Structural characterization of the *bglH* gene encoding a beta-glucosidase-like enzyme in an endophytic *Bacillus pumilus* strain

Andréa C. Bogas¹, Maria Angelica E. Watanabe¹, Aneli Barbosa², Laurival A. Vilas-Boas¹, Ana C. Bonatto³, Robert Dekker², Emanuel M. Souza³ and Maria Helena P. Fungaro¹

¹Centro de Ciências Biológicas, Universidade Estadual de Londrina, Londrina, PR, Brazil.
²Centro de Ciências Exatas, Universidade Estadual de Londrina, Londrina, PR, Brazil.
³Departamento de Bioquímica e Biologia Molecular, Universidade Federal do Paraná, Curitiba, PR, Brazil.

Abstract

A beta-glucosidase-like enzyme-encoding gene (*bglH*) of an endophytic *Bacillus pumilus* strain (CL16) was cloned using a shotgun genomic library constructed in *Escherichia coli*. The nucleotide sequence of the entire cloned fragment (2484 bp) was determined and characterized. An incomplete open reading frame (ORF) of 534 bp (ORF1) designated *bglP* and a complete ORF of 1419 bp (ORF2) designated *bglH*, located in the fragment, are organized in an operon. The protein deduced from 1419 bp (ORF2) had 472 amino acid residues without a characteristic signal peptide sequence, suggesting that the enzyme is localized in the cytoplasm. The amino acid sequence deduced from *bglH* gene had high similarity with *β*-glucosidases from the glycosyl hydrolase family 1. Over-expression of the *B. pumilus* *bglH* gene in *E. coli* showed a 54 kDa protein whose identity was confirmed by mass spectrometry (MALDI-TOF).

Key words: *Bacillus pumilus*, *bglH*, glucosidase, glycosyl hydrolase 1, PTS.

Received: November 28, 2005; Accepted: August 17, 2006.

Cellulose comprises the major carbohydrate polymer of the plant cell wall. It is an unbranched polymer composed of anhydro-1,4-D-glucopyranoside units linked by *β*-glucosidic bonds. Enzymatic degradation of cellulose within the polysaccharide matrix of the cell wall requires the synergism of multiple enzymes such as the cellulases, exo- (*cellobiohydrolase*) and endo-*β*-1,4-glucanases, and *β*-glucosidases (*cellobiases*) (Knauf and Moniruzzaman, 2004). The *β*-glucosidases are widespread in microorganisms where they metabolize various carbohydrate substrates, including cellobiose, produced as a consequence of cellulose hydrolysis, and aromatic *β*-glucosides such as arbutin and salicin that are produced by a variety of plants (Tajima et al., 2001; Spiridonov and Wilson, 2001; Park et al., 2002; Marques et al., 2003; An et al., 2005).

Considerable polymorphism in *β*-glucosidase forms, functions and kinetics has been reported (Ogunseitan, 2003). Although several cellulolytic enzymes released by phytopathogens have already been well-characterized, knowledge on the uptake and hydrolysis of carbohydrates by endophytic microorganisms is limited. Endophyte microorganisms colonize inner plant tissues, living symbiotically with the host species (Azevedo et al., 2000). These microorganisms have been investigated as a source of new genes and proteins for use in industrial processes (Stamford et al., 2001, 2002; Pleban et al., 1997; Reddy et al., 1996; Moy et al., 2002; Lima et al., 2004).

Recently, 15 endophytic strains of *Bacillus* spp. isolated from *Citrus* were evaluated for cellulolytic activity (Lima et al., 2004). The *Bacillus pumilus* strain CL16 showed high cellulase activity and was selected for further studies. We have cloned the *β*-1,4-endoglucanase *eglA* gene from strain CL16 and expressed it in *E. coli*. The endo-1,4-*β*-glucanase EglA has high thermostability, an important feature in biotechnical processes that require high temperatures (Lima et al., 2004).

During the study described in the present paper we isolated and characterized a new locus of *β*-glucoside sugar utilization genes from the endophytic *B. pumilus* CL16 strain.

Using the degenerated primers DEG1F (5’-ATRACC TACTgNAARTTRgg-3’) and DEG1R (5’gCRAANCCY AgHTARACggT-3’) designed based on the amino acid regions conserved amongst *β*-glucosidases reported for *Ba-
cillus spp, a 560 bp *B. pumilus* fragment was obtained and successfully cloned in the pUC18 vector. Nucleotide sequence, determined using the DYEnamic ET DYE Terminator Cycle Sequencing Kit (Amersham Biosciences, Germany) on MegaBACE 1000 (Amersham Pharmacia Biotech, Germany), showed high similarity with the *bglH* gene from *B. subtilis* subsp. *subtilis* strain 168 (E value = 1e-17).

This fragment was successfully used as a probe for screening the *bglH* gene in a shotgun genomic library constructed from *B. pumilus* strain CL16 (Lima et al., 2004). Hybridization was done using the DIG High Prime DNA Labeling and Detection Starter Kit II, according to the manufacturer’s instructions (Roche, Germany). Pre-hybridization (30 min) and hybridization (overnight) steps were at 42 °C. Only one positive transformant was recovered from 2400 colonies. The recombinant plasmid isolated from this clone was denoted pMH2. The presence of the *bglH* gene was confirmed by two steps of sequencing. Firstly, using the M13 primers (Amersham Biosciences, Germany) and then with a set of new primers, GLICO1 F (5’-TCCAgAgATTCTTggACAAgT-3’), GLICO2 R (5’-CACTTggAACAAATTggTgATg-3’) and GLm F (5’-gCATAAgCACggAATTgAgTC3-3’) designed specifically for a *bgl* internal segment.

Two open reading frames (ORF) were found to compose the insert: an incomplete ORF of 534 bp (ORF1) and a complete one with 1419 bp (ORF2), which presented high similarity with the *bglP* and *bglH* genes, respectively, both from *B. subtilis* strain 168 (Kunst et al., 1997). The *bglP* gene from *B. subtilis* encodes an aryl-β-glucoside-specific enzyme II of the phosphoenolpyruvate sugar: phosphotransferase system (PTS), whereas the activity of BglH from *B. subtilis* was only recently directly demonstrated (Setlow et al., 2004). These authors showed that the *bglH* gene from *B. subtilis* encodes an aryl-phospho-β-D-glucosidase and that this gene was induced by aryl-β-D-glucosides.

The nucleotide sequence of the entire insert (2484 bp) and the deduced protein sequence of the *bglH* gene from *B. pumilus* are shown in Figure 1. The ORF1 was upstream from ORF2, and separated by a 24 bp sequence. The absence of a promoter sequence between the ORFs identified by us, and their similarity with the operon described for *B. subtilis* strain 168 (Entrez Gene-NCBI server) suggest that they are organized into an operon. The nucleotide sequence downstream from ORF2 was compared to other sequences deposited at NCBI GenBank (BLASTX) but no similarity could be identified.

We found that the 1419-nucleotide-long ORF2 had a GC content of 41.6%. It was preceded by a potential ribosome binding site (AGGAGG) that was 9 bp upstream from the putative ATG start codon, but, as expected, with no adjacent promoter sequence. Downstream from the TAA stop codon, no sequence resembling a ρ-independent transcriptional terminator could be identified. The protein deduced from the ORF-complete sequence had 472 amino acid residues with an estimated molecular mass of 53.9 kDa and a isoelectric point of 4.97. The CDSearch program (NCBI server) revealed that BglH had a single domain consisting of the glycosyl hydrolase family 1 (GH1) sequence, covering 464 residues from amino acid 4 to 468.

The deduced *B. pumilus* BglH amino acid sequence was compared to other BglH sequences deposited at NCBI
GenBank (Figure 2) and found to be highly similar to homologous enzymes from several *Bacillus* species: *B. halodurans* (gi 10173219), *B. subtilis* strain 168 (gi 7435440), *B. cereus* (gi 52144364), *B. thuringiensis* (gi 49477027), *B. licheniformis* (gi 52082495), *B. clausii* (gi 56965550). The Glu\(^{175}\) and Glu\(^{369}\) residues, the catalytic nucleophile, conserved in the *B. pumilus* BglH, are characteristic of family 1 proteins that hydrolyze glucosidic bonds by acid/base catalysis (Withers and Aebersold, 1995).

The absence of a signal peptide sequence (SignalP 3.0 program) and the lack of potential transmembrane regions (TMpred program) in the BglH sequence suggest that the enzyme is localized in the cytoplasm, as is the case for most bacterial β-glucosidases (Bhatia *et al*., 2002 and references therein).

We constructed an expression plasmid for the over-production of β-glucosidase (Bgl) by amplifying the *bgl* open reading frame (ORF) using the polymerase chain reaction (PCR) and the MHF (5'-CACCATgAACAAgTAATTAgAAAATAACAT-3') and MHR (5'-TTAgTAATCCAAATTgTTCCCCCATTTg-3') primer pair. For DNA polymerization the AccuPrime Pfx enzyme (Invitrogen, USA) was used. The amplified product was cloned into the pENTR/SD/D-TOPO plasmid producing the entry vector of the Gateway Cloning System (Invitrogen, USA). From the entry vector the *bgl* gene was transferred to the expression vector pET-DEST42 by in vitro site-specific recombination. The recombinant expression plasmid containing the *bgl* gene was named the pAB1 plasmid. The cloned fragment was completely sequenced to confirm that no mutations were introduced during the amplification procedures.

Transformed *Escherichia coli* BL21 CodonPlus (DE3) cells harboring the pAB1 expression vector were grown on Luria-Bertani medium (LB) supplemented with 250 μg mL\(^{-1}\) ampicillin and incubated at 37°C until the log phase (OD\(_{600}\)nm = 0.2). After induction with 0.5 mmol L\(^{-1}\) IPTG (isopropyl-beta-D-thiogalactopyranoside) a strong band of 54 kDa was detected by sodium dodecyl sulfate polyacrylamide gel electrophoreses (SDS-PAGE) analysis (Laemmli, 1970), suggesting the over-expression of the *B. pumilus* bglH gene in *E. coli* (Figure 3). This molecular mass is consistent with that predicted from the amino acid sequence of the BglH protein. The over-expressed protein was completely purified and subjected to digestion with trypsin followed by peptide fingerprint analysis by matrix assisted laser desorption ionization time-of-flight mass spectrometry (MALDI-TOF/MS), in order to confirm the protein identity. The provided peptide masses were equivalent to the ones predicted, confirming that the over-expressed 54 kDa band was in fact the BglH protein of *B. pumilus*.

Crude protein extracts obtained from the *E. coli* cells harboring the pAB1 plasmid were assayed for activity against p-nitrophenyl-β-D-glucopyranoside and cellobiose. Each assay consisted of 0.5 mL 5 mM of p-nitrophenyl-β-D-glucopyranoside or 0.2% of cellobiose as substrate, 0.1 mL of 50 mM phosphate buffer (pH 5.8 to 7.5) or McIlvaine buffer (pH 3 to 7) and 0.05 mL of crude enzyme. The mixture was incubated for 1 h at 37°C and the activity of β-glucosidase towards p-nitrophenyl-β-D-glucopyranoside was estimated by measuring the amount of p-nitrophenol released at 400 nm. The activity of β-glucosidase toward cellobiose was estimated by measuring the glucose released by the glucose oxidase method (Glucose Enzyme Color Kit, Bio Diagnostica, Brazil). The enzymatic activity of Bgl against p-nitrophenyl-β-D-glucopyranoside was assayed in McIlvaine buffer (pH 7.0) using the same protocol. Little activity against synthetic aryl-β-D-glucosides and no activity against cellobiose were observed. The highest activity against p-nitrophenyl-β-D-
glucopyranoside (0.106 μmol min⁻¹ mL⁻¹) was observed using McIlvaine buffer at pH 7.0. This low activity was also observed using purified Bgl. The B. subtilis BglH also showed very low activity against non-phosphorylated β-glucosides. The enzyme activity against aryl-phospho-β-D-glucosides was not measured because there was no commercial supplier for this substrate.

Acknowledgments

We would like to thank Dr. Fábio de Oliveira Pedrosa (GENOPAR Consortium Coordinator) for providing us with laboratory facilities. This work was partially supported by the Brazilian agency Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq). A.C. Bogas is grateful to the Brazilian agency Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) for a scholarship.

References


Reddy PV, Lam CK and Belanger FC (1996) Mutualistic fungal endophytes express a protease that is homologous to proteases suspected to be important in fungal pathogenicity. Plant Physiol 111:1209-1218.


Associate Editor: Luis Carlos de Souza Ferreira