Gene, 167 (1995) 167–171

GENE 09405

Cloning, structural organization and regulation of expression of the *Penicillium chrysogenum paf* gene encoding an abundantly secreted protein with antifungal activity

(Molecular cloning; gene structure; secretory protein; transcriptional regulation)

Florentine Marx[a], Hubertus Haas[a], Markus Reindl[a], Georg Stöffler[a], Friedrich Lottspeich[b] and Bernhard Redl[a]

[a]*Institut für Mikrobiologie (Med. Fakultät), Universität Innsbruck, A-6020 Innsbruck, Austria; and* [b]*Max-Planck-Institut für Biochemie, D-82152 Martinsried bei München, Germany. Tel. (49-89) 85781*

Received by J.R. Kinghorn: 1 June 1995; Received/Accepted: 22 August/23 August 1995; Received at publishers: 28 September 1995

SUMMARY

An abundantly secreted, highly basic 12-kDa protein (PAF) was purified from the culture medium of *Penicillium chrysogenum* (*Pc*). Based on the N-terminal amino acid (aa) sequence of the protein, an oligodeoxyribonucleotide probe was derived and used for amplification of the encoding cDNA by PCR. This cDNA fragment encodes a Cys-rich preproprotein of 92 aa which appears to be processed to a mature product of 55 aa. The deduced aa sequence of the preproprotein reveals 42.6% identity to an antifungal protein (AFP) of *Aspergillus giganteus*. Agar diffusion tests confirmed that the *Pc* protein exhibits antifungal activity. In order to investigate the promoter region and the structural organization of the *paf* gene, a genomic 6-kb fragment was isolated and partially sequenced. Comparison of the nucleotide sequence of the genomic fragment and the cDNA clone revealed the presence of a coding region of 279 bp which is interrupted by two introns of 76 and 68 bp in length. In the promoter region, a typical TATA box, a motif resembling the fungal carbon catabolite repression element, as well as several putative GATA factor binding motifs, were found. Northern blot analysis indicated that the regulation of *paf* expression occurs at the level of mRNA transcription and is under control of carbon catabolite and nitrogen metabolite repression regulatory circuits.

INTRODUCTION

The high level of secretion of native proteins by filamentous fungi clearly identifies these organisms as attractive hosts for the production of secreted heterologous proteins (Van den Hondel et al., 1991; Jeenes et al., 1991). In addition, some like *Aspergillus* and *Penicillium* have been used in biotechnology for years and therefore advanced technologies for their large scale cultivation are available. A prerequisite for any expression system is a suitable promoter and some knowledge about the secretion process. In contrast to *Aspergillus*, only few promoters have been characterized in *Penicillium* so far. In order to isolate and analyze suitable promoters and to study

Correspondence to: Dr. B. Redl, Institut für Mikrobiologie (Medizinische Fakultät), Universität Innsbruck, Fritz-Pregl-Str. 3, A-6020 Innsbruck, Austria. Tel. (43-512) 507-3603; Fax (43-512) 507-2866;e-mail: bernhard.redl@uibk.ac.at

Abbreviations: *A.*, *Aspergillus*; aa, amino acid(s); AFP, *Ag* antifungal protein; *Ag*, *A. giganteus*; *An*, *A. niger*; bp, base pair(s); cDNA, complementary DNA; CREA, repressor mediating carbon catabolite repression in *A. nidulans*; HPLC, high-performance liquid chromatography; kb, kilobase(s) or 1000 bp; nt, nucleotide(s); oligo, oligodeoxyribonucleotide; ORF, open reading frame; PAF, *Pc* antifungal protein; *paf*, gene encoding PAF; PAGE, polyacrylamide-gel electrophoresis; *Pc*, *Penicillium chrysogenum*; PCR, polymerase chain reaction; RACE, rapid amplification of cDNA ends; re-, recombinant; RT, reverse transcription(ase); *tsp*, transcription start point(s); *UTR*, untranslated region(s).

SSDI 0378-1119(95)00701-6

the secretion potential of *Penicillium chrysogenum* (*Pc*), we have previously investigated some genes encoding extracellular enzymes, e.g., a phoshate-repressible acid phosphatase and a xylanase-encoding gene (Haas et al., 1992; 1993).

As a part of this study we examined the protein secretion profile of *Pc* and found that a 12-kDa protein (PAF) is one of the most abundant species present in the culture medium. Using aa sequence information of the purified protein and corresponding degenerate oligo probes we have isolated in a first step the encoding cDNA. Sequence analysis revealed that this gene shows significant homology to the *Aspergillus giganteus* (*Ag*) *afp* gene which encodes a protein with antifungal activity (Wnendt et al., 1994). Since the PAF-encoding gene might be a suitable target gene to establish a *Penicillium*-specific gene expression system for production and secretion of heterologous re-proteins, we analyzed this gene in detail.

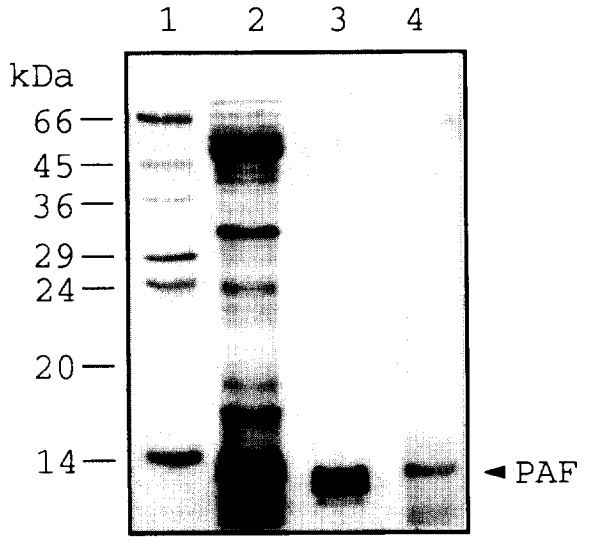

Fig. 1. 0.1% SDS-12% PAGE of fractions obtained during purification of PAF from *Pc*. Lanes: 1, molecular mass protein standards; 2, concentrated (1:20) culture medium; 3, flow through from Mono-Q-chromatography; 4, antifungal protein purified by reverse-phase HPLC. Proteins were stained with Coomassie brilliant blue. **Protein purification methods:** Inoculum cultures (approx. 10^9 spores) of the filamentous fungus *Pc*, strain Q176, were grown at 25°C for 48 h in 250 ml medium containing per litre: 3 g $NaNO_3$/0.5 g KCl/0.5 g $MgSO_4 \cdot 7H_2O$/0.1 g $FeSO_4 \cdot 7H_2O$ and 2% sucrose in 25 mM phosphate buffer (pH 5.8). The culture medium was harvested after 48 h by filtration and concentrated 1:20 by Millipore 10.000 concentrators (Millipore). After dialysis against 15 mM Tris·HCl (pH 7.6)/25 mM NaCl/0.15 mM EDTA it was loaded on a Mono-Q-HR 5/5 column (equilibrated with the same buffer). The protein was finally purified by reverse-phase HPLC chromatography or alternatively by chromatography on Mono-S-HR 5/5.

EXPERIMENTAL AND DISCUSSION

(a) Protein purification and N-terminal aa sequence analysis

Pc produces a number of secreted proteins when cultivated in a medium containing sucrose as sole carbon source (Fig. 1). One of these proteins with a molecular mass of approx. 12 kDa represents up to 30% of the extracellular protein fraction, as judged by SDS-PAGE (Fig. 1, lane 2) and densitometric scanning of the gel. This protein is absent when cells are grown in the same medium but containing 2% glucose instead of sucrose as sole carbon source (data not shown).

In order to investigate this abundantly secreted protein and to obtain aa sequence information for subsequent isolation of the encoding gene we have purified the protein by combination of ion-exchange chromatography and reverse-phase HPLC (Fig. 1, lanes 3, 4). The purified protein (Fig. 1, lane 4) was subjected to N-terminal aa sequence analysis in a gas/liquid phase sequenator providing identification of 21 N-terminal aa. The aa sequence of the N-terminus of the protein was as follows: AlaLysT yrThrGlyLysCysThrLysSerLy sAsnGluCysLys-TyrLysAsnAspAlaGly. A comparison of this partial aa sequence with data contained in several protein databases revealed significant homology (12 identical aa out of 21) with the N-terminal aa sequence of an antifungal protein (AFP) produced by *Ag* (Nakaya et al., 1990). To prove whether the *Pc* protein also exhibits antifungal activity, agar diffusion tests were performed using the purified protein and *A. niger* (*An*) as test organism. The *Pc* protein inhibited the growth of *An* (data not shown), thus indicating that this protein might be indeed equivalent to the *Ag* AFP.

(b) Isolation of the PAF-encoding cDNA and genomic clone

The *paf* cDNA was cloned using the RACE protocol according to Frohmann et al. (1988). For cloning of the 3′-part of the *paf* gene, RT of mRNA isolated from cells grown in a medium containing sucrose was performed. Cyclic thermal amplification of cDNA was done by *Taq* polymerase, an adapter primer and the oligo 5′-AARTAYAARAAYGAYGC (R = A or G; Y = C or T), which corresponds to the N-terminal aa sequence LysTyrLysAsnAspAla of the purified protein. The DNA fragment of approximately 450 bp obtained by this procedure was subcloned and sequenced. According to the 5′-sequence of this fragment the specific oligo 5′-TTATTATCCTTGGTGCAC (corresponding to nt 292 and nt 361–377 in Fig. 2B) was used for RT of mRNA. The cDNA was d(A) tailed and amplified by PCR using a $(dT)_{17}$ primer and a second nested oligo, 5′-TTTCCAGCATCGTTCTTG (nt 235–252 in Fig. 2B), as specific primer. The resulting 300-bp PCR product was

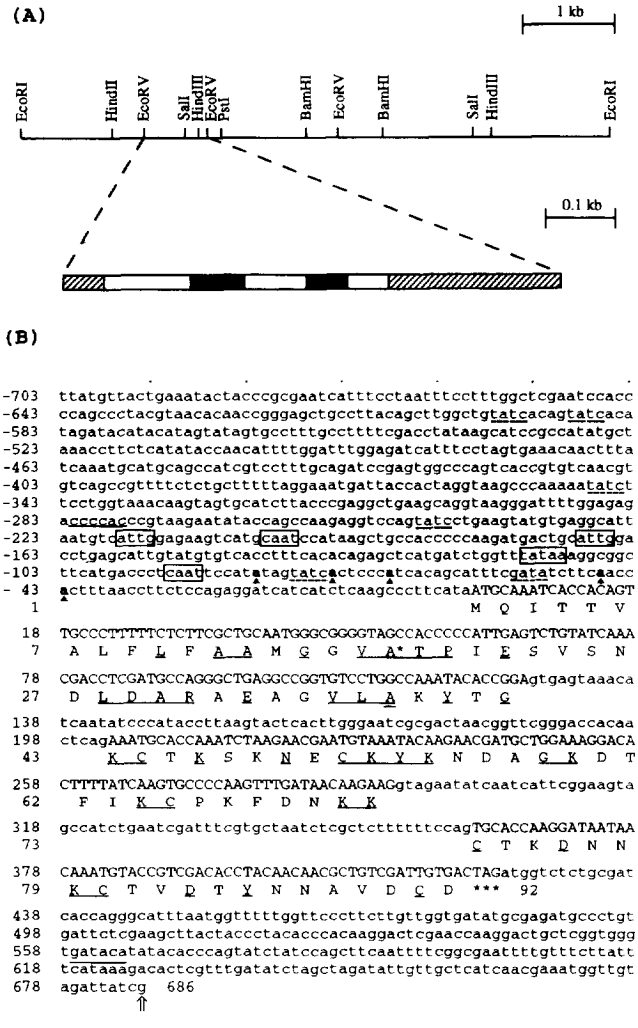

Fig. 2. Structural organization, nt sequence and deduced aa sequence of the *Pc paf* gene (GenBank accession No. U22944). (**A**): The upper part shows the restriction map of an approximately 6-kb *Eco*RI λ phage insert of *Pc* genomic DNA containing the *paf* gene. The lower part shows the intron/exon structure of the gene. Exons are represented by open boxes, while black boxes correspond to introns, 5′- and 3′-noncoding regions are shaded. (**B**): The entire nt sequence of the *paf* gene and deduced aa sequence. The nt numbering begins at the A of the start codon. The aa residues 1–18 comprise a putative signal peptide and the most favoured signal peptide cleavage site is indicated by an asterisk (*). The N-terminus of the mature protein is doubly underlined. The aa identical to aa of AFP from *Ag* are underlined. The five *tsp* are marked in bold letters and black arrowheads. The putative CAAT (or corresponding ATTG) and TATA sequences upstream from *tsp* are boxed. The putative CREA-binding site is underlined. GATA motifs are underlined by broken lines. The stop codon is marked by three asterisks. Noncoding sequences are shown in lower-case letters. The consensus sequence for polyadenylation is overlined. The polyadenylation site is indicated by an upward arrow.

subcloned and one of the re-clones was used for sequence analysis.

In order to isolate a genomic fragment containing the *paf* gene and the adjacent non-coding regions a λ re-phage gene library of *Pc* was screened using a radiolabeled PCR fragment of the cDNA clone. Four different clones which gave a strong hybridization signal were isolated out of 6×10^4 clones. Restriction mapping and hybridization experiments localized the *paf* gene on a 0.7-kb *Eco*RV fragment (Fig. 2A).

(c) Sequence analysis of the *paf* cDNA and the genomic fragment

Translation of the cDNA sequence confirmed that the cloned fragment encodes the purified protein, since the aa Ala^{38}–Gly^{58} perfectly matched those derived from direct sequencing of the N-terminus of the protein (Fig. 2B). The aa sequence deduced from the *paf* cDNA was used to search in several databases with the BLAST alignment computer program (Altschul et al., 1990). It exhibits 42.6% identical aa with the protein sequence of AFP from *Ag* (Wnendt et al., 1994). No significant sequence homology of PAF could be detected with a group of small, very basic and Cys-rich antifungal proteins from plants (Terras et al., 1992; 1993). It should be noted, that the aa sequence motif Asp^{76}–Cys^{91} is highly homologous (11 identical aa out of 16) to a motif which is present in the *Dictyostelium* prestalk morphogen DIF in several copies (Ceccarelli et al., 1987).

The *paf* cDNA encodes a protein precursor of 92 aa. The sequence of the first 18 aa shows characteristic properties of eukaryotic signal sequences (von Heijne, 1986). A signal sequence cleavage site prediction according to von Heijne (1986) strongly suggests that cleavage occurs after Ala^{18}. From comparison of the deduced aa sequence of the cDNA and the N-terminus determined by direct protein sequencing it is evident that following this typical signal peptide a prosequence of 19 aa is present, which is missing in the mature protein. A similar prosequence is also present in the AFP of *Ag* (Wnendt et al., 1994), in the killer toxins of *Saccharomyces cerevisiae* and *Ustilago maydis* (Bostian et al., 1984; Tao et al., 1990), and in several extracellular proteases from *An* (Jarai et al., 1994). It was speculated that this part of the prosequence might have a function in maintaining the precursor in an inactive form.

The 1389 bp of the genomic fragment of the *paf* gene is shown in Fig. 2. In addition to the 279-bp coding sequence, 703 bp of the *5′-UTR* as well as 263 bp of the *3′-UTR* have been sequenced. The *5′-UTR* shows characteristic features of a lower eukaryotic promoter including a perfect TATA box at nt −115 relative to the A of the putative start codon and a consensus region around the start codon (Kozak, 1987). Four copies of the CAAT sequence motif are present at nt −92, −169, −202 and −217. A nt sequence resembling the CREA protein-binding site consensus sequence, SYGGGG (S=C or G; Y=C or T), mediating carbon catabolite repression in *A. nidulans* (Kulmburg et al., 1993), is found at nt −282.

Additionally, several GATA motifs can be found in the *5'-UTR*, of which 2 pairs are closely spaced in a head to tail and head to head orientation, respectively (Fig. 2B). Such motifs are supposed to be involved in nitrogen metabolite repression (Marzluf, 1993). However, whether any of these sequence motifs have functional significance has to be proved by additional studies.

Comparison of the nt sequence of the cDNA and the genomic nt sequence revealed that the *Pc paf* gene consists of three exons and two introns with a length of 76 and 68 bp, respectively. In common with other fungal genes these introns are small and contain the consensus 5'-splice donor site (GT), the consensus 3'-splice acceptor site (AG) and the internal putative lariat formation element (RCTRAC). The size and position of the two *paf* introns are very similar to those found in *Ag afp* (Wnendt et al., 1994).

A sequence resembling the eukaryotic polyadenylation signal can be found at nt 619–624.

(d) Transcript mapping and regulation of *paf* expression

For determination of the *tsp*, in vivo transcripts from *Pc* were characterized by primer extension. As observed in most genes of filamentous fungi, multiple *tsp* were found (Gurr et al., 1987). The major *tsp* of the *paf* gene were determined at positions −43, −47, −69, −75, −83 nt upstream from the start codon (Fig. 2B).

Investigation of proteins present in the medium of *Pc* cells grown with glucose or sucrose as sole carbon source indicated that synthesis of PAF depends on the type of carbon source present in the growth medium. To test whether the expression of *paf* is regulated at the level of transcription and is mediated by carbon catabolite repression, RNA from cells grown with different carbon sources was isolated.

When Northern blots of these RNA samples were hybridized with a *paf* cDNA probe, a single band of approx. 0.7 kb was observed with RNA from cells grown in a medium supplemented with starch or sucrose, which was absent in cells grown with glucose (Fig. 3A). These data clearly demonstrate that *paf* expression is regulated at the level of transcription and indicate that it is controlled by carbon catabolite repression. However, *paf* expression is not exclusively regulated by carbon catabolite repression. From Fig. 3B it is evident that nitrogen metabolite repression, another global regulatory mechanism might also be an important factor in regulation of *paf* expression. Northern blot analysis of RNA isolated from cells grown with different nitrogen sources demonstrated that cells grown in medium supplemented with glutamine had only small amounts of *paf* transcripts, whereas cells grown with the less-favoured nitrogen source nitrate showed a much higher level of *paf* mRNA

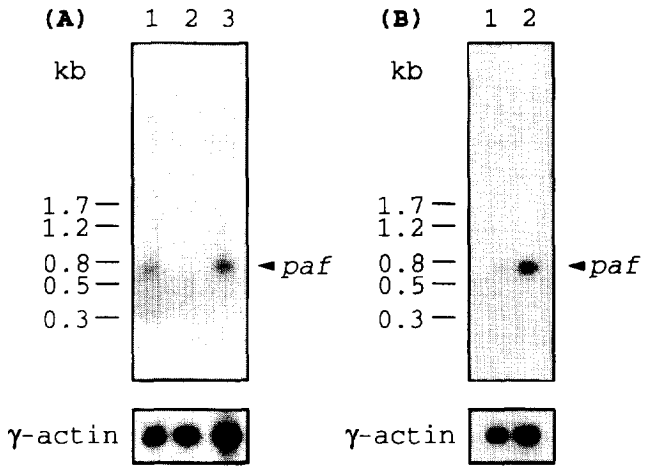

Fig. 3. Regulation of the *paf* gene expression. (**A**) Carbon catabolite repression of *paf* gene expression. RNA was isolated from cells grown on 2% starch (lane 1), 2% glucose (lane 2) or 2% sucrose (lane 3) as sole carbon source. (**B**) Nitrogen catabolite repression of *paf* gene expression. For RNA isolation cells were grown in medium supplemented with 35 mM glutamine (lane 1) or 35 mM $NaNO_3$ (lane 2) as sole nitrogen source. Total RNA was electrophoresed on 2.2 M formaldehyde-1.2% agarose gels, blotted onto Hybond N membranes (Amersham) and hybridized with a radiolabeled PCR fragment of *paf* cDNA. The Northern blots were stripped and rehybridized with a γ-actin probe from *A. nidulans* (Fidel et al., 1988). An RNA ladder (GIBCO/Life Technologies) was used as a size marker.

(Fig. 3B, lanes 1, 2). Since the *paf* promoter region contains several GATA motifs it might be suggested that the major nitrogen regulatory protein (NRE) of *Pc* (Haas et al., 1995), which was shown to bind to GATA motifs (Haas and Marzluf, 1995), is involved in the nitrogen control of *paf* gene expression.

(e) Conclusions

(*1*) An abundantly secreted 12-kDa protein produced by *Pc* cells, grown in a medium with sucrose as the sole carbon source, was purified and subjected to N-terminal aa sequence analysis. Subsequent isolation and sequencing of the encoding cDNA revealed that this protein is homologous to an antifungal protein of *Ag*. With exception of general physical characteristics – e.g. small size, highly basic and Cys-rich – these fungal proteins do not show significant aa sequence homology to antifungal proteins from plants and may thus represent a new family of antifungal proteins.

(*2*) The high amount of PAF produced, and the mode of regulation of expression, indicates that the *paf* gene might be a suitable target gene to establish a *Pc* specific secretion system for heterologous gene expression. The isolation and analysis of the entire *paf* gene including the regulatory regions will enable us to prove this feasibility.

(*3*) Expression of *paf* is repressed by the presence of a favoured carbon or nitrogen source in the growth medium, thus indicating that the gene is under control

of major wide-domain carbon and nitrogen regulatory genes.

ACKNOWLEDGEMENTS

B. R. was supported by the Dr. Legerlotz Foundation. We would like to thank Dr. P. Loidl for his strong encouragement and support of this project.

REFERENCES

Altschul, S.F., Gish, W., Miller, W., Myers, E.M. and Lipman, D.J.: Basic local alignment search tool. J. Mol. Biol. 215 (1990) 403–410.

Bostian, K.A., Eliott, Q., Bussey, H., Burn, V., Smith, A. and Tipper, D.J.: Sequence of the preprotoxin dsRNA gene of type-I killer yeast: multiple processing events produce a two-component toxin. Cell 36 (1984) 741–751.

Ceccarelli, A., McRobbie, S.J., Jermyn, K.A., Duffy, K., Early, A. and Williams, J.G.: Structural and functional characterization of a *Dictyostelium* gene encoding a DIF inducible, prestalk-enriched mRNA sequence. Nucleic Acids Res. 15 (1987) 7463–7476.

Fidel, S., Doonan, J.H. and Morris, N.R.: *Aspergillus nidulans* contains a single γ-actin gene which has unique intron locations and encodes a γ-actin. Gene 70 (1988) 283–293.

Frohmann, M.A., Dush, M.K. and Martin, G.R.: Rapid production of full-length cDNAs from rare transcripts: amplification using a single gene-specific oligonucleotide primer. Proc. Natl. Acad. Sci. USA 85 (1988) 8998–9002.

Gurr, S.J., Uncles, S.E. and Kinghorn, J.R.: The structure and organization of nuclear genes of filamentous fungi. In: Kinghorn, J.R. (Ed.), Gene Structure in Eukaryotic Microbes. IRL Press, Oxford, 1987, pp. 93–193.

Haas, H., Redl, B., Friedlin, E. and Stöffler, G.: Isolation and analysis of the *Penicillium chrysogenum phoA* gene encoding a secreted phosphate-repressible acid phosphatase. Gene 113 (1992) 129–133.

Haas, H., Friedlin, E., Stöffler, G. and Redl, B.: Cloning and structural organization of a xylanase-encoding gene from *Penicillium chrysogenum*. Gene 126 (1993) 237–242.

Haas, H., Bauer, B., Redl, B., Stöffler, G. and Marzluf, G.A.: Molecular cloning of *nre*, the major nitrogen regulatory gene of *Penicillium chrysogenum*. Curr. Genet. 27 (1995) 150–158.

Haas, H. and Marzluf, G.A.: NRE, the major nitrogen regulatory protein of *Penicillium chrysogenum* interacts with the promoter regions of nitrate assimilation and penicillin biosynthetic gene cluster. Curr. Genet. 28 (1995) 177–183.

Jarai, G., Van den Hombergh, H. and Buxton, F.P.: Cloning and characterization of the *pepE* gene of *Aspergillus niger* encoding a new aspartic protease and regulation of *pepE* and *pepC*. Gene 145 (1994) 171–178.

Jeenes, D.J., MacKenzie, D.A., Roberts, I.N. and Archer, D.B.: Heterologous protein production by filamentous fungi. Biotechnol. Genet. Eng. Rev. 9 (1991) 327–367.

Kozak, M.: An analysis of 5′-noncoding sequences from 699 vertebrate mRNAs. Nucleic Acids Res. 15 (1987) 8125–8148.

Kulmburg, P., Mathieu, M., Dowzer, C., Kelly, J. and Felenbok, B.: Specific binding sites in the *alcR* and *alcA* promoters of the ethanol regulon for the CREA repressor mediating carbon catabolite repression in *Aspergillus nidulans*. Mol. Microbiol. 7 (1993) 847–857.

Marzluf, G.A.: Regulation of sulfur and nitrogen metabolism in filamentous fungi. Annu. Rev. Microbiol. 47 (1993) 31–55.

Nakaya, K., Omata, K., Okahashi, I., Nakamura, Y., Kolkenbrick, H. and Ulbrich, N.: Amino acid sequence and disulfide bridges of an antifungal protein isolated from *Aspergillus giganteus*. Eur. J. Biochem. 193 (1990) 31–38.

Tao, J., Ginsberg, I., Banerjee, N., Held, W., Koltin, Y. and Bruenn, J.: *Ustilago maydis* KP6 killer toxin: structure, expression in *Saccharomyces cerevisiae*, and relationship to other cellular proteins. Mol. Cell. Biol. 10 (1990) 1373–1381.

Terras, F.R.G., Schoofs, H.M.E., De Bolle, M.F.C., Van Leuven, F., Rees, S.B., Vanderleyden, J., Cammue, B.P.A. and Broekaert, W.F.: Analysis of two novel classes of plant antifungal proteins from radish (*Raphanus sativus* L.) seeds. J. Biol. Chem. 267 (1992) 15301–15309.

Terras, F.R.G., Torrekens, S., Van Leuven, F., Osborn, R.W., Vanderleyden, J., Cammue, B.P.A. and Broekaert, W.F.: A new family of basic cysteine-rich plant antifungal proteins from *Brassicaceae* species. FEBS Lett. 316 (1993) 233–240.

Van den Hondel, C.A.M.J.J., Punt, P.J. and Van Gorcom, R.F.M.: Heterologous gene expression in filamentous fungi. In: Bennett, J.W. and Lasure, L.L. (Eds.), More Gene Manipulations in Fungi. Academic Press, London, 1991, pp. 396–428.

Von Heijne, G.: A new method for predicting signal sequence cleavage sites. Nucleic Acids Res. 14 (1986) 4683–4690.

Wnendt, S., Ulbrich, N. and Stahl, U.: Molecular cloning, sequence analysis and expression of the gene encoding an antifungal-protein from *Aspergillus giganteus*. Curr. Genet. 25 (1994) 519–523.