

The *nagA* gene of *Penicillium chrysogenum* encoding β -*N*-acetylglucosaminidase

Bruno Díez¹, Marta Rodríguez-Sáiz, Juan Luis de la Fuente, Miguel Ángel Moreno²,
José Luis Barredo^{*}

R&D Biology, Antibióticos S.A., Avenida de Antibióticos 59-61, 24009 León, Spain

Received 3 September 2004; received in revised form 2 November 2004; accepted 5 November 2004

First published online 19 November 2004

Edited by N. Gunde-Cimerman

Abstract

We purified the β -*N*-acetylglucosaminidase from the filamentous fungus *Penicillium chrysogenum* and its N-terminal sequence was determined, showing the presence of a mixture of two proteins (P1 and P2). A genomic DNA fragment was cloned by using degenerated oligonucleotides from the Nt sequences. The nucleotide sequence showed the presence of an ORF (*nagA* gene) lacking introns, with a length of 1791 bp, and coding for a protein of 66.5 kDa showing similarity to acetylglucosaminidases. The NagA deduced protein includes P1 and P2 as incomplete forms of the mature protein, and contains putative features for protein maturation: an 18-amino acid signal peptide, a KEX2 processing site, and four glycosylation motifs. The sequence just after the signal peptide corresponds to P2 and that after the KEX2 site to P1. The *nagA* transcript has a size of about 2.1 kb and is present until the end of the fermentation process for penicillin production. NagA is one of the most largely represented proteins in *P. chrysogenum*, increasing along the fermentation process. The suitability of the *nagA* promoter (*PnagA*) for gene expression in fungi was demonstrated by expressing the bleomycin resistance gene (*ble^R*) from *Streptoalloteichus hindustanus* in *P. chrysogenum*.

© 2004 Federation of European Microbiological Societies. Published by Elsevier B.V. All rights reserved.

Keywords: Glucosaminidase; Hexosaminidase; Acetylglucosaminidase; Acetylhexosaminidase; Chitin; Penicillin

1. Introduction

The main structural component of the cellular wall of most fungi is chitin, a β -1,4-linked homopolymer constituted by *N*-acetylglucosamine (GlcNAc) residues. Chitin accumulation in the genus *Penicillium* is influenced by

the composition and quantity of carbon and nitrogen sources, acidity of the medium, aeration rate, and age of the culture [1]. The carbohydrates of *Penicillium chrysogenum* mycelial cell wall (glucans and chitin) and those serving as storage materials (glycogen) decrease with the specific growth rate [2], whereas the contents of chitin remained constant during ageing in *P. expansum* [3]. The total content of neutral carbohydrates in the cell wall of *P. charlesii* is 69%, being glucosamine (12%) principally present as chitin [4].

The chitin degradation pathway present in many microorganisms involves two successive steps: exo- and endo-chitinases convert chitin into (GlcNAc)_n, which is then hydrolyzed to GlcNAc by β -*N*-acetylglucosami-

^{*} Corresponding author. Tel.: +34 987 895 826; fax: +34 987 895 986.

E-mail address: jbarredo@antibioticos.it (J.L. Barredo).

¹ Present address: Bioferma Murcia S.A., Poligono Industrial Las Salinas, Avda. de Europa s/n, 30840 Alhama de Murcia, Murcia, Spain.

² Present address: Puleva Biotech S.A., Camino de Purchil, 66, 18004, Granada, Spain.

dase (EC 3.2.1.52 according to IUBMB Enzyme Nomenclature). This enzyme is widely distributed throughout nature and catalyzes the removal of terminal β -1,4-linked to *N*-acetylglucosamine residues from the non-reducing ends of oligosaccharides and their conjugates. A β -*N*-acetylglucosaminidase which hydrolyzes chitin oligomers to GlcNAc was purified from *Aeromonas* sp. [5] and during autolysis from *Aspergillus nidulans* [6]. A β -*N*-acetylhexosaminidase released into the culture medium during autolysis of *P. chrysogenum* has been physiologically and enzymologically characterized as a homodimer quaternary structure of 66.4 kDa, which is formed under carbon catabolite repression and is not induced by either chitin or GlcNAc [7,8]. The β -*N*-acetylhexosaminidase is released into the broth without modification during the hyphal fragmentation phase in *P. chrysogenum* cultures [7].

The genes encoding β -*N*-acetylglucosaminidase from the fungi *A. nidulans* (*nagA*) [9], *Candida albicans* (*HEX1*) [10], and *Trichoderma atroviride* (= *Trichoderma harzianum* P1; *nag1*) [11] have been cloned and characterized. *HEX1* gene play no direct role in the dimorphic transition of *C. albicans* and its transcription appears to be temperature regulated [12]. The promoter region of *nag1* binds regulatory proteins to three definite motifs and directed mutagenesis of some motifs resulted in loss of in vivo inducibility by GlcNAc [13].

Fragmentation and autolysis of *P. chrysogenum* have been related with the production of high levels of *N*-acetyl- β -D-hexosaminidase [14]. Nitrogen and oxygen limitation promote extensive culture autolysis in submerged batch cultures of *P. chrysogenum*, becoming autolytic regions within some hyphae much more widespread as the process proceeded [15]. The empty (autolyzed and inactive) regions within the mycelia of an industrial strain of *P. chrysogenum* from a batch fermentation constitute up to 15% of the biomass during the growth phase and more than 50% by the end of fermentation. The increase in the percentage of empty regions coincides with a decrease in biomass and a fall in penicillin titre [16]. We investigated the pattern of proteins expressed during *P. chrysogenum* fermentation in order to correlate with mycelium autolysis, identifying β -*N*-acetylglucosaminidase as one of the more represented proteins. Our aim was to clone and characterize the *nagA* gene of *P. chrysogenum* as a tool for the construction of fungal expression vectors based on its promoter (*PnagA*).

2. Materials and methods

2.1. Microorganisms, phages and plasmids

P. chrysogenum penicillin overproducing strains A and E1 [17] from Antibióticos S.A. series were used to check the patterns of intracellular proteins. E1 strain

was used as a source of DNA. *Escherichia coli* SURE (Stratagene) and *E. coli* LE392 [18] were the hosts for λ GEM12 (Promega) phage derivatives. Phage λ GEM12 was used to construct a genomic library of *P. chrysogenum* E1. *E. coli* DH5 α [18] was the recipient for high-frequency plasmid transformation. pBluescript I KS (+) and pBC KS (+) phagemids (Stratagene) were used for subcloning and sequencing. General procedures for plasmid DNA purification, cloning and transformation of *E. coli* were those of Sambrook et al. [18].

2.2. Culture, β -*N*-acetylglucosaminidase purification and Nt sequencing

P. chrysogenum was cultured by flask fermentation as previously described [19]. Cell-free extracts were prepared from 180 h old mycelium grown in fermenter in penicillin producing conditions. Cells from 200 ml of culture broth were collected by filtration, washed twice with sterile saline solution, and suspended in 100 ml of 500 mM potassium phosphate buffer pH 8.2 containing 4 mM EDTA and 1 mM PMSF (phenylmethanesulfonyl fluoride). Mycelium was disrupted with 100 g of glass-bead (Ballotini, 0.25–0.5 mm diameter) using a Dynomillglass-bead mill apparatus (Willy A. Bachofen), and the crude extract was cleared by centrifugation at 17,000g for 20 min at 4 °C. β -*N*-acetylglucosaminidase activity was assayed as described [20]. Protein concentration was determined using bovine serum albumin as standard [21]. β -*N*-acetylglucosaminidase was purified essentially according to described procedures [22]. Briefly, the cleared extract was precipitated with ammonium sulphate and the active fraction (35–65% saturation) was sequentially subjected to the following purification steps: DEAE-Sephacel, hydroxylapatite, and concentration (Amicon PM10). Samples from the different purification steps were analysed by 10% denaturing SDS–polyacrylamide gel electrophoresis (SDS–PAGE) using the broad range molecular mass standards (Bio-Rad). Isoelectric focusing gels (IEF) were carried out with the Protean 111 mini IEF cell system (Bio-Rad). The Nt edge of the protein was sequenced by Eurosequence bv (Groningen, The Netherlands), determining two different amino acid sequences which were used to design three degenerated oligonucleotides: A (5'-TCGACGACGTGSACGTCSACGTTGTGGATGCC-3'), B (5'-CCGTAYTGSAGGTCRGCCTCGTTGTGACGAC-3'), and C (5'-GGGGCVGGGAGVG-GGTTGACYTG-3').

2.3. Construction and screening of a *P. chrysogenum* genomic library

The DNA of *P. chrysogenum* E1 was purified as previously described [23]. Fragments of genomic DNA (17–22 kb) were selected from *Sau*3AI partially digested

DNA, ligated to λ GEM12, and packaged with “Giga-pack II Gold” kit (Stratagene) to yield $\sim 5 \times 10^4$ pfus. For screening, the library was amplified in *E. coli* LE392, plated to obtain $\sim 4 \times 10^4$ pfus, and hybridized with the above described ^{32}P -ATP labelled oligonucleotides according to standard methods [18]. Recombinant phages were amplified in liquid medium to purify their DNA [18].

2.4. Nucleic acid hybridization and sequencing

Southern and Northern hybridizations were made with ^{32}P -dCTP labelled probes according standard procedures [18]. A 2.1 kb *Sa*I fragment including *PnagA* was used as a probe to analyze pALfleo10 transformants. Mycelium after 100 and 172 h from pilot-scale fed-batch fermentation for penicillin production was frozen in liquid nitrogen, and total RNA was recovered, resolved by agarose-formaldehyde gel electrophoresis, and hybridized according to standard procedures [18]. DNA fragments from *nagA* (980 bp *Bam*HI) and *pcbC* (937 bp *Nco*I) [24] genes of *P. chrysogenum* were used as probes for Northern analysis. Clones to be sequenced were constructed with the “Erase a base” system (Promega) and sequenced with an ABI PRISM 377 DNA sequencer (Applied Biosystems). DNA sequences were analyzed with the Dnastar and Winstar packages (DNASTAR Inc.), and compared to known sequences by using the BLASTP and BLASTX programs [25]. The codon-preference algorithm (Geneplot, Winstar) was used to detect the presence of ORFs. Multiple protein alignments were made with the Clustal V algorithm and pair wise alignments with the Lipman–Pearson method (MegAlign, Winstar). RNA secondary structures that can act as potential transcriptional terminators were predicted with the aid of mfold program [26]. The *nagA* DNA sequence of *P. chrysogenum* encoding β -N-acetylglucosaminidase has been deposited as GenBank Accession No. AF056977.

2.5. *ble*^R gene expression using *PnagA*

Plasmid pALfleo10 was constructed to express the bacterial phleomycin resistance gene (*ble*^R) from *Streptoalloteichus hindustanus* under the control of *PnagA*. To create an *Nco*I restriction site in the ATG translation start codon of *nagA*, the plasmid pALP307 (pBluescript I KS (+) with an insert of 892 bp including *PnagA*) was used as a template for PCR reactions with primers #8 (5'-CTCCATGGTGATAAGGTGAGTGACGATG-3'; *Nco*I site underlined) and -20 (5'-GTAAACGACGCCAGTG-3'). Optimized amplification reactions (100 μ l) contained about 20 ng of pALP307, 10 mM Tris–HCl pH 8.3, 1.5 mM MgCl₂, 50 mM KCl, 20 pmol of each primer, 1.25 U of Taq DNA polymerase and dNTPs 200 mM each. A Perkin–Elmer DNA Thermal

Cycler 480 was used for amplification with a program of 15 cycles (94 °C, 60 s (2 min for the first cycle); 53 °C, 30 s; 72 °C, 90 s (extending 3 s per cycle)). The amplified product was purified from agarose gel by “Qiaex II gel extraction kit” (Qiagen) and cloned into the *Eco*RV site of pBC KS (+). pALP427 was selected after sequencing several clones to verify the presence of the *Nco*I site and the absence of punctual mutations. Unidirectional deletions of pALP427 were done to remove an *Nco*I site located upstream *PnagA*, constructing pALP448 which includes *PnagA* as a 575 bp fragment. Both *ble*^R and the transcriptional terminator signal from the *TrpC* gene of *A. nidulans* (*TrpC*) were obtained as a 1.1 kb *Nco*I–*Apa*I fragment from pUT737 [27], and cloned into pALP448 double digested *Nco*I–*Apa*I. The resulting plasmid pALfleo10 includes the cassette *PnagA*–*ble*^R–*TrpC* conferring resistance to phleomycin–bleomycin in fungi and the *Cm*^R gene providing chloramphenicol resistance in *E. coli*. DNA sequence showed the correct in frame fusion *PnagA*–*ble*^R. *P. chrysogenum* transformants were obtained as previously described [28], using Czapek medium supplemented with 1 M sorbitol for protoplasts stability and with 30 $\mu\text{g ml}^{-1}$ phleomycin for selection.

3. Results

3.1. Protein purification, Nt edge sequence and protein identification

When analysing the patterns of intracellular proteins of *P. chrysogenum* penicillin overproducing strains A and E1 during their fermentation for penicillin production, one of the most largely represented proteins (64 kDa) was afterwards demonstrated to be β -N-acetylglucosaminidase (see below). A large amount of β -N-acetylglucosaminidase has an intracellular location increasing along the fermentation process (Fig. 1).

β -N-acetylglucosaminidase was purified from *P. chrysogenum* mycelium grown in penicillin production conditions showing the presence of a mixture of two proteins in a ratio of about 1:1, matching in size with the most represented protein in *P. chrysogenum* E1 extracts (64 kDa; Fig. 1). It was not possible to jointly sequence both Nt edges by standard methods, but since one of the proteins (P1) had a proline at position 2, it was sequenced making use of a blocking reagent (OPA) to avoid the sequence of the protein chain lacking a proline residue (P2). By exclusion, some preliminary sequence information of the Nt edge of both proteins was obtained: 22 residues from P1 ((Ala/Val)–Pro–Ser–Gly–Ile–His–Asn–Val–Asp–Val–His–Val–Val–Asp–Asn–Asp–Ala–Asp?–Leu–Gln–Tyr–Gly?), and 20 residues from P2 ((Ala/Val)–Gln–Val–Asn–Pro–Leu–Pro–Ala–Pro–Arg?–Arg?–Ile–Thr–???–Gly–Ser?–Ser?–Gly?–Pro?–(Ile/Thr)–

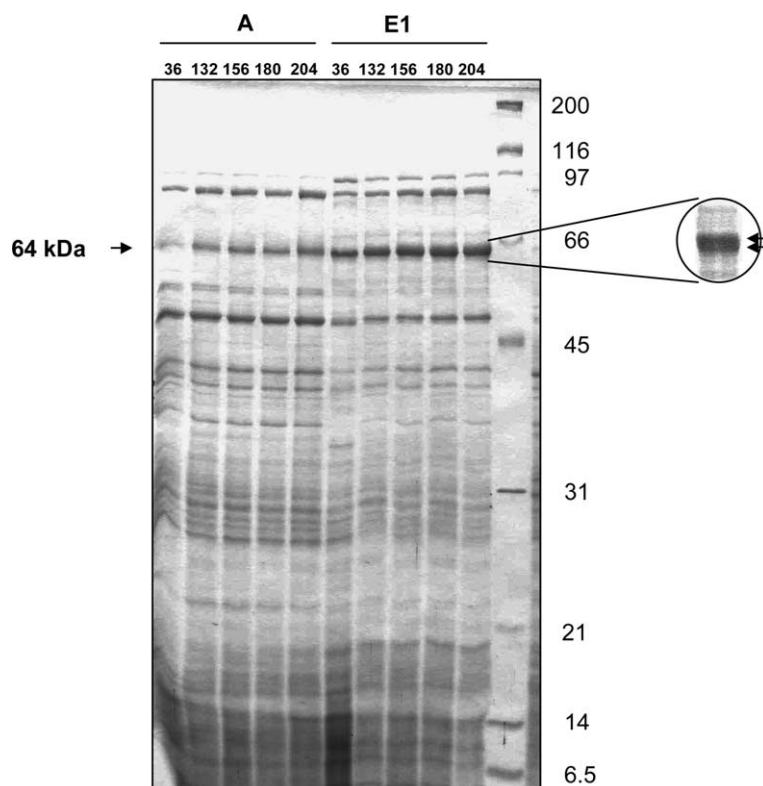


Fig. 1. SDS-PAGE analysis of intracellular proteins from *P. chrysogenum* showing the pattern of strains A and E1 along pilot-scale fermentation for penicillin production from 36 to 204 h. β -*N*-acetylglucosaminidase is indicated by arrows. The purified sample having a mixture of two proteins is shown on the right inside a circle. Protein broad range molecular mass standards (Bio-Rad) were used.

???–Val?). Residues with a question mark were estimated with around a 60% of confidence and residues with three question marks were not determined.

The purified sample was analyzed by 2D electrophoresis (IEF-PAGE), showing a narrow pI range of heterogeneity. To confirm this result slab IEF gels were done by adjusting the pH in the range from 5 to 7. In this case a mixture of three very close spots with a pI ranging from 5.1 to 5.3 was detected: (1) spot located to the more acidic zone, (2) spot located to the more basic zone, and (3) central spot. The N-terminal sequence of the spot corresponding to the more represented protein (3) was Ala–Pro–Ser–Gly–Ile–His–Asn–Val. This sequence was identical to the previously determined for P1, confirming Ala and Val as the first residues for P1 and P2, respectively. P1 and P2 N-terminal sequences were used for searching the databases to obtain information about their putative functions, showing both of them significant similarity to the β -*N*-acetylglucosaminidases from *A. nidulans* (BAB13330) and *Aspergillus oryzae* (BAC41255).

3.2. Cloning and characterization of the *nagA* gene from *P. chrysogenum*

A genomic library from *P. chrysogenum* was screened with three degenerated oligonucleotides, two of them corresponding to P1 (A and B) and one from P2 (C). Po-

sitive phages were isolated with oligonucleotide A, but oligonucleotides B and C did not show clear positive signals. Eleven phages were mapped, all showing the hybridization signal in a 3.2 kb *SacI* fragment and in a 2.1 kb *SalI* fragment, which were cloned into pBlue-script I KS (+) to give pALP344 and pALP295, respectively. Additionally, a 4.5 kb *SacI* fragment adjacent to the 3.2 kb *SacI* fragment was cloned into pBlue-script I KS (+) to give pALP346. A 5240 bp DNA sequence (GenBank [AF056977](#)) was determined and analyzed with the codon-preference algorithm showing a 1791 bp ORF lacking introns and encoding a protein with high similarity to β -*N*-acetylglucosaminidases (Table 1). The deduced polypeptide has a predicted molecular mass of 66,538 Da and a pI of 5.35. Both N-terminal sequences obtained for P1 (Ala–Pro–Ser–Gly–Ile–His–Asn–Val–Asp–Val–His–Val–Val–Asp–Asn–Asp–Ala–Asp–Leu–Gln–Tyr–Gly; amino acids from 99 to 120) and P2 (Val–Gln–Val–Asn–Pro–Leu–Pro–Ala–Pro–Arg–Asn–Ile–Thr–Trp–Gly–Ser–Ser–Gly–Pro–Ile–Gln–Val; amino acids from 19 to 40) were found in the deduced sequence, pointing to the proteolytic cleavage of the 64,792 Da (578 amino acids; pI 5.26) protein to give a 56,184 Da (498 amino acids; pI 5.01) protein (Fig. 2). Amino acids from 1 to 18 in the deduced protein were not found in the Nt sequences, suggesting a 1763 Da leader peptide (Table 1). The sequence corresponding to P2

Table 1

Features of β -N-acetylglucosaminidase encoding genes from *P. chrysogenum* (Pc; AF056977), *A. nidulans* (An; BAB13330), *A. oryzae* (Ao; BAC41255), *C. albicans* (Ca; P430077), *T. atroviride* (Ta; S72211), and *A. thaliana* (At; NP_172050)

	Pc	An	Ao	Ca	Ta	At
Gene length (bp)	1791	1809	1803	1686	1880	1743
Deduced product (aa)	596	603	600	562	580	580
Deduced product (kDa)	66.5	68.0	67.5	63.6	64.5	65.0
Deduced pI	5.3	4.6	4.7	5.3	5.7	6.2
Signal peptide (aa)	18	19	18	22	19	–
Potential N-glycosylation motifs	4	3	6	7	7	4
Similarity to NagA (%)	100	68	66	48	32	37

Similarity was estimated with the program MegAlign (Winstar) by the Lipman–Pearson protein alignment method.

was found just after the signal peptide, whereas P1 was located immediately downstream of a KEX2 (Lys–Arg) [29] processing site (Fig. 2).

General transcription signals were detected at the 5'-flanking region (Fig. 2): pyrimidine rich regions between positions 1106–1128 and 1182–1200, a potential TATA like box (ATAAATA at 1258) located 66 bp upstream from the translation start codon ATG (1324), and two potential CAAT like boxes at positions 1163 and 1244. Additionally, *creA* motifs ((G/C)(C/T)GG(A/G)G) [30] were found at 896 and 1024. RNA secondary structures that can act as potential transcriptional terminators were downstream [26].

The deduced amino acid sequence includes four potential N-glycosylation motifs [NX(T/S)] [31] at 29, 348, 496, and 591. A phylogenetic tree was constructed

showing the evolutive relationship between hexosaminidases from bacterial, fungal and higher eukaryote origins (Fig. 3).

3.3. Transcription analysis of the *nagA* gene

Northern analysis was performed to determine the size and transcription level of *nagA* in *P. chrysogenum* grown in penicillin production conditions. A gene from the penicillin biosynthetic pathway (*pcbC*) known to show strong transcription under these conditions [32] was used to compare. A band of about 2.1 kb, which agrees with a monocistronic transcript efficiently expressed, was detected (Fig. 4). *nagA* transcript level clearly increased from 100 to 172 h, whereas a steadier transcript level of *pcbC* (about 1.4 kb) was detected.

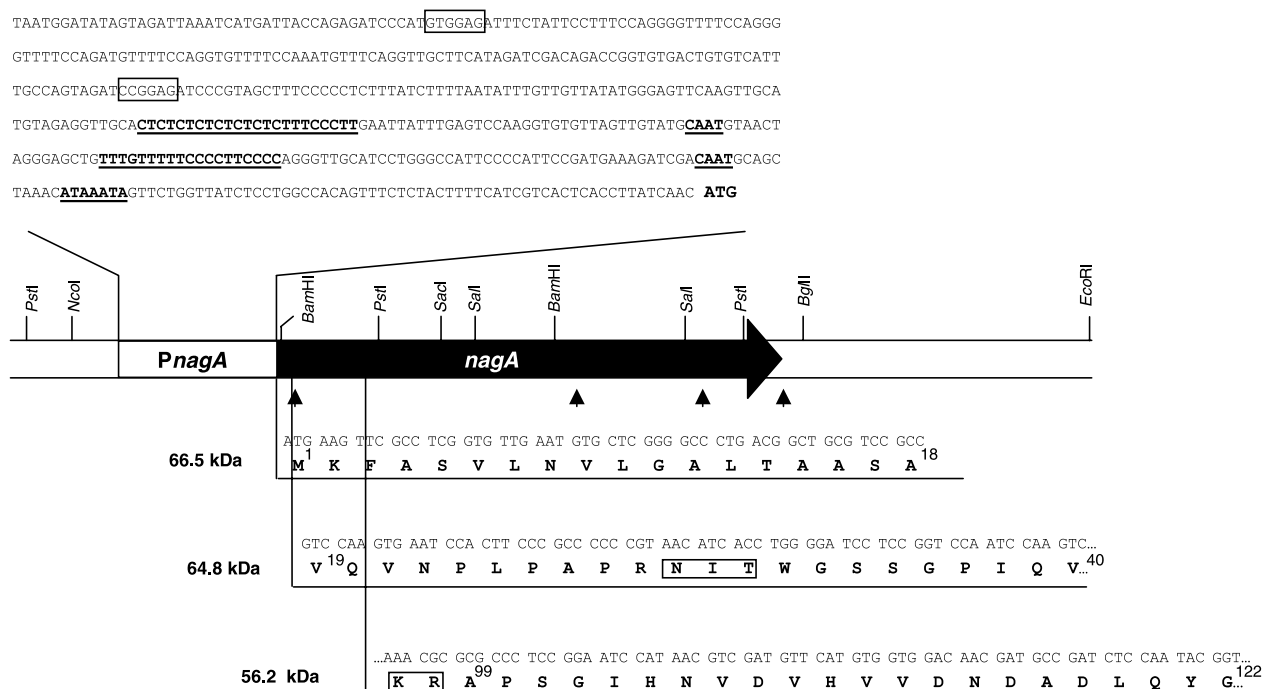


Fig. 2. Restriction map of the genomic DNA region of *P. chrysogenum* including the *nagA* gene. Little arrows show the location of putative glycosylation sites. Promoter region (*PnagA*): pyrimidine rich stretches, TATA box, and CAAT boxes are bold and underlined; *creA* motifs are boxed. Deduced protein: 66.5 kDa including leader peptide, 64.8 kDa lacking leader peptide (glycosylation site boxed), and 56.2 kDa KEX2 processed (KEX2 site boxed).

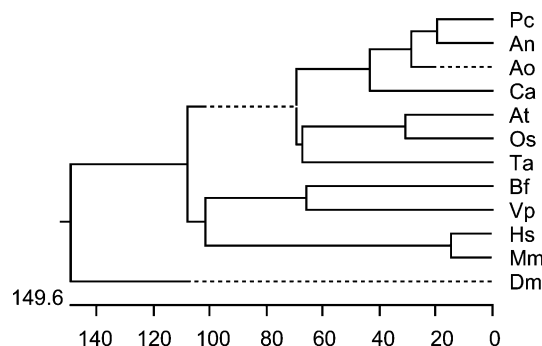


Fig. 3. Phylogenetic tree of hexosaminidases. Sequences from *P. chrysogenum* (Pc), *A. nidulans* (An), *A. oryzae* (Ao), *C. albicans* (Ca), *A. thaliana* (At), *O. sativa* (Os), *T. atroviride* (Ta), *B. fragilis* (Bf), *V. parahaemolyticus* (Vp), *H. sapiens* (Hs), *M. musculus* (Mm), and *D. melanogaster* (Dm) were analyzed by the Clustal method with the PAM250 residue weight table. Multiple alignment parameters were: gap penalty = 10 and gap length penalty = 10. Pairwise alignment parameters were: ktuple = 1, gap penalty = 3, window = 5 and diagonals = 5. The length of each pair of branches represents the distance between sequence pairs. The units at the bottom of the tree indicate the number of substitution events.

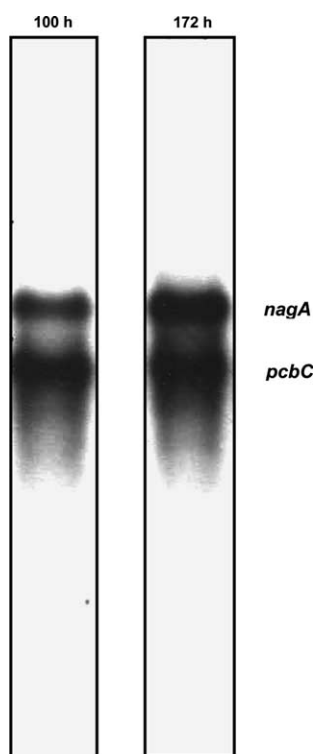


Fig. 4. *nagA* and *pcbC* transcripts after 100 and 172 h from pilot-scale fed-batch fermentation for penicillin production.

3.4. Gene expression under the control of *PnagA* in *P. chrysogenum*

The suitability of *PnagA* to be used for gene expression was tested in *P. chrysogenum* with the heterologous *ble^R* gene. Plasmid pALfleo10 (Fig. 5) contains the

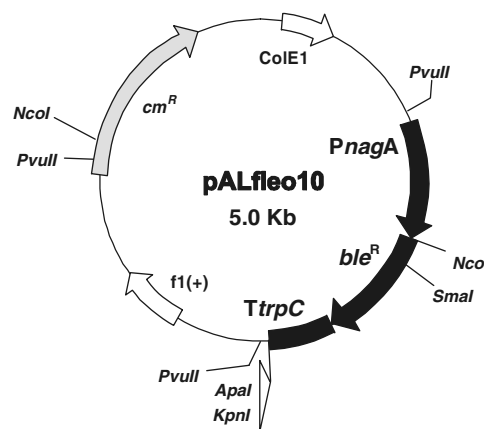


Fig. 5. Restriction map of the plasmid pALfleo10 used for *ble^R* gene expression under *PnagA* control in *P. chrysogenum*.

PnagA-ble^R-TtrpC cassette and the *Cm^R* gene for selection in *E. coli*. pALfleo10 transformants of *P. chrysogenum* were able to grow on 30–80 $\mu\text{g ml}^{-1}$ phleomycin. Phleomycin resistance levels were dependent on the strain tested and the number of functional copies of *PnagA-ble^R-TtrpC* integrated in their chromosomes. Southern analysis of 10 transformants showed hybridization bands of different size and intensity depending on the locus and number of integrations of pALfleo10, but only the hybridization signal corresponding to *nagA* (2.1 kb *SalI* fragment) appeared in the parental untransformed strain (not shown). These results show that *PnagA* is able to direct heterologous *ble^R* gene expression in *P. chrysogenum*.

4. Discussion

We cloned the *nagA* gene from the penicillin overproducing strain *P. chrysogenum* E1. The deduced polypeptide is similar at the amino acid level to β -*N*-acetylglucosaminidases, especially from fungi. The presence of introns was not detected in *nagA*, differing of many genes from *P. chrysogenum* but in agreement to the *nagA* gene of *A. nidulans*. The motifs for regulatory proteins described in the promoter region of the *nagI* gene of *T. atroviride* and involved in its induction [13] were not found in *PnagA*. The pyrimidine rich stretch suggests a putative transcription start site, since it appears in highly expressed genes immediately upstream of this site [33]. Most of the fungal genes have a TATA box located 30–50 bp upstream of the transcription start point [33], which agrees with the supposed functional TATA box located 34 bp upstream to the translation start point. The CCAAT box is present in the promoter region of about 30% of known eukaryotic genes located between 50 and 200 bp upstream of the transcription start site [33]. The location of the potential CCAAT like box is consistent (105 bp upstream).

Both, *nagA* transcript level and NagA protein increased along the fermentation process. This phenomenon is coherent with (i) the existence of fragmentation and autolysis in *P. chrysogenum* fed-batch fermentations to produce penicillin coupled with the production of β -N-acetylhexosaminidase [14], and (ii) the much more extensive lysis as the process proceeds [15,16]. *nagA* transcript level after 172 h is similar to that of *pcbC*, a penicillin biosynthetic gene highly expressed under penicillin production conditions [32]. Maximum *pcbC* transcription level occurs in 2–3 days old flask cultures of *P. chrysogenum* [32], but steady-state levels of mRNA were present until 7 days in a fermenter, probably because the nutrient feeding protocol in the fermenter extends fungal vegetative growth and penicillin production. *nagA* transcript size was similar to the reported for *HEX1* from *C. albicans* [10].

Hexosaminidases from lower eukaryotes (*P. chrysogenum*, *A. nidulans*, *A. oryzae*, *C. albicans*, and *T. atroviride*) are more closely related to the homologous protein from plant sources (*Arabidopsis thaliana* and *Oryza sativa*) than to the corresponding enzyme from bacteria (*Bacteroides fragilis* and *Vibrio parahaemolyticus*), *Drosophila melanogaster* and higher eukaryotes (*Homo sapiens* and *Mus musculus*).

Putative features for protein maturation were found in NagA, including an 18-amino acid signal peptide, a KEX2 processing site, and four glycosylation motifs. As described for *T. atroviride*, where the protein has an increase of Mr by 10 kDa due to linked glycosyl chains [11], P1 (56 kDa) glycosylation pattern could explain its migration near P2 (64 kDa) in SDS–PAGE analysis. Potential N-glycosylation motifs are also present in fungal sequences encoding β -N-acetylglucosaminidase. A significant bias for some basic dipeptides (Lys–Arg, Arg–Arg, and Arg–Lys) as targets of proteolytic cleavage by the KEX2 protease exists in fungi [29], being Lys–Arg the dipeptide found immediately upstream of P1. The KEX2 site has also been found in the sequences from *A. nidulans*, *A. oryzae* and *A. thaliana*.

One of the most largely represented proteins in *P. chrysogenum* fermentations for penicillin production is β -N-acetylglucosaminidase. A β -N-acetylhexosaminidase was described to accumulate inside *P. chrysogenum* during the stationary phase of growth and to release into the culture broth when the cell wall lysis progressed [7]. Nevertheless, the N-terminal sequence of this 66.4 kDa β -N-acetylhexosaminidase (APSNVDNXXDVXNDA) [8] was not found in *nagA* deduced sequence, suggesting the existence of two different enzymes potentially related to fragmentation and autolysis in *P. chrysogenum*.

Heterologous expression of the *S. hindustanus ble^R* gene under the control of *PnagA* conferred phleomycin resistance in *P. chrysogenum*. Similar resistance levels were obtained with plasmids expressing *ble^R* under the control of either *PgdhA* or *PpcbC* [34]. *pALfleo10* in-

cludes *Cm^R* as selection marker in *E. coli*, instead of the β -lactamase (*Ap^R*) marker present in other plasmids which can be potentially harmful to be used with β -lactam producers, and fulfils the requirements for stable integration in *P. chrysogenum*, allowing cloning and gene expression.

As a conclusion, the *nagA* gene encoding one of the most largely represented proteins in *P. chrysogenum* strains industrially used for penicillin production has been characterized, and its promoter region can be used as a tool to drive gene expression in fungal hosts.

Acknowledgements

The authors thank T. Carmona, P. Merino, M. Sandoval, E.G. Quiñones, C. Flórez and P. Oricheta for their excellent technical assistance.

References

- [1] Kostina, A.M., Babitskaia, V.G. and Lobanok, A.G. (1978) Chitin of mycelial fungi of the genus *Penicillium*. Prikl. Biokhim. Mikrobiol. 14, 586–593.
- [2] Henriksen, C.M., Christensen, L.H., Nielsen, J. and Villadsen, J. (1996) Growth energetics and metabolic fluxes in continuous cultures of *Penicillium chrysogenum*. J. Biotechnol. 45, 149–164.
- [3] Gómez-Miranda, B. and Leal, J.A. (1985) Carbohydrate stability during ageing in *Penicillium expansum* cell wall. Microbiologia 1, 67–75.
- [4] Bulman, R.A. and Chittenden, G.J. (1976) The mycelial cell wall of *Penicillium charlesii* G. Smith NRRL 1887. Biochim. Biophys. Acta 444, 202–211.
- [5] Ueda, M. and Arai, M. (1992) Purification and some properties of β -N-acetylglucosaminidase from *Aeromonas* sp. 10S-24. Biosci. Biotechnol. Biochem. 56, 1204–1207.
- [6] Reyes, F., Calatayud, J., Vázquez, C. and Martínez, M.J. (1989) β -N-acetylglucosaminidase from *Aspergillus nidulans* which degrades chitin oligomers during autolysis. FEMS Microbiol. Lett. 53, 83–87.
- [7] Pócsi, I., Pusztahelyi, T., Bogáti, M.S. and Szentirmai, A. (1993) The formation of N-acetyl- β -D-hexosaminidase is repressed by glucose in *Penicillium chrysogenum*. J. Basic Microbiol. 33, 259–267.
- [8] Pócsi, I., Pócsi, I. and Pusztahelyi, T. (1999) Physiological and enzymological characterization of the β -N-acetylhexosaminidase of *Penicillium chrysogenum*. J. Basic Microbiol. 39, 177–187.
- [9] Kim, S., Matsuo, I., Ajisaka, K., Nakajima, H. and Kitamoto, K. (2002) Cloning and characterization of the *nagA* gene that encodes β -N-acetylglucosaminidase from *Aspergillus nidulans* and its expression in *Aspergillus oryzae*. Biosci. Biotechnol. Biochem. 66, 2168–2175.
- [10] Cannon, R.D., Niimi, K., Jenkinson, H.F. and Shepherd, M.G. (1994) Molecular cloning and expression of the *Candida albicans* β -N-acetylglucosaminidase (*HEX1*) gene. J. Bacteriol. 176, 2640–2647.
- [11] Peterbauer, C.K., Lorito, M., Hayes, C.K., Harman, G.E. and Kubicek, C.P. (1996) Molecular cloning and expression of the *nag1* gene (N-acetyl- β -D-glucosaminidase-encoding gene) from *Trichoderma harzianum* P1. Curr. Genet. 30, 325–331.

- [12] Niimi, M., Niimi, K. and Cannon, R.D. (1997) Temperature-related expression of the vacuolar aspartic proteinase (APR1) gene and β -N-acetylglucosaminidase (*HEX1*) gene during *Candida albicans* morphogenesis. FEMS Microbiol. Lett. 148, 247–254.
- [13] Peterbauer, C.K., Brunner, K., Mach, R.L. and Kubicek, C.P. (2002) Identification of the *N*-acetyl-D-glucosamine-inducible element in the promoter of the *Trichoderma atroviride nag1* gene encoding *N*-acetyl-glucosaminidase. Mol. Genet. Genomics 267, 162–170.
- [14] Pusztahelyi, T., Pócsi, I. and Szentirmai, A. (1997) Aging of *Penicillium chrysogenum* cultures under carbon starvation: II: protease and *N*-acetyl- β -D-hexosaminidase production. Biotechnol. Appl. Biochem. 25, 87–93.
- [15] McNeil, B., Berry, D.R., Harvey, L.M., Grant, A. and White, S. (1998) Measurement of autolysis in submerged batch cultures of *Penicillium chrysogenum*. Biotechnol. Bioeng. 57, 297–305.
- [16] McIntyre, M., Eade, J.K., Cox, P.W., Thomas, C.R., White, S., Berry, D.R. and McNeil, B. (2001) Quantification of autolysis in *Penicillium chrysogenum* by semiautomated image analysis. Can. J. Microbiol. 47, 315–321.
- [17] Fierro, F., Barredo, J.L., Díez, B., Gutiérrez, S., Fernández, F.J. and Martín, J.F. (1995) The penicillin gene cluster is amplified in tandem repeats linked by conserved hexanucleotide sequences. Proc. Natl. Acad. Sci. USA 92, 6200–6204.
- [18] Sambrook, K.J., Fritsch, E.F. and Maniatis, T. (1989) Molecular Cloning. A Laboratory Manual, second ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
- [19] Barredo, J.L., Álvarez, E., Cantoral, J.M., Díez, B. and Martín, J.F. (1988) Glucokinase deficient mutant of *Penicillium chrysogenum* is derepressed in glucose catabolite regulation of both β -galactosidase and penicillin biosynthesis. Antimicrob. Agents Chemother. 32, 1061–1067.
- [20] Sullivan, P.A., McHugh, N.J., Romana, L.K. and Shepherd, M.G. (1984) The secretion of *N*-acetylglucosaminidase during germ-tube formation in *Candida albicans*. J. Gen. Microbiol. 130, 2213–2218.
- [21] Bradford, M.M. (1976) A rapid and sensitive method for the quantitation of micrograms quantities of protein utilizing the principle of protein-dye binding. Anal. Biochem. 72, 248–254.
- [22] Jones, C.S. and Kosman, D.J. (1980) Purification, properties, kinetics, and mechanism of β -N-acetylglucosaminidase from *Aspergillus niger*. J. Biol. Chem. 255, 11861–11869.
- [23] Specht, C.A., DiRusso, C.C., Novotny, C.P. and Ullrich, R.C. (1982) A method for extracting high-molecular-weight deoxyribonucleic acid from fungi. Anal. Biochem. 119, 158–163.
- [24] Barredo, J.L., Cantoral, J.M., Álvarez, E., Díez, B. and Martín, J.F. (1989) Cloning, sequence analysis and transcriptional study of the isopenicillin N synthase of *Penicillium chrysogenum* AS-P-78. Mol. Gen. Genet. 216, 91–98.
- [25] Altschul, S.F., Madden, T.L., Schäffer, A.A., Zhang, J., Zhang, Z., Miller, W. and Lipman, D.J. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucl. Acids Res. 25, 3389–3402.
- [26] Mathews, D.H., Sabina, J., Zuker, M. and Turner, D.H. (1999) Expanded sequence dependence of thermodynamic parameters improves prediction of RNA secondary structure. J. Mol. Biol. 288, 911–940.
- [27] Mullaney, E.J., Hamer, J.E., Roberti, K.A., Yelton, M.M. and Timberlake, W.E. (1985) Primary structure of the *trpC* gene from *Aspergillus nidulans*. Mol. Gen. Genet. 199, 37–45.
- [28] Díez, B., Álvarez, E., Cantoral, J.M., Barredo, J.L. and Martín, J.F. (1987) Selection and characterization of *pyrG* mutants of *Penicillium chrysogenum* lacking orotidine-5'-phosphate decarboxylase and complementation by the *pyr4* gene of *Neurospora crassa*. Curr. Genet. 12, 277–282.
- [29] Calmels, T.P.G., Martin, F., Durand, H. and Tiraby, G. (1991) Proteolytic events in the processing of secreted proteins in fungi. J. Biotechnol. 17, 51–66.
- [30] Cubero, B. and Scazzocchio, C. (1994) Two different, adjacent and divergent zinc finger binding sites are necessary for CRE-mediated carbon catabolite repression in the proline gene cluster of *Aspergillus nidulans*. EMBO J. 13, 407–415.
- [31] Marshall, R.D. (1972) Glycoproteins. Annu. Rev. Biochem. 41, 673–702.
- [32] Díez, B., Barredo, J.L., Álvarez, E., Cantoral, J.M., van Solingen, P., Groenen, M.A., Veenstra, A.E. and Martín, J.F. (1989) Two genes involved in penicillin biosynthesis are linked in a 5.1 kb *SalI* fragment in the genome of *Penicillium chrysogenum*. Mol. Gen. Genet. 218, 572–576.
- [33] Davis, M.A. and Hynes, M.J. (1991) Regulatory circuits in *Aspergillus nidulans*. In: More Gene Manipulations in Fungi (Bennet, J.W. and Lasure, L.L., Eds.), pp. 151–189. Academic Press, San Diego, CA.
- [34] Díez, B., Mellado, E., Rodríguez, M., Bernasconi, E. and Barredo, J.L. (1999) The NADP-dependent glutamate dehydrogenase gene from *Penicillium chrysogenum* and the construction of expression vectors for filamentous fungi. Appl. Microbiol. Biotechnol. 52, 196–207.