

A new subfamily of fungal subtilases: structural and functional analysis of a *Pleurotus ostreatus* member

Vincenza Faraco,¹ Gianna Palmieri,² Giovanna Festa,¹ Maria Monti,¹ Giovanni Sanna¹ and Paola Giardina¹

Correspondence
Paola Giardina
giardina@unina.it

¹Dipartimento di Chimica Organica e Biochimica, Università di Napoli 'Federico II', Complesso Universitario Monte S. Angelo, via Cinthia, 80126 Napoli, Italy

²ISPAAM, Consiglio Nazionale delle Ricerche, via Argine 1085, 80147 Napoli, Italy

Pleurotus ostreatus produces several extracellular proteases which are believed to be involved in the regulation of the ligninolytic activities of this fungus. Recently, purification and characterization of the most abundant *P. ostreatus* extracellular protease (*PoSI*) have been reported. The sequence of the *posl* gene and of the corresponding cDNA has been determined, allowing the identification of its pre- and pro-sequences. A mature protein sequence has been verified by mass spectrometry mapping, the *N*-glycosylation sites have been identified and the glycosidic moieties characterized. Mature *PoSI* shows a cleaved peptide bond in the C-terminal region, which remains associated with the catalytic domain in a non-covalent complex. Reported results indicate that this enzyme is involved in the activation of other *P. ostreatus* secreted proteases, thus suggesting its leading role in cascade activation mechanisms. Analyses of the *PoSI* sequence by homology search resulted in the identification of a DNA sequence encoding a new protease, homologous to *PoSI*, in the *Phanerochaete chrysosporium* genome. A new subgroup of subtilisin-like proteases, belonging to the pyrolysins family, has been defined, which includes proteases from ascomycete and basidiomycete fungi.

Received 29 June 2004
Revised 13 October 2004
Accepted 15 October 2004

INTRODUCTION

Subtilisin-like serine proteases, subtilases, have been classified into six families on the basis of their amino acid sequences (Siezen & Leunissen, 1997). Members of these families have been extensively studied and the crystallographic structures of some of them have been determined (Bode *et al.*, 1987; Jain *et al.*, 1998; Wright *et al.*, 1969). Site-directed mutagenesis and protein engineering have provided intimate knowledge of structure–function relationships in this class of enzymes (Sroga & Dordick, 2001; Ness *et al.*, 2002). The majority of the subtilases are synthesized as a precursor with a pre- and pro-sequence extension of the N-terminus of the mature protein (Bryan *et al.*, 1995). The pre-sequence acts as a signal peptide, driving translocation through the cell membrane, whilst the pro-sequence acts both as an intramolecular chaperone that guides the correct folding of the mature protein and as a protease self-inhibitor (Yabuta *et al.*, 2001, 2003). The pro-sequence is usually cleaved from the mature protein by autoproteolysis to produce active mature protease. In addition to a myriad of

prokaryotic subtilases, many members of this superfamily have been identified in eukaryotes such as fungi, plants, insects and mammals.

Pleurotus ostreatus and *Phanerochaete chrysosporium* are white-rot basidiomycetes, which belong to different subclasses of ligninolytic micro-organisms, producing distinct patterns of ligninolytic enzymes. They produce extracellular proteases, which are believed to be involved in the regulation of the ligninolytic activities of these fungi (Palmieri *et al.*, 2000; Dosoretz *et al.*, 1990). In particular, it has been demonstrated that different laccase isoenzymes from *Pl. ostreatus* can be specifically degraded or activated during fungal growth by proteases present in the culture broth (Palmieri *et al.*, 2000, 2003). In a recent study, we reported the purification and characterization of the main *Pl. ostreatus* extracellular protease *PoSI* (Palmieri *et al.*, 2001).

On the basis of structural and kinetic properties, *PoSI* appears to be a serine protease belonging to the subtilase family. This enzyme seems to play a key role in the regulation process of *Pl. ostreatus* laccase activities. A similar relationship was observed for lignin peroxidases (LiPs) in *Ph. chrysosporium*: in this case, the extracellular proteases caused an almost complete disappearance of LiP activity due to degradation of all LiP isoenzymes (Dosoretz *et al.*, 1990).

Abbreviations: MALDIMS, Matrix-assisted laser desorption ionization–mass spectrometry; PTH, phenylthiohydantoin.

The EMBL accession numbers for the nucleotide sequences reported in this paper are AJ634913 (*posl*) and AJ748587 (*pcsl*).

This paper reports evidence concerning the role played by *PoSl* in the activation of *Pl. ostreatus* extracellular proteases. The *posl* gene and cDNA were cloned and sequenced, and mass spectrometric analysis was employed to validate the deduced amino acid sequence and to identify post-translational modifications. Furthermore, analyses by homology search allowed us to define a new subtilase subfamily which includes proteases from ascomycete and basidiomycete fungi.

METHODS

Organisms and culture conditions. White-rot fungi, *Pleurotus ostreatus* (Jacq.Fr.) Kummer (type: Florida) (ATCC no. MYA-2306) and *Phanerochaete chrysosporium* Burdsall M1 (DSM 13583) were maintained through periodic transfer at 4 °C on potato glucose (2.4 %) agar plates (Difco) in the presence of 0.5 % yeast extract (Difco).

Pl. ostreatus cultures were carried out in the basal medium as previously described (Palmieri *et al.*, 1997), or with the addition of 150 µM copper sulphate or 1 mM vanillic acid (4-hydroxy-3-methoxybenzoic acid). *Ph. chrysosporium* cultures were carried out in 0.24 % potato glucose broth in the presence of 0.05 % yeast extract.

Protease assay. Protease activity was assayed using SucAAPpNA (*N*-succinyl-Ala-Ala-Pro-Phe-*p*-nitroanilide; Sigma) or azoalbumin as substrates as follows:

- The assay mixture contained 5 mM SucAAPpNA, 10 mM CaCl₂ and 50 mM Tris/HCl buffer, pH 8.0, in a final volume of 1 ml. Hydrolysis of the substrate was followed by absorbance increase at 405 nm ($\epsilon_{405} = 8800 \text{ M}^{-1} \text{ cm}^{-1}$).
- A 400 µl volume of 15 mg ml⁻¹ azoalbumin (Sigma) in 50 mM MOPS buffer, pH 7.0, was incubated with 250 µl of the enzyme

sample at 37 °C for 30 min. To stop the reaction, 650 µl 20 % TCA was added and the undigested precipitated substrate was removed by centrifugation at 10 000 r.p.m. for 15 min. Six hundred and fifty microlitres of supernatant was added to 350 µl 10 M NaOH and the absorbance at 440 nm was measured. The control assay was performed without any enzyme in the reaction mixture and used as reference. One unit of enzyme activity was defined as the amount of enzyme needed to increase the A_{440} by 0.01.

Zymographic analysis. Samples underwent electrophoresis in 10 % gelatin-containing polyacrylamide gel at alkaline pH under non-denaturing conditions. The separating and stacking gels contained 12.5 % acrylamide solution in 50 mM Tris/HCl, pH 9.5, and 9 % acrylamide solution in 18 mM Tris/HCl, pH 7.5, respectively. The electrode reservoir solution was 25 mM Tris/190 mM glycine, pH 8.4. After electrophoresis, gels were incubated for 16 h at 37 °C in 50 mM Tris/HCl, pH 7.6, buffer containing 200 mM NaCl and 5 mM CaCl₂. Gels were then stained for 30 min with 30 % methanol/10 % acetic acid containing 0.5 % Coomassie Brilliant Blue R-250 and destained in the same solution without dye. Clear bands on the blue background represent areas of gelatinolysis.

Cloning of the *posl* gene. Amplification experiments of *Pl. ostreatus* genomic DNA were performed using the oligonucleotide pair 1, as primers (Table 1). The 800 bp fragment obtained was cloned into the pGEM-T Easy Vector (Promega) and sequenced. This fragment, labelled by the random priming method, was used as probe to screen a *Pl. ostreatus* genomic library (Giardina *et al.*, 1995). A further screening of the genomic library was performed using, as probe, a 700 bp fragment obtained by *KpnI* digestion of the amplified cDNA. Colony hybridization experiments were carried out in 5 × SSC at 65 °C (where 1 × SSC is 0.15 M NaCl, 0.015 M sodium citrate).

Cloning of *posl* and *pcsl* cDNAs. Mycelia from *Ph. chrysosporium* and copper-supplemented *Pl. ostreatus* cultures were collected after 2 days of growth. Total RNA was extracted from lyophilized mycelia, as described by Lucas *et al.* (1977). Reverse transcription

Table 1. Oligonucleotide pairs used in amplification experiments

Y = T/C, R = G/A, V = G/C/A, B = G/T/C, D = G/A/T, E = T/C/A, P = A/G/C/T.

Oligonucleotide pair	Sequence	Annealing temperature	
1	<i>posl</i> nterm <i>posl</i> 1 down	5'-GGECCPGAYGAYCCPGC-3' 5'-TCVGTCCADCCRTCPGC-3'	56 °C
2	pep segn <i>posl</i> nter rev <i>posl</i>	5'-ATGAAGGGCGTACTCGTGTGG-3' 5'-GAGCCGGATCGTCAGGGCC-3'	64 °C
3	<i>posl</i> spec nterm <i>posl</i> 1D down	5'-CCACCCGACTCAGAGTCC-3' 5'-GCCTCCCAAAGAAAGGGTG-3'	58 °C
4	<i>posl</i> 1D up dT- <i>NoI</i>	5'-CACCTTTCTTTGGGAGGC-3' 5'-AATTCGCGCCGCTTTTTTTTTTTTTTTT-3'	60 °C
5	5' <i>pcsl</i> phan <i>pcsl</i> phan rev	5'-CGGTCAGACCCATGAGGC-3' 5'-GGACGAGGAATTCTCCGTCG-3'	58 °C
6	<i>pcsl</i> phan fw 3' <i>pcsl</i> phan	5'-CGACGGAGAATTCCTCGTCC-3' 5'-TCAGGAGCTCGGGGCATTG-3'	58 °C
7	nter rev <i>posl</i> d(T)RACE	5'-GAGCCGGATCGTCAGGGCC-3' 5'-GACCACGCGTATCGATGTGCGACTTTTTT- TTTTTTTTTV-3'	65 °C
8	nter rev <i>posl</i> Anchor-RACE	5'-GAGCCGGATCGTCAGGGCC-3' 5'-GACCACGCGTATCGATGTGCGAC-3'	65 °C

reactions were performed using Super Script II RNase H⁻ Reverse transcriptase (Invitrogen) and oligonucleotide dT-*NotI* as primer, following the manufacturer's instructions. Amplification experiments for each *posl* cDNA fragment were performed using oligonucleotide pairs 2, 3 and 4 at the corresponding annealing temperatures (Table 1). Amplification experiments for each *psl* cDNA fragment were performed using oligonucleotide pairs 5 and 6 at the corresponding annealing temperatures (Table 1). The amplified fragments were cloned in the pGEM-T Easy Vector and sequenced.

First-strand cDNA synthesized from *Pl. ostreatus* total RNA using the gene-specific primer *posl* 1D down (Table 1) was used to perform rapid amplification of 5' cDNA end (5'-RACE). Terminal transferase (Roche) was used to add a homopolymeric A-tail to the 3' end of the cDNA. Tailed cDNA was then amplified by PCR, using oligonucleotide pair 7 (Table 1), and the resulting product was reamplified using oligonucleotide pair 8 (Table 1), cloned in pGEM-T Easy Vector and sequenced.

DNA preparation, subcloning and restriction analyses were performed by standard methods according to Sambrook *et al.* (1989). Sequencing by the dideoxy chain-termination method was performed by the CEINGE Sequencing Service (Naples, Italy) using universal and specific oligonucleotide primers.

Amino acid sequence analysis. Automated N-terminal degradation of the protein was performed using a Perkin-Elmer Applied Biosystem 477A pulsed liquid protein sequencer equipped with a model 120A phenylthiohydantoin analyser for the on-line identification and quantification of phenylthiohydantoin (PTH) amino acids.

Matrix-assisted laser desorption ionization-mass spectrometry (MALDIMS) analysis. MALDIMS analyses were carried out with a Voyager DE MALDI Time of Flight mass spectrometer (PerSeptive Biosystems) on both the protein and the peptide mixtures obtained from proteolytic hydrolyses.

Molecular mass determination of whole protease was performed by loading a mixture of sample solution and 3,5-dimethoxy-4-hydroxycinnamic acid (sinapinic acid) on a sample slide and drying the slide *in vacuo*. Mass range was calibrated using apomyoglobin from horse heart (mean molecular mass, 16952.5 Da) and human serum albumin (mean molecular mass, 66431.0 Da).

Peptide mixtures derived from each hydrolysis were lyophilized and successively dissolved in 0.2% trifluoroacetic acid, and 1 μ l sample solution was mixed with 1 μ l 10 mg ml⁻¹ α -cyano-4-hydroxycinnamic acid (CHCA) solution on a sample slide and left to dry under vacuum. The CHCA solution was prepared in acetonitrile:0.1% TFA, 70:30 (v/v). Mass spectra were acquired in a linear mode, and calibrated from 379.35 and 5734.59, the *m/z* values of the CHCA dimer and bovine insulin, respectively.

Enzymic hydrolysis. A *PoSl* sample (1 nmol) was reduced with 100 mM DTT at 37 °C for 2 h under a nitrogen atmosphere in 0.25 M Tris/HCl (pH 8.5), 1.25 mM EDTA, containing 6 M guanidinium chloride, and then alkylated for 30 min by using an excess of iodoacetamide at room temperature in the dark. The protein sample was desalted by loading the reaction mixture onto a PD-10 prepacked column (Pharmacia), equilibrated and eluted in 0.4% ammonium bicarbonate, pH 8.5.

Enzymic digestions with trypsin and endoproteinase Glu-C were carried out in 0.4% ammonium bicarbonate, pH 8.5, at 37 °C overnight using an enzyme/substrate molar ratio of 1:50.

The tryptic mixture was deglycosylated by incubation with 0.15 U of peptide *N*-glycosidase F (Boehringer Mannheim) in 0.4% ammonium bicarbonate, pH 8.5, at 37 °C for 16 h. Samples were fractionated on a prepacked cartridge Sep-Pak C18 (Waters); peptide fractions were collected manually.

RESULTS

Protease induction and activity regulation

Different substances, already tested as inducers of *Pl. ostreatus* laccase production, were analysed with respect to their ability to induce protease production; among them, vanillic acid was shown to be the best protease inducer. Fig. 1 shows the time-course of extracellular protease activity from *Pl. ostreatus* culture broth supplemented with 1 mM vanillic acid compared to that of basal or copper-supplemented culture media. SucAAPFpNa, a subtilisin-like protease substrate, or azoalbumin, a non-specific

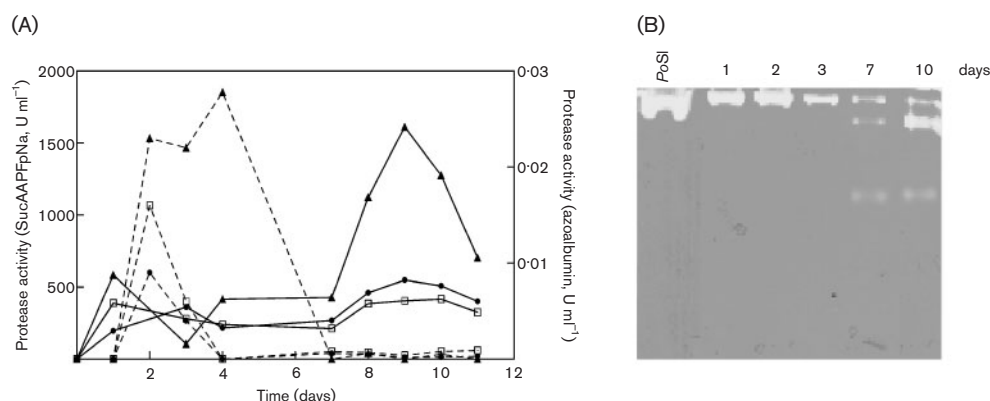


Fig. 1. (A) Time-course of *Pl. ostreatus* extracellular protease activity from cultures supplemented with copper sulphate (□) or vanillic acid (▲), in comparison to basal medium culture (●). Protease activity was determined using SucAAPFpNa as substrate (dashed line) or azoalbumin (solid line). (B) Gelatin zymography of samples collected at different growth times from vanillic acid-supplemented culture.

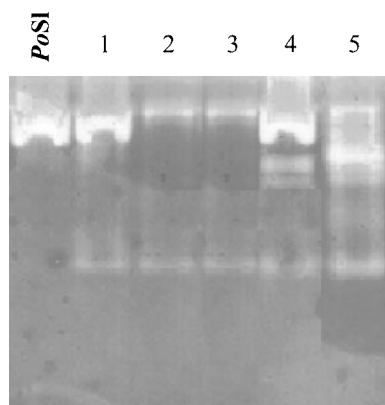


Fig. 2. Zymogram of filtered culture broth (supplemented with vanillic acid) incubated with 1 mM PMSF or 10 mM CaCl_2 . Lanes: 1, 3-day-old filtered culture broth (sample A); 2, 3, sample A after PMSF addition (lane 2) and 24 h incubation (lane 3); 4, sample A after CaCl_2 addition and 24 h incubation; 5, 9-day-old filtered culture broth.

protease substrate, was used as substrate. The gelatin zymography of samples from vanillic acid culture is shown in the same figure. The protease activity determined using SucAAPFPNa as substrate, as well as the zymogram analysis, indicated that the maximum activity of *PoSI* was reached in the first days of growth. New active protease bands appeared at longer growth times, and a remarkable increase of total protease activity (assayed with azoalbumin) was observed. The occurrence of these new protease activities at prolonged growth times could arise from their late production or from maturation of inactive precursors. In order to distinguish between these two hypotheses, samples of culture broth at the third day of fungal growth were filtered through a 0.2- μm -pore-size membrane and incubated for 24 h at 28 °C in the presence of 1 mM PMSF (an irreversible inhibitor of serine proteases) or 10 mM CaCl_2 , which strongly increases *PoSI* activity (Palmieri *et al.*, 2001). Fig. 2 shows the zymogram analysis of these samples together with culture broth samples collected after 3 or 9 days of growth. No *PoSI* activity signal was detected in the PMSF-treated sample, and the protease activity bands of this sample were unaltered after 24 h incubation. On the other hand, when the incubation was performed in the presence of CaCl_2 , new protease activity bands were switched on, displaying a pattern similar to that observed at prolonged fungal growth time (i.e. the ninth day).

Cloning and sequencing of the *posl* gene and cDNA

Sequences of *PoSI* N-terminus and of three tryptic peptides have previously been reported (Palmieri *et al.*, 2001). On the basis of the sequences of the N-terminus and that of a tryptic peptide (boxed peptides in Fig. 3), oligonucleotide primer mixtures were designed (*posl* nterm and *posl* 1 down, Table 1) and used in amplification experiments performed

using *Pl. ostreatus* genomic DNA as template. The amino acid sequence encoded by the 800 bp amplified fragment contained the sequence of another tryptic peptide, homologous to the conserved sequence around the Asp residue of the catalytic triad in the subtilase family. To amplify the *PoSI*-encoding cDNA, two amplification experiments were performed using oligonucleotide pairs 3 and 4 (Table 1), designed on the basis of the nucleotide sequence of the amplified gene fragment. Two amplified fragments of 500 bp and 2000 bp, corresponding to 5' and 3' regions of *posl* cDNA, respectively, were cloned and sequenced.

A *Pl. ostreatus* genomic library was screened using the gene-amplified fragment. One of the positive clones analysed (named 6C4, 2500 bp) encompassed the 5' coding sequence and extended 750 bp upstream from the codon corresponding to the N-terminal of the mature protein. To complete the 3' coding region of the *posl* gene, a 700 bp fragment of the amplified 2000 bp cDNA was used as probe for further screening of the genomic library. One of the positive clones analysed (5–13) overlapped clone 6C4 for 1600 bp and extended 60 bp at the 3' non-coding region. Thus, the whole *posl* gene sequence was determined. A putative start codon (ATG 166–168) was identified in the 5' gene region, 600 bp upstream from the codon corresponding to the N-terminal of the mature protein. The corresponding cDNA fragment was amplified (oligonucleotide pair 2, Table 1), cloned and sequenced. The strategy used to clone the *posl* gene and cDNA is shown in Fig. 4. All known *PoSI* peptide sequences were identified in the encoded sequence. Comparison of cDNA and gene sequences allowed determination of the gene structure, with the coding sequence interrupted by 20 introns (Fig. 4).

A 5'-RACE experiment was performed allowing the identification of the transcription initiation site at nucleotide 145, 21 bp upstream from the ATG 166–168, thus confirming it as the translation start codon.

Validation of the *PoSI* primary structure and characterization of its glycoside moiety

Comparison of the *PoSI* amino acid sequence with that of the mature protein N-terminus allowed us to identify the putative *PoSI* pre-propeptide. The putative pre-peptide (–124 –108) was identified as a secretion signal by the program SignalP; consequently the propeptide was established as peptide (–107 –1).

The mature protein N-terminal sequence had previously been determined from the electroblotted purified protein (GPDDPALPPD...) (Palmieri *et al.*, 2001). Edman degradation of the purified native protein actually gave rise to two equimolar PTH amino acids for each step. Other than the above N-terminal sequence, a second one (AQVPTLGTVEFL), corresponding to a polypeptide chain starting from A690 of *PoSI*, was detectable. Hence, it can be inferred that the F689-A690 peptide bond is hydrolysed in the mature protein and the cleaved peptide remains associated with the

-124 MKGVLVWSIL SAGCALAVTP LSSVKRVTNL PIVANKFIVE VDSTSDIPTK
 -74 RSNPHDALYE SLRKRAIGFK VDREFNTQGL FVGAALTLSD AQDVGQIQST
 -24 TGVKSIRPVR AFDPRPKPVKV QVVK GDDPA LPPDSESTHV ITGVDKLHAQ
 27 GITGAGIKIG ILDTGIDYTH PTLGGAFGPG NKVIGGFDLV GDDYDGTNTP
 77 VPDPDPLDQC AGHGHVAGI IGANPDNAFN* ISGVAFQASL SAYRVFGCVG
 127 FVTDDVLVDA LLLGFNEGQD ILTSLGCGAD GWTBSVSAVV ASRIAATGKV
 177 VTIAAGNDGA SGAWYTSSPG NGIDVISVAS LDNTVVPLQT AQVHGVTNDP
 227 IIYQDTFPLP INDTRPIFAT STDVTVVDDA CNPLPDSTPD LSGFVVIVRR
 277 GTCTFVQKLG NVAAGKAKVT LIYDNGSGFG AIDTGAFVAS FIQAADGEFL
 327 VQQFASGQNV SLSFPQSGGL IQFPDPAGGL ISSFTSYGFS NDDFFKPAVA
 377 APGGNILSTL PVNLGLFGIE SGTSMATPFV AGSAALLFQA KGTSAAVGRS
 427 ARTLFETTAQ RVPSTHTDGD PLQTLTQQA GLINVFNAIH TTIIVSPGEL
 477 ILNDTAHFKG TQQTIVRNAG SSAKTYTLRH IPAGTAVTVT PGTIFPADGP
 527 VPLSTDFASV SLSTSKFTLA PGKTQVTAR FTFPAAADPS TFPVFSGFIQ
 577 IESGTEQVQV SYLGLKASLK DKQVIDNTDF FFGVPTPVLV DPNGEVQTSF
 627 RNYSFLASDF PTLIFRLAFG SPKVVVDLVS PTINFKPTLN TRALGDTHGS
 677 FFSFPHRVKT GSFAQVPTLG TVFELDFTSR NNDPDTDNGF NTVAI DPPTF
 727 ANGTTIPNGQ YRLLLRALRV TGDPTNEADF ESFLSPIIGV NAP

Fig. 3. *PoSI* sequence. The numbers represent the positions of the amino acid residues starting from the N-terminus of the mature *PoSI* protein. Regions verified by peptide mass mapping are underlined. *N*-Glycosylated asparagine residues are marked with one star, asparagine residues found both glycosylated and non-glycosylated are marked with two stars. Boxed peptide sequences were used to design oligonucleotide-primer mixtures for the amplification experiments. The arrow indicates the peptide bond cleaved in the mature protein.

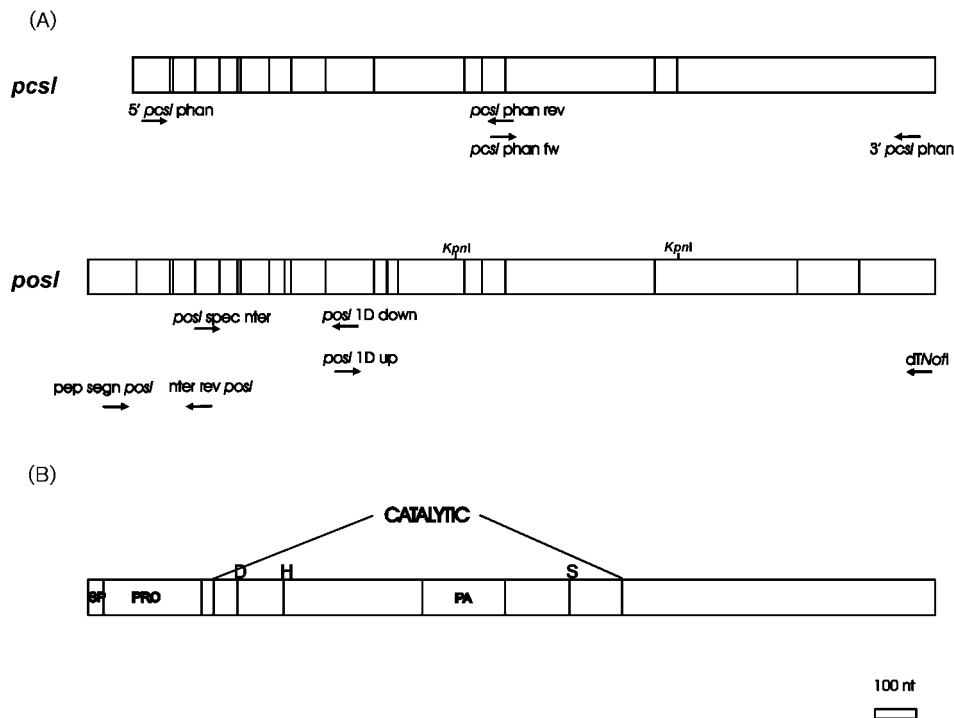


Fig. 4. (A) Schematic representation of *posI* and *pcsI* cDNAs. Arrows show the positions of oligonucleotide primers used for cDNA amplifications. *KpnI* restriction sites at the ends of the probe used for genomic library screening are also shown. The positions of introns are shown by vertical bars. (B) *PoSI* domains. SP, signal peptide; PRO, propeptide; CATALYTIC, catalytic domain; PA, PA domain; D, H, S, catalytic active site residues.

catalytic domain in a non-covalent complex, since no Cys residues in the cleaved peptide are present.

The *PoSl* amino acid sequence is characterized by seven putative glycosylation sites. As previously reported (Palmieri *et al.*, 2001), MALDIMS analysis of permethylated *N*-linked sugar released by hydrolysis with *N*-glycosidase F showed the occurrence of several high mannose moieties with molecular mass ranging from 1580.0 and 2603.6 *m/z* and identified as Hex₅HexNAc₂ and Hex₁₀HexNAc₂. The other molecular ions at *m/z* 1784.0, 1987.5, 2191.5 and 2394.6 were identified as homologous structures having between three and six mannose residues linked to the pentasaccharide core.

To verify the deduced *PoSl* amino acid sequence, the *PoSl* protein was reduced, alkylated and then digested with trypsin, and the peptide mixture was directly analysed by MALDIMS, allowing the identification of several peptides (Fig. 3). This analysis also showed three different clusters of signals each characterized by a pattern of *m/z* values differing in 162 Da, in agreement with the presence of the high mannose glycosidic moiety, as described above. The assignments of the different molecular masses to the corresponding peptides (Table 2) led to the identification of Asn106, Asn479 and Asn728 as *N*-glycosylation sites. As shown in Fig. 3, the two residues Asn479 and Asn728 were present in both glycosylated and unglycosylated state. A further picture of *PoSl* *N*-glycosylation sites was obtained after deglycosylation of peptide mixture by *N*-glycosidase F. Indeed, new signals were detected by MALDIMS analysis at 3440.1, 5087.1, 5484.1 and 6361.2 *m/z*, corresponding to the expected molecular mass of peptides 707–738, 438–485, 59–112 and 59–120, respectively, each increased by 1 Da, since the *N*-glycosylated Asn residues are converted to Asp

following *N*-glycosidase F treatment. Moreover, an additional peak at *m/z* 1793.0 was present after deglycosylation. This mass value is in agreement with the expected molecular mass of the peptide 628–642 (containing the potential *N*-glycosylation site Asn628) increased by 1 Da. This result led to the assignment of Asn628 as an *N*-glycosylation site, even though the MALDIMS spectrum of the tryptic peptide mixture did not allow us to detect the cluster of *m/z* peaks related to the Asn628 glycoforms. Hence, two out of seven putative *PoSl* *N*-glycosylation sites were found to be glycosylated, Asn479 and Asn728 were found to be both glycosylated and unglycosylated, whilst Asn238 and Asn301 were only found unmodified. MALDIMS analyses did not allow mapping of the last putative *N*-glycosylation site, Asn335. This could be due to an incomplete extraction of the peptides from the gel or to the well-known MALDI suppression phenomena.

To validate further regions of the protein sequence, the tryptic peptide mixture was submitted to enzymic hydrolysis with the endoproteinase Glu-C. The data obtained allowed us subsequently to verify 84% the *PoSl* primary structure, as shown in Fig. 3.

Homology search analysis and *pcsI* cDNA cloning and sequencing

PoSl shows a high level of identity with a hypothetical protein from *Neurospora crassa* (38%, BLAST E score = 2.1×10^{-121}), a subtilisin-like serine protease from the fungus *Metarhizium anisopliae*, Pr1C (39%, E score = 2.0×10^{-114}), and a minor extracellular serine protease from *Bacillus subtilis*, Vpr (31%, E score = 7.1×10^{-41}).

Ph. chrysosporium is the only basidiomycete whose genome has been sequenced to date. BLAST analysis of this genome

Table 2. Theoretical and measured molecular mass values for *PoSl* glycopeptides

MH⁺, singly protonated ion. Values in the theoretical and measured MH⁺ columns are in Da.

Peptide	Glycosylation site	Theoretical MH ⁺	Measured MH ⁺	Glycoforms
628–642	Asn628	1792.0	–	–
707–738	Asn728	4655.5	4657.7	Hex ₅ HexNAc ₂
		4817.6	4820.4	Hex ₆ HexNAc ₂
		4979.7	4982.7	Hex ₇ HexNAc ₂
		5141.8	5143.6	Hex ₈ HexNAc ₂
59–112	Asn106	6699.9	6702.0	Hex ₅ HexNAc ₂
		6862.0	6864.1	Hex ₆ HexNAc ₂
		7024.1	7025.6	Hex ₇ HexNAc ₂
438–485	Asn479	6788.9	6790.4	Hex ₈ HexNAc ₂
		6951.0	6952.8	Hex ₉ HexNAc ₂
		7113.1	7114.6	Hex ₁₀ HexNAc ₂
59–120	Asn106	7576.8	7579.1	Hex ₅ HexNAc ₂
		7738.9	7741.1	Hex ₆ HexNAc ₂
		7901.0	7902.9	Hex ₇ HexNAc ₂
		8063.1	8064.1	Hex ₈ HexNAc ₂

(<http://genome.jgi-psf.org/whiterot1/whiterot1.home.html>) resulted in the identification of two *PoSl* homologous sequences, about 112 kb from each other. One of these, designated pc.18.58.1, appeared to encode a protein which was very similar to *PoSl*. RNA was extracted from *Ph. chrysosporium* and specific cDNA was amplified using oligonucleotide pairs designed on the basis of the genomic sequence and protein homology. The amplified cDNA was longer than pc.18.58.1 at its 5' terminus, even if it did not contain the start translational codon. Moreover, the alignment between pc.18.58.1 and AJ748587 showed a 6 aa insertion in the former, due to an incorrect exon end. The deduced amino acid sequence corresponding to the amplified cDNA showed an identity of 65 % with *PoSl*. Fourteen out of 15 introns in the gene sequence were in the same positions as those of the *posl* gene (Fig. 4).

DISCUSSION

The extracellular subtilisin-like protease *PoSl* from *Pl. ostreatus* was believed to be involved in the activation/degradation mechanism of laccase isoenzymes (Palmieri *et al.*, 2000, 2001). This mechanism seemed to be indirectly controlled by *PoSl*, whose activity appeared to be necessary but not sufficient for laccase post-translational regulation. In this paper we investigated the effect of *PoSl* on other *Pl. ostreatus* extracellular proteases. The addition of vanillic acid to the fungal culture led to an increase both of total protease activity and of *PoSl* activity, with respect to any of the other conditions examined, such as copper-supplemented culture (Fig. 1). In vanillic acid-amended cultures, some protease activities were only detectable at long fungal growth times (Fig. 1). These proteases were also found to be activated after incubation of the early growth time culture broth in the absence of mycelia, but in the presence of active *PoSl*. Therefore these proteases were secreted as inactive forms and subsequently activated in a process governed by *PoSl*.

The *PoSl* coding sequence was determined and the deduced protein primary structure confirmed by MALDIMS mapping. This approach has proven useful for the identification of the *N*-glycosylation sites and characterization of the glycoside moiety of the protein. *PoSl* is homologous to many other serine proteases, especially in the areas surrounding the amino acids that are known to be involved in the active site of subtilisin (D, H and S).

The first amino acid (G1) of the mature protein was identified, allowing determination of the *PoSl* pro-sequence. Pro-sequences are usually removed from the catalytic domain by self-digestion or by another protease upon completion of protein folding. The peptide bond K⁻¹-G¹, which connects the *PoSl* pro-sequence to its catalytic domain, should be cleaved by a protease with trypsin-like specificity different from the known *PoSl* substrate preference (Palmieri *et al.*, 2001). Moreover, no significant activity of *PoSl* towards the chromogenic peptide D-Val-Leu-Lys-*p*-nitroanilide was

observed (data not shown). On the other hand, *PoSl* has a D residue in the position corresponding to E156 in subtilisin BPN'. Subtilases with a negative charge on this residue are reported to have an additional ability to cleave after K at the P1 position (Gron *et al.*, 1992; Voorhorst *et al.*, 1997). Furthermore, all the interactions of the substrate P4-P4' residue side chains with the S1 pocket can contribute to binding; thus it cannot be excluded that *PoSl* zymogen can undergo autolysis to self-remove the pro-peptide from unprocessed protein during the maturation process.

PoSl shows a high identity with hypothetical proteins from *N. crassa* and *Trichoderma reesei*, and Pr1C, a 'bacterial-type' subtilisin-like serine protease from the fungus *M. anisopliae*. Eleven subtilisin-like proteins were identified in the latter fungus (Bagga *et al.*, 2004), ten of which displayed identities ranging from 93 to 98 %, being classified as proteinase K-like subtilisins. The unique exception was the 'bacterial-type' subtilisin Pr1C, which is largely divergent from the others. Among the bacterial subtilases, Vpr, a minor extracellular serine protease from *B. subtilis* (Sloma *et al.*, 1991), shows the highest identity with *PoSl*. This bacterial subtilase belongs to the pyrolysins family, according to the classification of Siezen & Leunissen (1997).

Interestingly, a *Ph. chrysosporium* genomic sequence encoding a hypothetical protein, *PcSl*, showing the highest identity with *PoSl* (65 %), has been identified and the corresponding cDNA was amplified and sequenced, thus demonstrating the expression of the *pcsl* gene in this fungus. The *posl* and *pcsl* genes shared 57 % of their sequences and their intron/exon structures were very similar (Fig. 4). Furthermore, all the glycosylated Asn identified in *PoSl* was located in consensus sequences conserved in *PcSl*, suggesting common structural features between the two proteins. Indeed, a 76 kDa protease, inhibited by PMSF, has previously been reported in *Ph. chrysosporium* culture broth. This enzyme is the most abundant protease produced in the presence of both excess and limiting nitrogen source (Dass *et al.*, 1995). These data suggest that this protease is the one encoded by the *pcsl* gene. Hence, a new subgroup of subtilisin-like proteases from ascomycete and basidiomycete fungi, belonging to the pyrolysins family, can be defined, including the proteases *PoSl* and *PcSl* from the basidiomycete fungi *Pl. ostreatus* and *Ph. chrysosporium*, the hypothetical protein from *N. crassa*, and Pr1C from *M. anisopliae*. In Fig. 5 the alignment of these protease sequences is shown.

The spacing between the catalytic H and S residues in *PoSl*, as well as for several other members of the pyrolysins family and all those of the new fungal pyrolysins subfamily, is approximately 150 aa longer than that of subtilisins (Voorhorst *et al.*, 1996). This insert region has significant similarity to sequences of other protein classes (M8/M33 zinc peptidases, Ring-type Zinc Finger proteins, vacuolar sorting receptor, transferrin receptor), and has been named as a protease-associated (PA) domain (Mahon & Bateman, 2000). The role of the PA domain remains unclear; however, it has been supposed that the position and the size

PoSl	MKGVLVWSIILSAGCALAVTPLSSVKRVNLPVIVANKFIVEVDSTSDIPTKRNPNDALYE	-65
PcSl	-----KIGQTHEALYS	
<i>N. crassa</i>	---MVRRLGLATTLAAASFAQAAHQKAPAVVPGAYIVEYEDSHDPTSILASIKGDATIR	
PriC	-----	
PoSl	SLRKRAIGFKVDREFNTQGLFVGAALFLSDAQDVGQIQSTTGVKSIIRPVRAFDRPKPVKV	-5
PcSl	HMRKRNVGFOVDKEFNPSGLFVGAAVTLQNAADDKNQVAMPGVQAIRPVTLIPPKPVFS	
<i>N. crassa</i>	KDIRHELFGASFOFKDLNKADDLASKVAAMSGVKALYPVRRYSIPEHTVH-----STGS	
PriC	MKLDYELFKGVSQVLDHVDKAHEKAARLADTPAVKVVYPVQLFNNMPKPKVEVIAKDGTR	
PoSl	QVVKGPDDPALFPDSESTHVITGVDKLHAQGITGAGIKIGILDGTIDYTHPTLGGAFGPG	56
PcSl	HVVTKGSDPAVPADTESTHVITGVDKLHAQGITGKGIKIGIIDTGIDFTHPDLGGGIGPG	
<i>N. crassa</i>	AVQEVVAKRDGTNDTFSPLHMTQVNNKFRDSGITGKGIKIAVIDTGVDVYLLHPALGGCFGPG	
PriC	PGLLSSRADDGADIFSPHVMTOVDKLRKGVTKGVKIAVVDTGIDYKHPALGGCFGHG	
PoSl	NKVIIGFDLVGDDYDGTNTVPDPDPLDQCAGHGTAVAGIIGANPDNAFNISGVAFQASL	116
PcSl	FKIIGGFDFVGDAFNSTVPDPFSLDQCNGHGTAVAGIIGANPDNAFNISGVAFDASI	
<i>N. crassa</i>	CLVSYGTDLVGDDFNSTVPDPDMDTONGHSHVLGLLSANTNNPYGIIGAAPDVTL	
PriC	CLVAFGTDLVGDAYDGSNTEHPDPPDMD-CGGHSHVAGIVAAQEN-TFGTGAAPGATL	
PoSl	SAYRVFGCVGFVTDVLDVALLLGFNEGQDILTLSLGGADGWTEVSVAVASRIATGKV	176
PcSl	TSYRIFGCSGSTDDVIVDALLRGVSEGQDILTMSLGGPDGWTENTASVSSRISDMGKI	
<i>N. crassa</i>	GAYRVFGCSGDVGNLIEAYLKAYDDGSDIITASIGGASGWPEDSWAAVSRIVEKGVF	
PriC	GAYRVFGCKQAGNDVLIAAFNOAYQDGANIITASIGGSGWSEDPWAEAVSRIVDKGVF	
PoSl	VTIAAGNDGASCAWYTSPPNGIDVISVASLDNTVVE-LQTAQVHGVTHDPIIYQDTPFL	235
PcSl	VTIAAGNDGADGMFFTSPPGNAIDAVISVASLDNTVVE-LQATVHGVTHDPIIYQDTPFL	
<i>N. crassa</i>	CLVSAGNDGATGIFYASTAANGKRVTAVASVDNLIAPALLSEASYSVANGSLSTFGFTAG	
PriC	CTVSAGNEGAEGIFYASTAANGRRVSAIASYDNVQTPSLIYNKYQIDNGADTKFGYVPS	
PoSl	E----INDTRPIFATSTDVIVDDACNPLPDSTPDLSGFVIVVRRGTCTFVQKLNVAAK	291
PcSl	E----ITDTLPIFATSTNTTIDDDACDPLPDSTPDLSKFIVVRRGTCTFVQKLTNIAAK	
<i>N. crassa</i>	SPSAWANVSLPVWSVNFNTADAANGCEAFDDTDLKSKYIIVLIRRGCTCTFVQKQANAAAK	
PriC	EPASWDGVTLEPAWSNLDPTIPDDGCDFFAANTPDLKSKYIIVLIRRGCSSEAFQKVNNAVAK	
PoSl	GAKVTLIYDN-GSGFGAIDTGAFA-----SFIQAADGEFLVQQFASGQNVSLSFPQ--	342
PcSl	GGNVSLIYDN-GNGEADIDVGNETS-----ALIQAADGEFLVQQFASGANVSLSFPQ--	
<i>N. crassa</i>	GAKYIIYYNN-ASGSTKVDVSAVAD--VKAAMVTSETGAAWIKALQAGTQVTVNMADPE	
PriC	GAKYIVIVYNNNAVGAIPMDLTGVPAGSIKAASM-DGTTGATFINALKDKGKILTKMVS PQ	
PoSl	---SGGLIQFPDPAGGLISSFTSYGFSNDFFFKPAVAAPGGNILLSTLPVNLGLFGIESGT	399
PcSl	---TGASTQFPDPTGGLISSFTSYGPTNDMEFFKPAVAPGGNILLSTFPVPLGSEAVLSGT	
<i>N. crassa</i>	TAPKNLNNFNTATPGFLSTYTSWGPTEYVDVKPQISSPGMILSTYPRALGSYAVLSGT	
PriC	KTDSEVSTSNNTITGGALSTFTSWGPTWEMDTKQFGAVGGNVLSTYPRALGSYAVLSGT	
PoSl	SMATPFVAGSALLFQAAGTSAAVGRSARTLFETTAQRVPSHTDGDPLQTLTQQAGGLI	459
PcSl	SMATPFMAGVSALLFGLKGNTPVGRSARDLFETTAQLVPSHSDGDPLQTLVQQAGGLV	
<i>N. crassa</i>	SMACPLAAATWALVMQKRGTKDKPVLENLFSATAHPNLFNDGKTYPMLAPVAQQAGGLI	
PriC	SMSCPQTAGIIALIHQVRGTYDPELIQNLLSANANPQLFNDGKTFYDFLAPVPOQGGGLV	
PoSl	NVENIHTTTIVSPGELILNDTAHFKGTQOETVRNAGSSAKTYTLRHIPAGTAVVTV-PG	518
PcSl	NAFQALTTDIIITPPELLTNDTANFKSLHTFKIKNTGKTAESFKISHVPAGTAVISLQ-NG	
<i>N. crassa</i>	QAWDAANANALLSVSSISFNDEHFKPLQSFVNTNGKKAIVTYQLGHTSAAATAYTFANDT	
PriC	QAYDAAYATLLSPSSLISFNDEHFVETLSEKLNQNTDSKGIITYKITHTEAMMYALG-DG	
PoSl	TIFPADGPVPLSTDFASVLSLTSKFTLAPGKTQVTVARFTPPAAADPSTFPVFSGFIOIE	578
PcSl	TAFFNDGPVPLTSQFATVKFSETSVTVHPGQTQEIITAHITPESGVDPVLPVFSGFIOIE	
<i>N. crassa</i>	SIGPAAFPNELVDAKATLVETPAKLTLPNGQKKTVTVLAIPELGLDAKRLPVYSYITLN	
PriC	SPSVQQFPNDAVQAAATIKLSETSVTLNGGQSKSISVSPTPPQGLDAKRLALWSGYIVIN	
PoSl	SGTEQ-VQVSYLGLKASLKDQVIDNTDFFFGVPTFVLTDPNCEVQTSFRNYS-----	630
PcSl	SATET-LHVTYLGVAAALKDARVDDTDVFFGVDPVLTDPAGDFLTNATNFT-----	
<i>N. crassa</i>	GTDSTGYSLPYQGVGSMRSVTVLDKQNSYLSQSSDATYAPVAAGTFTFLPAGKANDTL	
PriC	GTDGTSLSLPYQGLTGLSHNSAVLGANNTWISKSTDKKSNVPPNSTFVVPAGNAG---	
PoSl	FLASDFPTLIFRLAFGSPKVVVDLVSPTINFKPTLNTRALGDTHGSFFSFPHRVKTCGSFA	690
PcSl	FVGTDPQLMLRFAFGTELLRADLVDPAEKIATILNKRAD-----AAKGSAA	
<i>N. crassa</i>	YATTVYPTIVLTLMSGSAEVAHADVVNSKG-----	
PriC	-SNDTLPQITVSLYLGRKIRADIVPLTTCPSKNL-----TTEFQ	
PoSl	QVPTLGTVFELDFTSRNNDDPTDNGFNTVAIDPPTFANGTTIPNGQYRLLLRALRVTDGPF	750
PcSl	AIKTLGSLFEFDEQPRNTEDEDDGTDFNVFELDTPVFANGTTIPNGLYKILLRVLKVTGDF	
<i>N. crassa</i>	--KTIQQLTFPARWNP-----GTFEWNWDG-ALSDGYAPADTYKITLRAKLIYNS	
PriC	GIKTIQOPYNFPALWGTR-----GLNTPFDWG-RLDSGNYPAPKQYKVVVRAKIRIFGDE	
PoSl	TNEADFESFLSPIIGVNAE----- 769	
PcSl	NNEADFESWLSPIIGVNAE-----	
<i>N. crassa</i>	KWPLDWETQTEPFTIKYAAKSKRAFTA	
PriC	KKKDDWDVSTSEALHIKYQ-----	

Fig. 5. Sequence alignment. The amino acid sequence of PoSl is compared with those of PcSl from *Ph. chrysosporium*, a putative protein from *N. crassa* (accession no. Q7RZV0) and PR1C from *M. anisopliae* (accession no. Q8X1Y7). Gaps are denoted by dashes. The regions highlighted in grey are conserved among PoSl and the analysed proteins. The amino acid residues that form the catalytic triad are denoted by an asterisk. The numbers represent the positions of the amino acid residues starting from the N-terminus of the mature PoSl protein.

of the inserted domain might allow interference with substrate access and thus it could be involved in substrate specificity (Luo & Hofmann, 2001).

All the members of the new fungal pyrolysins subfamily have a long C-terminal extension, which is less conserved than the catalytic region (Siezen & Leunissen, 1997; Voorhorst *et al.*, 1996). Mature PoSl shows a cleaved peptide bond in this region, probably caused by an autocatalytic event, on the basis of its specificity. The cleaved peptide remains associated with the catalytic domain in a non-covalent complex, as demonstrated by the two N-terminal sequences obtained from the native protein and confirmed by the MALDIMS characterization of the protein. This finding suggests that the cleaved bond belongs to an exposed loop of the protein and that, after the proteolytic event, the newly generated peptide remains tightly bound to the core of the protein.

The occurrence of this class of proteases in a wide variety of fungal genera could be suggestive of the crucial physiological role played by these proteins in the micro-organisms which produce them. With regard to *Pl. ostreatus*, the role proposed for PoSl, a member of this subfamily, would be to start a cascade of proteolytic reactions, leading to the activation of other extracellular proteases.

ACKNOWLEDGEMENTS

This work was supported by grants from the Ministero dell'Università e della Ricerca Scientifica (Progetti di Rilevante Interesse Nazionale, PRIN) and from Centro Regionale di Competenza BioTekNet. The authors thank Professor Gennaro Marino for helpful discussions and for the original suggestion of the occurrence of a proteolytic cascade.

REFERENCES

- Bagga, S., Hu, G., Screen, S. E. & St Leger, R. J. (2004). Reconstructing the diversification of subtilisins in the pathogenic fungus *Metarhizium anisopliae*. *Gene* **324**, 159–169.
- Bode, W., Papamokos, E. & Musil, D. (1987). The high-resolution X-ray crystal structure of the complex formed between subtilisin Carlsberg and eglin c, an elastase inhibitor from the leech *Hirudo medicinalis*. Structural analysis, subtilisin structure and interface geometry. *Eur J Biochem* **166**, 673–692.
- Bryan, P., Wang, L., Hoskins, J., Ruvinov, S., Strausberg, S., Alexander, P., Almog, O., Gilliland, G. & Gallagher, T. (1995). Catalysis of a protein folding reaction: mechanistic implications of the 2.0 Å structure of the subtilisin-prodomain complex. *Biochemistry* **34**, 10310–10318.
- Dass, S. B., Dosoretz, C. G., Reddy, C. A. & Grethlein, H. E. (1995). Extracellular proteases produced by the wood-degrading fungus *Phanerochaete chrysosporium* under ligninolytic and non-ligninolytic conditions. *Arch Microbiol* **163**, 254–258.
- Dosoretz, C. G., Dass, B., Reddy, A. & Grethlein, H. E. (1990). Protease-mediated degradation of lignin peroxidase in liquid cultures of *Phanerochaete chrysosporium*. *Appl Environ Microbiol* **56**, 395–400.
- Giardina, P., Cannio, R., Martirani, L., Marzullo, L., Palmieri, G. & Sannia, G. (1995). Cloning and sequencing of a laccase from the lignin degrading basidiomycete *Pleurotus ostreatus*. *Appl Environ Microbiol* **61**, 2408–2413.
- Gron, H., Meldal, M. & Breddam, K. (1992). Extensive comparison of the substrate preferences of two subtilisins as determined with peptide substrates which are based on the principle of intramolecular quenching. *Biochemistry* **26**, 6011–6018.
- Jain, S. C., Shinde, U., Li, Y., Inouye, M. & Barman, H. M. (1998). The crystal structure of an autoprocessed Ser221Cys-subtilisin E-propeptide complex at 2.0 Å resolution. *J Mol Biol* **284**, 137–144.
- Lucas, M. C., Jacobson, J. W. & Giles, N. H. (1977). Characterization and in vitro translation of polyadenylated messenger ribonucleic acid from *Neurospora crassa*. *J Bacteriol* **130**, 1192–1198.
- Luo, X. & Hofmann, K. (2001). The protease-associated domain: a homology domain associated with multiple classes of proteases. *Trends Biochem Sci* **26**, 147–148.
- Mahon, P. & Bateman, A. (2000). The PA domain: a protease-associated domain. *Protein Sci* **9**, 1930–1934.
- Ness, J. E., Kim, S., Gottman, A., Pak, R., Krebber, A., Borchert, T. V., Govindarajan, S., Mundorff, E. C. & Minshall, J. (2002). Synthetic shuffling expands functional protein diversity by allowing amino acids to recombine independently. *J Nat Biotechnol* **20**, 1251–1255.
- Palmieri, G., Giardina, P., Bianco, C., Scaloni, A., Capasso, A. & Sannia, G. (1997). A novel white laccase from *Pleurotus ostreatus*. *J Biol Chem* **272**, 31301–31307.
- Palmieri, G., Giardina, P., Bianco, C., Fontanella, B. & Sannia, G. (2000). Copper induction of laccase isoenzymes in the ligninolytic fungus *Pleurotus ostreatus*. *Appl Environ Microbiol* **66**, 920–924.
- Palmieri, G., Bianco, C., Cennamo, G., Giardina, P., Marino, G., Monti, M. & Sannia, G. (2001). Purification, characterization, and functional role of a novel extracellular protease from *Pleurotus ostreatus*. *Appl Environ Microbiol* **67**, 2754–2759.
- Palmieri, G., Cennamo, G., Faraco, V., Amoresano, A., Sannia, G. & Giardina, P. (2003). Atypical laccase isoenzymes from copper supplemented *Pleurotus ostreatus* cultures. *Enzyme Microb Technol* **33**, 220–230.
- Sambrook, J., Fritsch, E. F. & Maniatis, T. (1989). *Molecular Cloning: a Laboratory Manual*, 2nd edn. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory.
- Siezen, R. J. & Leunissen, J. A. (1997). Subtilases: the superfamily of subtilisin-like serine proteases. *Protein Sci* **6**, 501–523.
- Sloma, A., Rufo, G. A., Jr, Theriault, K. A., Dwyer, M., Wilson, S. W. & Pero, J. (1991). Cloning and characterization of the gene for an additional extracellular serine protease of *Bacillus subtilis*. *J Bacteriol* **173**, 6889–6895.
- Sroga, G. E. & Dordick, J. S. (2001). Generation of a broad esterolytic subtilisin using combined molecular evolution and periplasmic expression. *Protein Eng* **14**, 929–937.

Voorhorst, W. G., Eggen, R. I., Geerling, A. C., Platteeuw, C., Siezen, R. J. & Vos, W. M. (1996). Isolation and characterization of the hyperthermostable serine protease, pyrolysin, and its gene from the hyperthermophilic archaeon *Pyrococcus furiosus*. *J Biol Chem* **271**, 20426–20431.

Voorhorst, W. G., Warner, A., de Vos, W. M. & Siezen, R. J. (1997). Homology modelling of two subtilisin-like proteases from the hyperthermophilic archaea *Pyrococcus furiosus* and *Thermococcus stetteri*. *Protein Eng* **10**, 905–914.

Wright, C. S., Alden, R. A. & Kraut, J. (1969). Structure of subtilisin BPN' at 2.5 angstrom resolution. *Nature* **221**, 235–242.

Yabuta, Y., Takagi, H., Inouye, M. & Shinde, U. (2001). Folding pathway mediated by an intramolecular chaperone: propeptide release modulates activation precision of pro-subtilisin. *J Biol Chem* **276**, 44427–44434.

Yabuta, Y., Subbian, E., Oiry, C. & Shinde, U. (2003). Folding pathway mediated by an intramolecular chaperone. A functional peptide chaperone designed using sequence databases. *J Biol Chem* **278**, 15246–15251.