying masses of plant material, and other accumulations of organic matter wherein the warm, humid and aerobic environment provides the basic conditions for its development (Allen & Emerson, 1949). It has generally recognized as safe (GRAS) status, making it a biotechnologically relevant organism to rival *Aspergillus* and *Trichoderma* species for the production of important industrial enzymes. *T. emersonii* has been shown to secrete a wide array of hydrolytic enzymes under defined growth conditions, and possesses complete xylan- and cellulase-degrading systems that consist of diversified carbohydrases (Heneghan *et al.*, 2007; Murray *et al.*, 2001, 2003; Reen *et al.*, 2003; Tuohy *et al.*, 1993, 1994, 2002). Previously, we have shown that the presence of peptides in growth media induces expression of an extracellular peptidase with broad substrate specificity in *T. emersonii* (O'Donoghue *et al.*, 2008). In this study we examine intracellular proteolytic activity in *T. emersonii* under similar growth conditions. In order to better understand the role of intracellular peptidases we describe the purification and classification of the enzyme with the highest detectable activity, a prolyl aminopeptidase (PAP).

PAPs are widespread in nature but little is known about their biological function. The unique structure of proline prevents its cleavage by many broad-specificity aminopeptidases (Cunningham & O'Connor, 1997). Accordingly, the need to hydrolyse such bonds has led to the evolution of a specific subset of enzymes (PAPs) capable of cleaving N-terminal proline. Many bacteria and fungi that produce PAPs are pathogenic. A role for PAP in infectivity has been reported (Zhang *et al.*, 2007), and it has also been proposed as a viable drug target (Felipe *et al.*, 2005). Proline, the product of PAP, has multiple physiological functions in micro-organisms, plants and animals, and its accumulation has been implicated in stress responses in yeast and in the suppression of apoptosis in fungi (Takagi, 2008). PAP is an important enzyme industrially, as fungi that possess

intracellular proline aminopeptidase-type activity have been extensively used in flavour development of food products (Bolumar *et al.*, 2003; Fuke & Matsuoka, 1993).

To date, characterization of PAPs has mainly been carried out with enzymes from bacterial sources (Supplementary Table S3). A number of PAP-type activities in fungi have been reported (Bolumar *et al.*, 2003; Fuke & Matsuoka, 1993; Hiwatashi *et al.*, 2004), but no corresponding protein sequences have been assigned for these enzymes. However, more recently, one study by Basten *et al.* (2005) identified a gene encoding PapA from *Aspergillus niger* and characterized the recombinant protein.

This work describes the purification and characterization of a native fungal PAP. Cloning and sequence analysis of its gene and interrogation of genome databases allowed the identification of many additional homologous members of this unusual family of peptidases. Furthermore, we show that mRNA levels are regulated by the composition of the growth medium and propose a direct link between extracellular peptidases that yield proline-capped peptides and the intracellular uncapping activity of TePAP.

METHODS

All chemicals and reagents were obtained from Sigma–Aldrich unless otherwise stated.

Fungal growth. *T. emersonii* was obtained from laboratory stocks and routinely subcultured on Sabouraud dextrose agar (SDA) plates at 45 °C. Mycelial plugs from SDA plates were used to inoculate liquid medium for peptidase production that contained 0.5% (w/v) yeast nitrogen base without amino acids and ammonium sulfate (YNB), supplemented with 2% (w/v) glucose or glycerol and 0.5% (w/v) casein, casitone (BD Biosciences), Casamino acids (BD Biosciences) or ammonium sulfate, as indicated. All liquid cultures were grown at 45 °C on an orbital shaker at 220 r.p.m.

Preparation of fungal extract and peptidase screen. Mycelia were harvested, frozen in liquid N₂ and disrupted by grinding with a mortar and pestle in the presence of glass beads. The resultant biomass was resuspended in buffer A (0.05 M Tris/HCl, pH 7.4), containing 200 µM DTT, 0.001% (v/v) Tween 80 and 5 U DNase (Roche Applied Science), and incubated overnight at 4 °C with gentle mixing. Glass beads and cellular debris were removed by centrifugation at 15 000 g for 30 min and the supernatant was filtered through a 0.22 µm pore-size filter (Millipore). Peptidase activity was assessed using 1 mM *p*-nitroanilide (pNA) peptides listed in Supplementary Table S1. Assays were performed in 0.3 M Na₂HPO₄/NaH₂PO₄, pH 7.4, for 30 min at 50 °C. The reaction was terminated by the addition of an equal volume of 30% (v/v) acetic acid and absorbance was measured using a Wallac Victor multilabel counter at 405 nm. Three replicates of each test and control were assayed and appropriate enzyme, substrate and reagent controls were included.

Enzyme purification. For protein purification, *T. emersonii* was cultured for 72 h in medium containing glucose and casitone. Mycelia were harvested as outlined above, and the intracellular extract was concentrated and dialysed against buffer A in a stirred ultrafiltration cell fitted with a 10 kDa membrane (Millipore). Protein present in the crude extract was recovered by centrifugation at 15 000 g for 30 min following sequential precipitation with 45 and

85 % (w/v) ammonium sulfate. The final precipitate was resuspended in buffer A and desalted on a Sephadex G25 column. Proteolytically active fractions were applied to a Phenyl Sepharose CL-4B column and protein was eluted using a linear gradient of 0–5 % ethylene glycol in buffer A. Active fractions were pooled, adjusted to 0.2 M NaCl, reapplied to the column and eluted in a decreasing salt gradient (0.2–0 M NaCl). Enzyme-active fractions were pooled and dialysed against buffer A and applied to a DEAE-Sepharose CL-6B column. Adsorbed protein was eluted in a gradient of 0–0.35 M NaCl, concentrated in a Centricon 10 (Millipore) and applied to a Sephacryl S-300 HR column. Protein was eluted in buffer A containing 0.1 M NaCl. The protein concentration in chromatography fractions was measured routinely by determining A_{280} , while pooled fractions were quantified using a modified Lowry method (Bensadoun & Weinstein, 1976) standardized against BSA fraction V. A molar absorption coefficient of $8800 \text{ M}^{-1} \text{ cm}^{-1}$ was used to calculate activity against P-pNA (Varmanen *et al.*, 1996).

Enzyme characterization. *TePAP* specificity was analysed using peptidyl-pNA substrates as outlined above. All assays were performed in triplicate and activity was calculated relative to P-pNA. Optimal pH for *TePAP* activity was determined using the following buffer solutions: McIlvaine buffer (pH 3.8–6.4), sodium phosphate buffer (pH 6.6–7.8), Tris/HCl buffer (pH 8–9.2) and sodium carbonate/bicarbonate buffer (pH 9.2–10.4). The optimum temperature for activity was determined over a range from 20 to 90 °C. Temperature stability was determined by pre-incubating purified *TePAP* for 30 min at varying temperatures. All temperature assays were performed at pH 7.4 and enzyme stored at 4 °C served as a stability control. The effects of peptidase inhibitors, amino acid-modifying reagents, metal ions and chelators on purified *TePAP* were analysed

using P-pNA, following pre-incubation for 10 min with the target reagent at room temperature (Table 1b).

Determination of kinetic parameters. Activity of *TePAP* (1 nM) was monitored using L-proline-7-amino-4-methylcoumarin (Pro-AMC) as substrate in a Wallac Victor multilabel counter (Perkin Elmer) set at excitation (λ_{ex}) 360 nm and emission (λ_{em}) 440 nm. Kinetic constants, K_m and V_{max} were calculated by nonlinear regression using KaleidaGraph (Synergy Software), following continuous assay at pH 7.4 and 45 °C with 1–150 μM Pro-AMC.

Protein characterization. Protein purity and molecular mass were evaluated by reducing SDS-PAGE in a 10 % Tris-glycine gel stained with 0.02 % (w/v) Coomassie brilliant blue G-250. The protein band was excised from the gel and trypsin-digested, and internal peptide sequences were determined by liquid chromatography (LC)-MS/MS (Jiménez *et al.*, 2001) using a Micromass QTOF2 mass spectrometer (at the University of Nottingham, UK). Zymography was performed by native PAGE in an 8 % Tris-glycine gel and incubated in 100 μM Pro-AMC, 0.3 M Tris/HCl, pH 7.2, at 50 °C for 5 min (Yasothornsrikul & Hook, 2000). Active enzyme was visualized using a Fluor-S Multimager (Bio-Rad).

DNA extraction and cloning of *TePAP*. DNA was isolated (Raeder & Broda, 1985) and RNase-treated. A partial gene sequence encoding *TePAP* was PCR-amplified from genomic DNA using degenerate primers (5'-TTRCARGGNGGCCNGG-3' and 5'-CCNCCRAARC-TYTGNC-3', where N=A, C, T or G; Y=C or T; and R=A or G) designed from highly conserved regions of fungal PAPs (GenBank accession numbers CAC40647 and CAC40648). The PCR product was subcloned into the pGEM-T Easy vector (Promega), propagated in *E. coli* JM109 cells and sequenced. A DIG-labelled DNA probe (Roche Applied Science) was generated using gene-specific oligonucleotides (5'-ATCCTCTCTCTCGACCAACGCGG-3' and 5'-AATCCGCCGA-AACTCTGCCCCA-3'). This probe was used to screen a *T. emersonii* genomic library (Grassick *et al.*, 2004) for bacteriophage containing the full-length *Tepap* gene. Positive clones were sequenced using 5'-CCGCGTTGGTTCGAGGAAGAGGAT-3' and 5'-TCGGGCAGAG-TTTCGGCGGATT-3' oligonucleotides.

RNA extraction and Northern blot analysis. Total RNA was isolated from *T. emersonii* with TRI-Reagent (Molecular Research Centre). RNA was treated with DNase to remove residual genomic DNA and converted to cDNA using M-MLV reverse transcriptase (Applied Biosystems). The complete coding sequence of *Tepap* was confirmed by PCR amplification and sequencing of cDNA using gene-specific oligonucleotides (5'-TCCTAGTTCTGCGGGTCAGAC-3' and 5'-TGTCTGGACTGGAGACTATG-3'). Northern blot analyses were performed on total RNA as outlined previously (Murray *et al.*, 2003), using the DIG-labelled *Tepap* probe and detection with anti-DIG alkaline phosphatase antibody (Roche Molecular Biochemicals). Membranes were stripped in 0.2 M Tris/HCl, pH 7.5, 50 % (v/v) formamide at 80 °C and hybridized with DIG-labelled 18S rRNA probe (GenBank accession no. D88321.2). Images were visualized using a Fluor-S MultiImager (Bio-Rad).

Sequence and phylogenetic analyses. Sequences of known, biochemically characterized PAP enzymes from the BRENDA database (www.brenda-enzymes.org) were retrieved using the Universal Protein Resource (UniProt, 2007). CLUSTAL_X was used to align protein sequences (www.clustal.org), and sequences were manually modified where required. CLUSTAL_X was also used to construct an unrooted phylogenetic tree by the neighbour-joining method, and 1000 bootstrap trials were used (Larkin *et al.*, 2007). The tree was displayed with TreeView. To identify putative fungal PAPs, *TePAP* was used to query completed fungal genomes available at www.broad.mit.edu, www.jgi.doe.gov and www.yeastgenome.org.

Table 1. (a) Substrate specificity and (b) inhibition profile of purified *TePAP*

(a) Relative activity was expressed as a percentage of P-pNA activity. No activity was detected against K-pNA, E-pNA, M-pNA, AAP-pNA, Suc-AAA-pNA and Suc-AAPF-pNA. (b) Greater than 90 % activity was observed in the presence of MgSO_4 , MnSO_4 , MnCl_2 , LiSO_4 , CdCl_2 , KI, ZnSO_4 , CaCl_2 , 1,10-phenanthroline, pepstatin, bestatin, leupeptin, EDTA, AEBSF and E-64. All assays were carried out in triplicate and are expressed as mean $\pm$ SEM.

(a) Substrate	Relative activity (%)
P- <i>p</i> NA	100 ± 0.8
A- <i>p</i> NA	2 ± 0.5
V- <i>p</i> NA	2 ± 0.4
L- <i>p</i> NA	≤ 1 ± 0.2
F- <i>p</i> NA	≤ 1 ± 0.2
G- <i>p</i> NA	≤ 1 ± 0.1
GP- <i>p</i> NA	0

(b) Inhibitor	Concentration (μM)	Relative activity (%)
DEPC	1000	3 ± 0.4
PCMB	100	6 ± 2.8
HgCl ₂	500	31 ± 2.6
Iodoacetic acid	200	52 ± 1.3
TLCK	1000	62 ± 2.4
NEM	100	71 ± 1.6
TPCK	1000	89 ± 0.6

BLASTP analysis was used to identify other protein sequences with significant homology, while secondary structure was predicted by PSIPRED (McGuffin *et al.*, 2000).

RESULTS

Intracellular peptidase screen

The effects of medium composition on intracellular peptidase activities in *T. emersonii* mycelial lysates was assessed following 72 h growth and results are summarized in Supplementary Table S1. Hydrolysis of Suc-Ala-Ala-Pro-Phe-*p*NA (Suc-AAPF-*p*NA), Suc-Ala-Ala-Ala-*p*NA (Suc-AAA-*p*NA), P-*p*NA, Lys-*p*NA (K-*p*NA) and Phe-*p*NA (F-*p*NA) was observed under all culture conditions tested; activities against all other substrates were ≤ 1 IU mg^{-1} (Supplementary Table S1). Activity against P-*p*NA was the predominant activity in mycelia and was maximal in cells grown in a combination of glucose and casitone. Levels of the other two mono-aminopeptidase-type activities and the chymotrypsin-type activity (Suc-AAPF-*p*NA) were lower under all conditions tested. Moderate levels of the keratinase-type activity (Suc-AAA-*p*NA) were detected.

TePAP purification

P-*p*NA activity in mycelia increased with time over a 60 h growth period after inoculation, and maximum activity was observed between 60 and 84 h of growth (data not shown). Extracellular P-*p*NA activity was not detected under these conditions (data not shown). Purification of a PAP was therefore carried out from a crude extract prepared from *T. emersonii* mycelia harvested after 72 h growth in medium containing glucose and casitone. A homogeneous enzyme preparation (TePAP) was purified 78-fold by ammonium sulfate fractionation, followed by five successive chromatographic steps to a final yield of

0.6%, as outlined in Supplementary Table S2. The estimated M_r of 270 000 for the native enzyme was determined by both gel filtration chromatography and zymography (Fig. 1). The observation of a single protein band with an M_r of 46 000 following SDS-PAGE suggests that the enzyme exists in its native state as a homo-hexamer.

Enzymic characterization of TePAP

TePAP displayed 50-fold greater specificity towards P-*p*NA compared with Ala-*p*NA (A-*p*NA) and Val-*p*NA (V-*p*NA), and ≥ 100 fold greater specificity compared with all other aminopeptidase substrates tested (Table 1). No activity was detected against di- and tri-peptidyl *p*NA substrates. In addition, purified TePAP was shown to be incapable of cleaving K-*p*NA, F-*p*NA, Suc-AAPF-*p*NA or Suc-AAA-*p*NA, indicating that it was not responsible for the other peptidase activities detected in *T. emersonii* mycelial lysates. TePAP was most active against P-*p*NA between pH 7 and pH 8, with the highest activity observed at pH 7.4 (Fig. 2). At pH 7.4, the optimal temperature for activity was 50 °C, but activity dropped rapidly above this temperature with only 50% remaining at 60 °C. Under optimal assay conditions (pH 7.4, 50 °C), TePAP retained ~40% activity after 1 h and was completely inactivated after 2 h. Under optimal pH conditions, K_m , k_{cat} and k_{cat}/K_m values of 101 μM , 213.7 s^{-1} and $2.1 \times 10^6 \text{ M}^{-1} \text{ s}^{-1}$ respectively, were obtained using a fluorescent Pro-AMC substrate.

Inhibition of TePAP activity

The effects of amino-acid-modifying reagents and class-specific peptidase inhibitors on TePAP activity are shown in Table 1. Bestatin, a specific inhibitor of aminopeptidases, had no effect on activity, nor did the serine and aspartic peptidase inhibitors 4-(2-aminoethyl) benzenesulfonyl fluoride hydrochloride (AEBSF) and pepstatin,

Fig. 1. Purification of TePAP by size-exclusion chromatography using Sephacryl S-300 HR. Activity (black line) and protein abundance (grey line) are indicated. Arrows correspond to the peak fractions for gel filtration standards: ferritin, 440 000 Da; catalase, 240 000 Da; amylase, 200 000 Da. Fractions 68–78 were pooled and analysed by reducing SDS-PAGE followed by Coomassie staining (lane 2), and by native PAGE followed by zymography using Pro-AMC as substrate (lane 4). Denatured and native molecular mass markers are shown in lanes 1 and 3, respectively.

Fig. 2. Effect of temperature and pH on the activity and stability of purified *TePAP*. (a) Effect of pH on *TePAP* activity. (b) Effect of temperature on *TePAP* activity. (c) Effect of temperature on *TePAP* stability at 4 °C (broken line), 50 °C (solid line) and 60 °C (dotted line). *P*-pNA was used for all assays. For assays see Methods.

respectively. The histidine-modifying reagent diethylpyr-carbonate (DEPC) caused a 97% reduction in enzyme activity. Cysteine peptidase inhibitors such as *p*-chloro-

mercuribenzoate (PCMB), *N*-ethylmaleimide (NEM), iodoacetic acid and HgCl_2 all inhibited *P*-pNA hydrolysis to varying extents, although E-64 [*trans*-epoxy-succinyl-L-leucylamido-(4-guanidino)-butane] did not. The serine/cysteine inhibitors *N*-tosyl-L-lysine chloromethyl ketone (TLCK) and *N*-tosyl-L-phenylalanine chloromethyl ketone (TPCK) inhibited *TePAP* moderately. No effect on activity was detected in the presence of a range of metal salts or the chelators EDTA and 1,10-phenanthroline, suggesting that *TePAP* activity is not metal-dependent.

Cloning and sequence analysis of *Tepap*

Peptide sequencing of *TePAP* revealed that the enzyme is related to *Aspergillus niger* PapA and displays homology to a number of putative and unassigned fungal sequences (Fig. 3a). Degenerate oligonucleotides were designed from regions of homology that flank the *TePAP* peptides in other PAPs. These oligonucleotides facilitated PCR amplification and cloning of a 407 bp fragment from *T. emersonii* genomic DNA encoding *Tepap*. This sequence was used to design a specific DNA probe to screen a *T. emersonii* genomic library. The *Tepap* gene and protein-coding region were obtained by sequencing of a genomic clone and cDNA, respectively (GenBank accession no. AF439997). The cDNA sequence of *Tepap* is composed of nine exons, which encode a polypeptide of 448 amino acids with a predicted molecular mass of 50.6 kDa. Signal P (Bendtsen *et al.*, 2004) analysis determined that *TePAP* does not possess a secretory signal peptide. Conserved domain analysis of the sequence revealed it to be an α/β hydrolase fold serine peptidase containing a Ser-Asp-His active site triad. The nucleophilic serine is in the Gly-X-Ser-B-Z motif common to serine peptidases, where X represents any amino acid, and B and Z represent amino acids with a bulky R group and a small R group, respectively (Fig. 3b).

Expression profile of *Tepap*

Total RNA was extracted from *T. emersonii* after 24 and 48 h growth in a variety of media, as outlined in Fig. 4. Transcript levels of *Tepap* were determined by hybridization with a DNA probe specific to exon 3. These studies showed that low levels of mRNA were detectable after 24 h growth under all conditions examined. However, *Tepap* mRNA accumulated after a further 24 h growth in medium containing both glucose and casitone, but not when either compound was present alone. Substitution of glycerol for glucose failed to reproduce the expression levels seen in glucose + casitone medium. Addition of an easily metabolized nitrogen source such as ammonium sulfate had no effect on *Tepap* expression levels.

Sequence analysis and secondary structure prediction

Analysis of enzymes in the BRENDA database yielded 30 PAPs that have been biochemically characterized. As of

Fig. 3. Sequence analysis of *TePAP*. (a) Protein alignment of *TePAP* with known bacterial and fungal PAPs from *Aspergillus niger*, *PapA*; *Aeromonas sobria*, *AsPAP*; and *X. campestris*, *XcPIP*. Amino acids conserved in *TePAP* are highlighted in grey. Predicted helices (cylinders) and strands (arrows) in multimeric PAPs corresponding to confirmed elements in *XcPIP* are shown in dark grey. Predicted helices unique to all multimeric PAPs are shown in light grey. Peptide sequences determined by liquid chromatography-MS/MS are underlined. Regions used to design degenerate oligonucleotides are indicated with dashed arrows. (b) Catalytically and functionally important residues in PAPs. Residues with important functions have been deciphered from structural studies with *XcPIP*. These residues are listed in (b) and are indicated by ▲ and * in (a). The (semi)-conservation of these residues in the four aligned PAPs from (a) is indicated.

May 2009, 12 of these enzymes had been assigned protein sequences while we retrieved a further seven sequences by homology from available genomic sequences (see Supplementary Table S3). As well as being capable of cleaving Pro bonds, all PAPs were described as being

sensitive to one or more thiol-modifying reagent(s). The subunit number of the native form of each enzyme was also recorded. Aligning *TePAP* with a representative monomeric bacterial (*Xanthomonas campestris*, *XcPIP*), a multimeric bacterial (*Aeromonas sobria*, *AsPAP*) and a

Fig. 4. Northern analysis of total RNA for *Tepap* extracted from *T. emersonii*.

multimeric fungal (*Aspergillus niger*, PapA) enzyme showed specific regions of high homology, but overall sequence identities of the full-length proteins were 13, 40 and 72 % respectively (Fig. 3a). A secondary structure prediction of each protein identified 12 α -helices and eight β -strands that correlate with the known structure of XcPIP (Medrano *et al.*, 1998). An additional four helices (M α 1–4) are present in *TePAP*, PapA and AsPAP, and are predicted to be unique to the multimeric PAPs.

A phylogenetic tree of all 19 available PAP sequences from biochemically characterized enzymes was constructed. The tree revealed the presence of two major subfamilies that are distinguishable by the presence or absence of M α 1–4 (Fig. 5). Interestingly, the MEROPS database has grouped PAPs into two subfamilies on the basis of similarity to *Neisseria gonorrhoeae* PAP (S33.001) or AsPAP (S33.008). We define here that this classification correlates exactly with the presence of four additional helices (M α 1–4) in all

S33.008 subfamily proteins and their absence in S33.001 enzymes. S33.008 proteins are found in eukaryotes (fungi and plants) and some bacteria, while S33.001 enzymes are only found in bacteria.

BLASTP searches performed with *TePAP* revealed hypothetical proteins in the genomes of many fungi. The sequences closest to *TePAP* were identified in *Aspergillus fumigatus* (71 % identity), *Neosartorya fischeri* (71 %) and *Aspergillus clavatus* (70 %). Matching sequences were also present in model fungal species (e.g. *Neurospora crassa*, *Aspergillus nidulans*), human pathogens (e.g. *Candida albicans*, *Paracoccoides brasiliensis*) and plant pathogens (e.g. *Fusarium graminearum*, *Magnaporthe grisea*). These proteins were aligned based on the presence of the putative catalytic serine, aspartic and histidine residues identified by homology (Supplementary Table S4). Notable fungal genomes lacking PAP homologues are listed in the legend of Supplementary Table S4.

Fig. 5. Phylogenetic tree of characterized PAPs from all known organisms. The tree was constructed using a CLUSTAL_X alignment of entire PAP protein sequences. Asterisks indicate S33.001 proteins that are multimeric. The branch lengths are proportional to sequence divergence and the scale of 0.1 amino acid changes per site is indicated in the bottom left. The numbers displayed are bootstrap values (1000 replicates). See Supplementary Table S3 for the full unabbreviated names.

DISCUSSION

Fungi have the ability to digest polymeric compounds encountered in their environment into smaller constituents that are readily taken up by mycelia as carbon and nitrogen nutrient sources. This characteristic is reflected in the high levels of hydrolytic enzyme activities that have been found in both the mycelial lysates and culture filtrates of many fungi. Due to their opportunistic ecological niche, saprophytes such as *T. emersonii* have the ability to utilize a variety of biopolymers and therefore display a wide range of enzymic activities. Proteins and peptides are the most abundant naturally occurring nitrogen sources and due to protein sequence heterogeneity, a range of peptidases with varying specificities are required to carry out the degradation of these proteins and peptides into smaller more readily assimilated peptides.

Previous work has shown that peptide hydrolysates such as caseitone (an enzymic hydrolysate of casein) are excellent inducers of a major extracellular peptidase in *T. emersonii* (O'Donoghue *et al.*, 2008). To expand our knowledge of the spectrum of peptidases produced by this fungus, we screened cell lysates generated under similar growth conditions for the presence of peptidases. This screen identified three mono aminopeptidase-type activities (active against K-pNA, F-pNA and P-pNA), as well as chymotrypsin-type (Suc-AAPF-pNA) and keratinase-type (Suc-AAA-pNA) activities. Purification and characterization of the predominant intracellular activity produced by *T. emersonii*, which was determined to be a proline aminopeptidase (*TePAP*), was successfully achieved.

The inhibition profile of purified *TePAP* indicated that histidine-alkylating and thiol-modifying reagents inhibit P-pNA hydrolysis, suggesting the involvement of essential, though not necessarily catalytic, Cys and His residues. However, sequence analysis identified the enzyme as a serine peptidase, complete with a conserved catalytic triad (Fig. 3). The unusual inhibitor profile of all PAPs regarding their sensitivity to thiol-modifying reagents (Supplementary Table S3) and resistance to serine peptidase inhibitors has complicated their classification. Earlier studies indicated that inhibition of *Serratia marcescens* PAP (*SmPAP*) by PCMB was the result of modification of cysteine residues located in or around the P₁ pocket (Ito *et al.*, 2000). It should be noted, however, that no cysteine was found to be conserved in all PAPs, therefore further work is required to elucidate the cysteine peptidase-like inhibition patterns of these enzymes. The discrepancy observed between the near total inhibition of *TePAP* by PCMB but only partial inhibition by NEM has also been observed with other bacterial PAPs (Kitazono *et al.*, 1996). This is consistent with PCMB being more reactive than NEM while also being capable of alkylating amines (Morty *et al.*, 2005).

PAPs are not obligate 'proline aminopeptidases', as many members identified to date are capable of cleaving residues that consist of structural elements of the pyrrolidine ring of

proline, such as alanine and sarcosine. In fact, some PAPs are capable of cleaving such substrates at higher rates relative to Pro-X bonds (Inoue *et al.*, 2003). *TePAP*, however shows 50-fold lower activity against A-pNA relative to P-pNA (Table 1). The residues responsible for catalysis and hydrophobic pocket formation are indicated in Fig. 3(b). Some conservation is observed across all kingdoms, and it is possible that subtle differences in the micro-environment of this pocket account for the varying secondary cleavage preference noted for individual PAPs.

It is well established that the presence of extracellular inducers can affect expression levels of fungal hydrolases via specific pathways, such as cellulase induction by cellulose (Doi, 2008). The production and regulation of secreted peptidases in fungi are strongly influenced by culture conditions, although conditions for induction and/or repression appear to be enzyme- and species-specific (Polaina & MacCabe, 2007). Much less information exists regarding the regulation of intracellular fungal peptidases. The results presented here establish that *TePAP* is regulated at the level of transcription by environmental signals. In contrast, earlier studies of intracellular aminopeptidases from *Aspergillus* have indicated that expression is constitutive (Basten *et al.*, 2001, 2003, 2005). This study definitively shows induction of *TePAP* when the peptide hydrolysate from casein (a protein rich in proline) is present in growth medium supplemented with a favoured carbon source (glucose). Peptides from a similar casein hydrolysate (peptone) have been shown to induce lipase production in the yeast *Yarrowia lipolytica*, and this induction has been attributed to the presence of specific peptides (Fickers *et al.*, 2004). It remains possible that a similar induction pathway and therefore peptidase inducer exists for fungal peptidases like *TePAP*. It is interesting to note that peptides larger than six amino acids cannot be transported across the membrane of yeast (Milewski *et al.*, 2001), implying that extensive 'pre'-hydrolysis of proteins/peptides present in the microbial environment is required for efficient uptake. This concept supports our previous studies of secreted pepsin- and glutamic-type peptidases from *T. emersonii*. TGP1, a glutamic peptidase, readily cleaves Phe-Pro bonds as well as a broad range of non-proline-containing peptide bonds (O'Donoghue *et al.*, 2008), while PEP1 (pepsin-type enzyme) cleaves Phe-Pro, Lys-Pro and Arg-Pro bonds (results not shown). The combined action of the two enzymes would generate multiple short peptides containing N-terminal proline residues for uptake. Further evidence to support this concerted digestion was derived by comparison of the expression patterns of TGP1 and *TePAP*. The highest expression level of *tgp1* occurs after 24 h growth on glucose/caseitone medium and is largely downregulated to basal levels by 48 h. However, under identical conditions, *TePAP* is only induced after 48 h growth, which correlates with the likely timeframe required for TGP1 to reduce extracellular protein to peptides of a suitable size for intracellular uptake. Furthermore, there is the possibility that other as yet undiscovered peptidase(s) in addition to

TGP1 also provide suitable peptides for subsequent uptake and hydrolysis by PAP.

Previously, PAPs have been broadly classified into two groups on the basis of their existence as monomeric or multimeric proteins (Polaina & MacCabe, 2007) and also tentatively by substrate specificity (Yoshimoto *et al.*, 1999). This study definitively shows that PAPs occupy at least two subfamilies that correlate well with multimerization. The 3D structures of only two PAPs have been solved (both of which are functional as monomers), from the Gram-negative bacteria *Xanthomonas campestris* (XcPIP; Medrano *et al.*, 1998) and *Serratia marcescens* (Yoshimoto *et al.*, 1999). These proteins are folded into two contiguous domains (Fig. 3): a central domain consisting of a β -sheet (strands β 1–8) flanked by six helices (α A–F) that forms a canonical α/β hydrolase fold; and a cap domain located between β 6 and α D that consists entirely of helices (α D1–6). In general, α/β hydrolase enzymes share similar topological features without extensive sequence homology, and large insertions are common (Ollis *et al.*, 1992). Indeed, multimeric PAP sequences share low sequence identity and are ~120 amino acids larger than their monomeric counterparts. Secondary structure analysis predicted TePAP to be composed of an α/β hydrolase fold similar to XcPIP with additional helices in the cap domain. This larger cap domain is a common feature of all PAPs in the S33.008 subfamily and therefore may play a role in multimerization.

The existence of PAPs in vastly diverse organisms and unrelated phyla (e.g. actinomycetes and proteobacteria) suggests that these unusual enzymes carry out conserved roles central to biological processes in these organisms. The presence of PAP activity may confer an adaptive advantage, as competing micro-organisms lacking this enzyme (such as *Saccharomyces cerevisiae*) would be unable to completely degrade high-proline-content peptides. Proline-rich proteins are abundant in nature and include plant cell wall proteins and glutens. However, the significance of proline liberation and its functions within this distinct set of micro-organisms remains to be elucidated. Bacterial PAPs are found in both S33.001 and S33.008 subfamilies with some actinomycetes and proteobacteria in opposing clades. This suggests that a subset of bacterial PAPs are more closely related to their eukaryotic counterparts than to other bacterial PAPs.

The reported role of PAPs in bacteria as virulence factors (Felipe *et al.*, 2005) suggests that invasive organisms use this peptidase to degrade proline-rich host proteins, such as collagen in animals. Interestingly, phylogenetic analysis suggests two subclades within the S33.001 family, one of which consists of PAPs from pathogenic species that include *X. campestris*, *S. marcescens* and *N. gonorrhoeae*, while the other consists predominantly of non-pathogenic organisms. This subgrouping is not apparent in S33.008, as PAPs have only been characterized from non-pathogenic species.

In conclusion, TePAP is the second PAP to be characterized and cloned from the fungal kingdom. Comparative genomic analysis described here predicts the widespread occurrence of PAPs in many fungal species (Supplementary Table S4). The presence of additional sequence motifs (α M1–4), identified in this study, indicates that all eukaryotic PAPs exist as multimers. Although biochemical studies with TePAP show many similarities to PAPs from other organisms, we have identified two novel features of this peptidase. Firstly, unlike other fungal aminopeptidases, TePAP is not constitutively expressed. Possessing an inducible PAP may enable *T. emersonii* to adapt to and thrive on a greater diversity of nutrient sources. Secondly, this study reveals a relationship between extracellular protein degradation and intracellular induction of peptidases. The correlation between the induction and specificity of both TGP1 and TePAP suggests the existence of a pathway that regulates concerted peptidase action.

ACKNOWLEDGEMENTS

This work was supported in part by Enterprise Ireland Research Innovation Fund, Higher Education Authority PRTLI and Department of Agriculture, Food and Fisheries awards to M.G.T., and in part by National Institutes of Health Grant AI067423 to C.S.C. C.S.M. was a recipient of a National University of Ireland Galway research scholarship and A.J.O'D received funding awards from the Irish Research Council for Science, Engineering, and Technology, Enterprise Ireland, County Cork Vocational Education Committee and an education abroad program fellowship from the National University of Ireland, Galway.

REFERENCES

- Allen, P. J. & Emerson, R. (1949). Guayule rubber: microbiological improvement by shrub retting. *Ind Eng Chem* **41**, 346–365.
- Basten, D. E., Visser, J. & Schaap, P. J. (2001). Lysine aminopeptidase of *Aspergillus niger*. *Microbiology* **147**, 2045–2050.
- Basten, D. E., Dekker, P. J. & Schaap, P. J. (2003). Aminopeptidase C of *Aspergillus niger* is a novel phenylalanine aminopeptidase. *Appl Environ Microbiol* **69**, 1246–1250.
- Basten, D. E., Moers, A. P., Ooyen, A. J. & Schaap, P. J. (2005). Characterisation of *Aspergillus niger* prolyl aminopeptidase. *Mol Genet Genomics* **272**, 673–679.
- Bendtsen, J. D., Nielsen, H., von Heijne, G. & Brunak, S. (2004). Improved prediction of signal peptides: SignalP 3.0. *J Mol Biol* **340**, 783–795.
- Bensadoun, A. & Weinstein, D. (1976). Assay of proteins in the presence of interfering materials. *Anal Biochem* **70**, 241–250.
- Bolumar, T., Sanz, Y., Aristoy, M. C. & Toldra, F. (2003). Purification and characterization of a prolyl aminopeptidase from *Debaryomyces hansenii*. *Appl Environ Microbiol* **69**, 227–232.
- Cunningham, D. F. & O'Connor, B. (1997). Proline specific peptidases. *Biochim Biophys Acta* **1343**, 160–186.
- Doi, R. H. (2008). Cellulases of mesophilic microorganisms: cellulosome and noncellulosome producers. *Ann N Y Acad Sci* **1125**, 267–279.
- Felipe, M. S., Andrade, R. V., Arraes, F. B., Nicola, A. M., Maranhão, A. Q., Torres, F. A., Silva-Pereira, I., Poças-Fonseca, M. J., Campos, E. G. & other authors (2005). Transcriptional profiles of the human

pathogenic fungus *Paracoccidioides brasiliensis* in mycelium and yeast cells. *J Biol Chem* **280**, 24706–24714.

Fickers, P., Nicaud, J. M., Gaillardin, C., Destain, J. & Thonart, P. (2004). Carbon and nitrogen sources modulate lipase production in the yeast *Yarrowia lipolytica*. *J Appl Microbiol* **96**, 742–749.

Fuke, Y. & Matsuoka, H. (1993). The purification and characterization of prolyl aminopeptidase from *Penicillium camemberti*. *J Dairy Sci* **76**, 2478–2484.

Grassick, A., Murray, P. G., Thompson, R., Collins, C. M., Byrnes, L., Birrane, G., Higgins, T. M. & Tuohy, M. G. (2004). Three-dimensional structure of a thermostable native cellobiohydrolase, CBH IB, and molecular characterization of the *cel7* gene from the filamentous fungus, *Talaromyces emersonii*. *Eur J Biochem* **271**, 4495–4506.

Heneghan, M. N., McLoughlin, L., Murray, P. G. & Tuohy, M. G. (2007). Cloning, characterisation and expression analysis of α -glucuronidase from the thermophilic fungus *Talaromyces emersonii*. *Enzyme Microb Technol* **41**, 677–682.

Hiwatashi, K., Hori, K., Takahashi, K., Kagaya, A., Inoue, S., Sugiyama, T. & Takahashi, S. (2004). Purification and characterization of a novel prolyl aminopeptidase from Maitake (*Grifola frondosa*). *Biosci Biotechnol Biochem* **68**, 1395–1397.

Inoue, T., Ito, K., Tozaka, T., Hatakeyama, S., Tanaka, N., Nakamura, K. T. & Yoshimoto, T. (2003). Novel inhibitor for prolyl aminopeptidase from *Serratia marcescens* and studies on the mechanism of substrate recognition of the enzyme using the inhibitor. *Arch Biochem Biophys* **416**, 147–154.

Ito, K., Inoue, T., Kabashima, T., Kanada, N., Huang, H. S., Ma, X., Azmi, N., Azab, E. & Yoshimoto, T. (2000). Substrate recognition mechanism of prolyl aminopeptidase from *Serratia marcescens*. *J Biochem* **128**, 673–678.

Jiménez, C. R., Huang, L., Qiu, Y. & Burlingame, A. L. (2001). In-gel digestion of proteins for MALDI-MS fingerprint mapping. *Curr Protoc Protein Sci Chapter 16*, Unit 16.4.

Kato, N., Murakoshi, Y., Kato, M., Kobayashi, T. & Tsukagoshi, N. (2002). Isomaltose formed by α -glucosidases triggers amylase induction in *Aspergillus nidulans*. *Curr Genet* **42**, 43–50.

Kitazono, A., Kitano, A., Kabashima, T., Ito, K. & Yoshimoto, T. (1996). Prolyl aminopeptidase is also present in *Enterobacteriaceae*: cloning and sequencing of the *Hafnia alvei* enzyme-gene and characterization of the expressed enzyme. *J Biochem* **119**, 468–474.

Kredics, L., Antal, Z., Szekeres, A., Hatvani, L., Manczinger, L., Vagvolgyi, C. & Nagy, E. (2005). Extracellular proteases of *Trichoderma* species. A review. *Acta Microbiol Immunol Hung* **52**, 169–184.

Larkin, M. A., Blackshields, G., Brown, N. P., Chenna, R., McGettigan, P. A., McWilliam, H., Valentin, F., Wallace, I. M., Wilm, A. & other authors (2007). CLUSTAL W and CLUSTAL_X version 2.0. *Bioinformatics* **23**, 2947–2948.

McGuffin, L. J., Bryson, K. & Jones, D. T. (2000). The PSIPRED protein structure prediction server. *Bioinformatics* **16**, 404–405.

Medrano, F. J., Alonso, J., Garcia, J. L., Romero, A., Bode, W. & Gomis-Ruth, F. X. (1998). Structure of proline iminopeptidase from *Xanthomonas campestris* pv. *citri*: a prototype for the prolyl oligopeptidase family. *EMBO J* **17**, 1–9.

Milewski, S., Mignini, F., Prasad, R. & Borowski, E. (2001). Unusual susceptibility of a multidrug-resistant yeast strain to peptidic antifungals. *Antimicrob Agents Chemother* **45**, 223–228.

Morty, R. E., Shih, A. Y., Fulop, V. & Andrews, N. W. (2005). Identification of the reactive cysteine residues in oligopeptidase B from *Trypanosoma brucei*. *FEBS Lett* **579**, 2191–2196.

Murray, P. G., Grassick, A., Laffey, C. D., Cuffe, M. M., Higgins, T., Savage, A. V., Planas, A. & Tuohy, M. G. (2001). Isolation and characterization of a thermostable endo-beta-glucanase active on 1,3-

1,4-beta-D-glucans from the aerobic fungus *Talaromyces emersonii* CBS 814.70. *Enzyme Microb Technol* **29**, 90–98.

Murray, P. G., Collins, C. M., Grassick, A. & Tuohy, M. G. (2003). Molecular cloning, transcriptional, and expression analysis of the first cellulase gene (*cbh2*), encoding cellobiohydrolase II, from the moderately thermophilic fungus *Talaromyces emersonii* and structure prediction of the gene product. *Biochem Biophys Res Commun* **301**, 280–286.

O'Donoghue, A. J., Mahon, C. S., Goetz, D. H., O'Malley, J. M., Gallagher, D. M., Zhou, M., Murray, P. G., Craik, C. S. & Tuohy, M. G. (2008). Inhibition of a secreted glutamic peptidase prevents growth of the fungus *Talaromyces emersonii*. *J Biol Chem* **283**, 29186–29195.

Ollis, D. L., Cheah, E., Cygler, M., Dijkstra, B., Frolow, F., Franken, S. M., Harel, M., Remington, S. J., Silman, I. & other authors (1992). The α/β hydrolase fold. *Protein Eng* **5**, 197–211.

Polaina, J. & MacCabe, A. P. (2007). *Industrial Enzymes: Structure, Function and Applications*. Dordrecht: Springer.

Polizeli, M. L., Rizzatti, A. C., Monti, R., Terenzi, H. F., Jorge, J. A. & Amorim, D. S. (2005). Xylanases from fungi: properties and industrial applications. *Appl Microbiol Biotechnol* **67**, 577–591.

Raeder, U. & Broda, P. (1985). Rapid preparation of DNA from filamentous fungi. *Lett Appl Microbiol* **1**, 17–20.

Reen, F. J., Murray, P. G. & Tuohy, M. G. (2003). Molecular characterisation and expression analysis of the first hemicellulase gene (*bxl1*) encoding β -xylosidase from the thermophilic fungus *Talaromyces emersonii*. *Biochem Biophys Res Commun* **305**, 579–585.

Stricker, A. R., Mach, R. L. & de Graaff, L. H. (2008). Regulation of transcription of cellulases- and hemicellulases-encoding genes in *Aspergillus niger* and *Hypocrea jecorina* (*Trichoderma reesei*). *Appl Microbiol Biotechnol* **78**, 211–220.

Takagi, H. (2008). Proline as a stress protectant in yeast: physiological functions, metabolic regulations, and biotechnological applications. *Appl Microbiol Biotechnol* **81**, 211–223.

Tuohy, M. G., Puls, J., Claeysens, M., Vrsanska, M. & Coughlan, M. P. (1993). The xylan-degrading enzyme system of *Talaromyces emersonii*: novel enzymes with activity against aryl β -D-xylosides and unsubstituted xylans. *Biochem J* **290**, 515–523.

Tuohy, M. G., Laffey, C. D. & Coughlan, M. P. (1994). Characterization of the individual components of the xylanolytic enzyme system of *Talaromyces emersonii*. *Bioresour Technol* **50**, 37–42.

Tuohy, M. G., Walsh, D. J., Murray, P. G., Claeysens, M., Cuffe, M. M., Savage, A. V. & Coughlan, M. P. (2002). Kinetic parameters and mode of action of the cellobiohydrolases produced by *Talaromyces emersonii*. *Biochim Biophys Acta* **1596**, 366–380.

UniProt (2007). The Universal Protein Resource. *Nucleic Acids Res* **35**, D193–D197.

Varmanen, P., Rantanen, T. & Palva, A. (1996). An operon from *Lactobacillus helveticus* composed of a proline iminopeptidase gene (*pepI*) and two genes coding for putative members of the ABC transporter family of proteins. *Microbiology* **142**, 3459–3468.

Yasothornsrikul, S. & Hook, V. Y. (2000). Detection of proteolytic activity by fluorescent zymogram in-gel assays. *Biotechniques* **28**, 1166–1173.

Yoshimoto, T., Kabashima, T., Uchikawa, K., Inoue, T., Tanaka, N., Nakamura, K. T., Tsuru, M. & Ito, K. (1999). Crystal structure of prolyl aminopeptidase from *Serratia marcescens*. *J Biochem* **126**, 559–565.

Zhang, L., Jia, Y., Wang, L. & Fang, R. (2007). A proline iminopeptidase gene upregulated in planta by a LuxR homologue is essential for pathogenicity of *Xanthomonas campestris* pv. *campestris*. *Mol Microbiol* **65**, 121–136.

Edited by: M. Tien