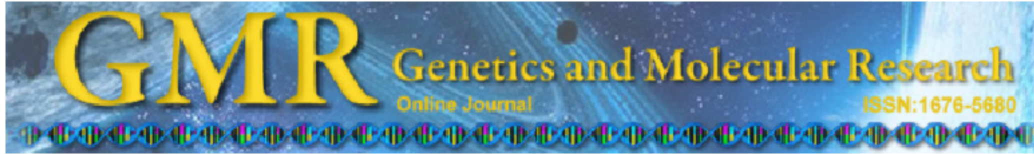

Molecular cloning and characterization of the glyceraldehyde-3-phosphate dehydrogenase gene from *Penicillium expansum* PE-12

T. Zhang[1], Z. Qi[2], Q.S. Yu[1] and K.X. Tang[1]

[1]Plant Biotechnology Research Center,
School of Agriculture and Biology,
Shanghai Jiao Tong University, Shanghai, China
[2]School of Life Science and Biotechnology, Shanghai Jiao Tong University,
Shanghai, China

Corresponding author: K.X. Tang
E-mail: kxtang@sjtu.edu.cn / kxtang1@yahoo.com

Genet. Mol. Res. 12 (3): 2442-2454 (2013)
Received June 24, 2012
Accepted August 31, 2012
Published February 6, 2013
DOI http://dx.doi.org/10.4238/2013.February.6.2

ABSTRACT. *Penicillium expansum* produces large amounts of lipase, which is widely used in laundry detergent and leather industry. We isolated the glyceraldehyde-3-phosphate dehydrogenase gene (*PeGPD*) from *P. expansum* PE-12 through reverse transcriptase PCR and 5'-3'-rapid amplification of cDNA ends (RACE-PCR). The gene is 1266 bp long, including an ORF of 1014 bp, encoding a polypeptide chain of 337 amino acids. A phylogenetic tree based on GPD proteins showed that *P. expansum* is close to *Aspergillus* species, but comparatively distant from *P. marneffei*. Southern blot results revealed a single copy of *PeGPD*, and expression analysis gave evidence of high expression levels. *PeGPD* genes have potential for genetic engineering of *P. expansum* for industrial lipase production.

Key words: Glyceraldehyde-3-phosphate dehydrogenase gene; *Penicillium expansum*; 5'-3'-RACE-PCR; Codon bias; Homology modeling; Southern blot

 ©FUNPEC-RP www.funpecrp.com.br

INTRODUCTION

Penicillium expansum is of considerable importance from a biotechnological aspect as a consequence of its content of versatile lipase (Dai and Xia, 2005). The lipase secreted by *P. expansum* (called PEL) can effectively hydrolyze the 3 ester bonds of triacylglycerol under alkaline condition (Bian et al., 2005) and is therefore widely used in laundry detergent and leather industry (www.leveking.com). However, back to the first discovery of PEL, the lipase productivity of this species was rather low, which made it impractical for commercial production (Stocklein et al., 1993). Later on, several strategies were adopted to maximize the production level of PEL including successive mutagenesis and selection (Zhang et al., 2001), fermentation condition optimization (Dai and Xia, 2005), and heterologous expression (Yuan et al., 2003; Zhang et al., 2010). After 2 decades of mutagenesis and selection accompanied by fermentation condition optimization, the industrial production of PEL was realized. However, due to excessive mutagenesis and selection, *P. expansum* mutants, such as the industrial strain *P. expansum* PE-12, became blunt to physical and chemical mutagens, which was a tremendous obstacle for further enhancement of its lipase productivity. Under such circumstances, reasonable genetic modification of the industrial strain based on homologous expression should be a promising strategy to resolve this problem.

However, to the best of our knowledge, no homologous gene promoters have been reported in *P. expansum*. Previous observations in yeast and higher eukaryotes have demonstrated that the glyceraldehyde-3-phosphate dehydrogenase (GPD, EC 1.2.1.12) gene (*GPD*) is under the control of a strong constitutive promoter. In such case, the expression level of the GPD protein could reach up to 5% of the soluble cellular protein (Piechaczyk et al., 1984). Furthermore, 2-5% of the poly(A)+ RNA in yeast is *GPD* mRNA (Holland and Holland, 1978). Therefore, the promoters of *GPD* genes were widely used in homologous expression for strain improvements (Li et al., 2000; Kajita et al., 2004).

As part of the effort to develop a homologous expression system for *P. expansum* PE-12 based on a *GPD* promoter, we report here for the first time, the cloning and characterization of an ORF, which encodes a GPD. Molecular organization, codon usage, copy number, and expression analysis are discussed.

MATERIAL AND METHODS

Strains, media, and oligonucleotides

P. expansum PE-12 is an industrial strain that has been subjected to several generations of mutagenesis to increase its lipase productivity. The strain was cultivated in fermentation medium containing 5 g/L starch, 40 g/L soybean powder, 2 g/L Na_2HPO_4, 3 g/L K_2SO_4, 3 g/L $MgSO_4$, 5 g/L $CaCO_3$, and 0.15 g/L $FeSO_4$. *Escherichia coli* DH5α (Takara, Japan) was grown in LB medium (10 g/L tryptone, 5 g/L yeast extract, 10 g/L sodium chloride) and handled as described elsewhere (Sambrook and Russell, 2001).

Rapid amplification of cDNA end (RACE)-PCR oligonucleotides were designed using Primer premier 5.0, and all primers were synthesized by BBI (China). Molecular biological kits were used according to manufacturer manuals.

RNA isolation

P. expansum PE-12 was grown in fermentation medium at 26°C for 72 h in an orbital shaker. The mycelia were collected by centrifugation, ground in liquid nitrogen, lysed with Trizol® (Invitrogen, USA), deproteinated by chloroform and finally added to a silica column for RNA purification (Tiangen, China). After DNase treatment for 10 min, the RNA was eluted with RNase-free water.

Cloning of the full-length cDNA of the *PeGPD* gene

Cloning of *PeGPD* was by RACE primers GPD-F and GPD-R (Table 1), which were designed according to the conservative regions of the homologous *GPD* genes from GenBank. The partial conservative fragment of the *PeGPD* gene was obtained by RT-PCR (AMV RT-PCR kit, Takara, Shiga, Japan) and subsequently used to design gene-specific primers (GPD3-1, GPD3-2, GPD5-1, and GPD5-2) for RACE. The cDNA was synthesized using the SMART technology (SMART™ RACE cDNA Amplification kit) for 3'- and 5'-RACE (Clontech Laboratories, Inc.). For 5'-RACE, RNA was reversely transcribed with the 5'-RACE CDS primer and SMART II oligonucleotide (provided in the kit). Cloning of *PeGPD* was by 5'-RACE primers GPD5-1 and GPD5-2, which were designed according to the conserved ORF of the *PeGPD* gene. For 3'-RACE, RNA was reversely transcribed with the 3'-RACE CDS Primer A (provided in the kit). Based on the conserved fragment of the *PeGPD* gene, the specific primers GPD3-1 and GPD3-2 were designed and synthesized. The first round of PCR was carried out with primer GPD5-1 and Universal Primer A Mix (provided in the kit). The PCR product was then diluted 50-fold for subsequent amplification with primer GPD5-2 and Nested Universal Primer A (provided in the kit). 3'- and 5'-RACE was performed according to the SMART™ RACE cDNA Amplification Kit user manual. The full-length *PeGPD* was deduced from the obtained sequences and subsequently amplified by proofreading RT-PCR amplification with primers GPDfull5 and GPDfull3. All the primers used in RACE are listed in Table 1.

Table 1. Primers used for PCR amplification in this study.

Primers	Purpose	Primer sequence (5' to 3')
GPD-F	Clone	GTYGTCGCTGTCAACGACCCCY(C/T)
GPD-R	Clone	GACVAGCTTGATGAAGTTGGV(A/G/C)
GPD3-1	3'-RACE	GTGCTTCCTACGAGGAGATCAAG
GPD3-2	3'-RACE	CAAGGCCGCTTCCGAGAACGGCG
GPD5-1	5'-RACE	CATCCTTGGTGTAAGACTTGTTGTT
GPD5-2	5'-RACE	CCATGACGAACATGGGGGCATC
3'-RACE CDS primer A	RACE	AAGCAGTGGTATCAACGCAGAGTAC(T)30VN
5'-RACE CDS primer A	RACE	(T)25VNAAGCAGTGGTATCAACGCAGAGTACGCGGG
SMART II™ A Oligo UPM	RACE	Long (0.4 M): CTAATACGACTCACTATAGGGCAAGCAGTGGTATCAACGCAGAGT Short (2 M): CTAATACGACTCACTATAGGGC
NUP	RACE	AAGCAGTGGTATCAACGCAGAGT
GPDfull5	Clone	ATGGTTGTCCAGGTCGGTAT
GPDfull3	Clone	CTACTGGCCATCGACACC
GPD-SouthF	Southern blot	TTCACCACTCAGGAGAAGGC
GPD-SouthR	Southern blot	CGATGCGGCAGGTAAGGT
GPDRT-F	RT-PCR	TATCAACGGTTTCGGTCG
GPDRT-R	RT-PCR	GGATGGCAGCGGGGTCAC
PELRT-F	RT-PCR	CCCTTAACGCCTTCCCCA
PELRT-R	RT-PCR	AGCGTACATGCCATTGCC

Y(C/T) V(A/G/C).

Cloning of the genomic sequence

Total genomic DNA was isolated from *P. expansum* PE-12 by the cetyltrimethylammonium bromide method and used as template in PCR amplification with primers, designed corresponding to the *PeGPD* full-length cDNA, GPDfull5 and GPDfull3 (Table 1) to check the existence of intron(s). The amplification reactions were denatured at 94°C for 5 min, followed by 35 cycles of amplification (94°C for 30 s, 56°C for 30 s, and 72°C for 120 s) and ended after 72°C for 10 min. The target band was recovered and cloned into pMD18-T vector, followed by sequencing.

Bioinformatic analysis

The percentage identities of the *PeGPD* with other known *GPD* genes were calculated using the BLAST program at the National Center for Biotechnology Information (NCBI, http://www.ncbi.nlm.nih.gov/BLAST). The amino acid sequence deduction as well as the theoretical molecular weight and isoelectric point (pI) calculation were carried out at website (http://www.expasy.org). The phylogenetic analysis of the PeGPD protein and GPDs from other species was performed through alignment with CLUSTAL X (1.81). An evolutionary tree from the full-length amino acid sequence available from animals, plants and microorganisms was conducted by the neighbor-joining method (Saitou and Nei, 1987) using MEGA version 3.1 (Kumar et al., 2004). The codon usage frequency of the *PeGPD* gene was calculated with the Countcodon program of the NCBI Codon Usage Database (http://www.kazusa.or.jp/codon/).

Secondary structure of the PeGPD protein was predicted using the SOMPA server (http://npsa-pbil.ibcp.fr/cgi-bin/npsa_automat.pl?page=npsa_sopma.html). A homology model of the PeGPD was further developed to investigate its catalytic and NAD^+-binding domain. A protein-protein BLAST search to PDB was performed using PeGPD for potential structural templates to develop the homology model. The template selected was the GPD crystal structure from *Oryctolagus cuniculus* (PDB: 1J0X_R), which had 73% sequence identity to the query sequence. The model was built using the alignment model program at the SWISS-MODEL website (http://swissmodel.expasy.org/). The PeGPD model was superposed on the template, and the Cα and backbone root mean square deviation values were calculated using FATCAT (Ye and Godzik, 2003) for validation of the model. The model was further checked with a Ramachandran plot using PROCHECK (Laskowski et al., 1993).

Southern blot analysis

The copy number of the *GPD* gene in *P. expansum* PE-12 was determined as follows. Genomic DNA was digested with the restriction enzymes *Eco*RI, *Hin*dIII, *Kpn*I, *Dra*I, and *Bam*HI. Digestion products were then separated on a 1.0% agarose gel and transferred to a Hybond™-N^+ membrane (Amersham Pharmacia, UK). The membrane was probed with a labeled DNA fragment containing the structural *PeGPD* gene according to the Amersham protocol. Probes were amplified with the primers GPD-SouthF and GPD-SouthR and subsequently labeled using the Alkphos Direct Labeling kit (Amersham). Hybridization was carried out with high stringency wash at 55°C and signal detection was performed using the CDP-Star detection system (Amersham), according to manufacturer instructions.

Gene expression analysis

The value of *GPD* promoter in engineering the industrial strain *P. expansum* PE-12 was determined by comparing the transcriptional strength of *GPD* and *PEL* through semi-quantitative RT-PCR. The RNA sample was digested with DNase I (RNase-free) prior to use. Aliquots of 400 ng total RNA were employed in the reverse transcriptase reaction using a random hexamer primer for the synthesis of first-strand cDNA. The amplification reactions of RT-PCR were performed with gene-specific primers GPDRT-F and GPDRT-R for *GPD* and primers PELRT-F and PELRT-R for *PEL* under the following conditions: pre-denaturation at 94°C for 5 min, followed by 29 cycles of 30 s at 94°C, 30 s at 56°C and 30 s at 72°C. RT-PCR images were captured by a UVP transilluminator.

RESULTS AND DISCUSSION

Characterization of the full-length cDNA and genomic DNA sequences of *PeGPD*

Using the degenerate primers GPD-F and GPD-R, a specific fragment of about 850 bp was obtained, and the sequence alignment displayed a high similarity to *GPDs* from other fungi. Based on the sequence of this fragment, 2 pairs of primers were designed for 5'- and 3'-RACE. Following the steps described, the full-length cDNA of the *PeGPD* gene (GenBank accession No. JN389439) was obtained, which was subsequently verified by PCR and sequencing. This new *PeGPD* gene was 1306 bp long comprising a 126-bp 5'-untranslated region (UTR), 1014-bp ORF encoding a 337-amino acid residue protein, and a 166-bp 3'-UTR (Figure 1). The 5'-UTR contains one putative TATA box and 3 CAAT boxes. A pyrimidine-rich stretch was found at -69 bp upstream of the start codon. In a number of filamentous fungi, transcription initiation sites appear in or immediately downstream of such pyrimidine-rich sequence (Vastag et al., 2004). In the terminator region, a putative polyadenylation signal (ATTAA) was identified about 122 bp downstream of the stop codon.

The PCR for genomic sequence resulted in a specific fragment of 1558 bp, which was 252 bp longer than that of the coding sequence. Compared to the cDNA, the genomic gene contained four introns. The lengths of the 4 introns were 56, 70, 64, and 62 bp (Figure 2). This result coincided well with the exon-intron arrangement of *GPD* genes from *Curvularia lunata* and *Cochliobolus heterostrophus*, both of which also showed a similar genomic pattern (Thirach et al., 2008).

Codon usage in *PeGPD*

The codon bias in *GPD* has been previously reported in *Ganoderma lucidum* (Fei et al., 2006), *Rhizomucor miehei* (Vastag et al., 2004), and *Ermothecium ashbyi* (Sengupta and Chandra, 2011). In this research, we also noticed a strong codon bias which may be related to the abundance of the corresponding tRNA molecules (Table 2). In *PeGPD* only 35 of 61 possible sense codons have been utilized, and moreover, only 30 of these 35 codons have been used more than once. According to previous research, this bias could suggest a strong promoter activity and high expression level (Van Bogaert et al., 2008). Several amino acids exhibit a strict preference for a single codon such as Phe (TTC), His (CAC), Lys (AAG), Glu

 ©FUNPEC-RP www.funpecrp.com.br

(GAG), and Cys (TGC). When a choice among A, T, C, and G in the wobble position exists, there is a strong preference for C. Codon bias revealed by this study might benefit downstream research on over-expression of valuable protein, such as PEL, in *Penicillium* hosts, because it provides a reference for the codon engineering of the target protein (Gustafsson et al., 2004).

```
-126 attgcgtcagtcatccattcattaatcctcccctcccccagatcttctttttttttcttctcttctttctccatcccaat
     tctcccctttcatcaattacttcactatatacataaaccgcaatc
1    ATG GTT GTC CAA GTC GGT ATC AAC GGT TTC GGT CGT ATT GGC CGC ATT GTC TTC CGC   57
1     M   V   V   Q   V   G   I   N   G   F   G   R   I   G   R   I   V   F   R   19
58   AAT GCT ATC AAC AAC CCC GAT GTT GAG GTT GTC GCT GTC AAC GAC CCC TTC ATT GAG   114
20    N   A   I   N   N   P   D   V   E   V   V   A   V   N   D   P   F   I   E   38
115  ACT CAC TAC GCT GCC TAC ATG CTC AAG TAT GAC TCC ACC CAC GGC CAG TTC AAG GGT   171
39    T   H   Y   A   A   Y   M   L   K   Y   D   S   T   H   G   Q   F   K   G   57
172  GAG ATC GAG CAC GTT GAG GAC GGT CTC ATC GTC AAC GGC AAG AAG GTT ACC TTC TTC   228
58    E   I   E   H   V   E   D   G   L   I   V   N   G   K   K   V   T   F   F   76
229  GCT GAG CGT GAC CCC GCT GCC ATC CCC TGG GGC AAG GCC GGT GCC GCC TAC ATT GTC   285
77    A   E   R   D   P   A   A   I   P   W   G   K   A   G   A   A   Y   I   V   95
286  GAG TCC ACC GGT GTC TTC ACC ACT CAG GAG AAG GCC TCT GCT CAC TTG AAG GGT GGT   342
96    E   S   T   G   V   F   T   T   Q   E   K   A   S   A   H   L   K   G   G   114
343  GCT AAG AAG GTC ATC ATC TCC GCT CCC TCC GCC GAT GCC CCC ATG TTC GTC ATG GGT   399
115   A   K   K   V   I   I   S   A   P   S   A   D   A   P   M   F   V   M   G   133
400  GTC AAC AAC AAG TCT TAC ACC AAG GAT GTC AAC GTT CTC TCC AAC GCC TCT TGC ACC   456
134   V   N   N   K   S   Y   T   K   D   V   N   V   L   S   N   A   S   C   T   152
457  ACC AAC TGC TTG GCT CCT CTC GCC AAG GTC ATC AAC GAC AAC TTC GGT ATC GTT GAG   513
153   T   N   C   L   A   P   L   A   K   V   I   N   D   N   F   G   I   V   E   171
514  GGT CTC ATG ACC ACC GTC CAC TCC TAC ACC GCC ACC CAG AAG ACC GTC GAT GCT CCC   570
172   G   L   M   T   T   V   H   S   Y   T   A   T   Q   K   T   V   D   A   P   190
571  TCC TCC AAG GAC TGG CGT GGT GGC CGT ACT GCG GCC CAG AAC ATT ATC CCC TCC TCC   627
191   S   S   K   D   W   R   G   G   R   T   A   A   Q   N   I   I   P   S   S   209
628  ACC GGT GCC GCC AAG GCT GTC GGC AAG GTT ATC CCC GAG CTC AAC GGC AAG CTC ACC   684
210   T   G   A   A   K   A   V   G   K   V   I   P   E   L   N   G   K   L   T   228
685  GGT ATG GCC ATG CGT GTC CCC ACC GCC AAC GTC TCC GTT GTC GAC CTT ACC TGC CGC   741
229   G   M   A   M   R   V   P   T   A   N   V   S   V   V   D   L   T   C   R   247
742  ATC GAG AAG GGT GCT TCC TAC GAG GAG ATC AAG GCC GTC GTC AAG GCC GCT TCC GAG   798
248   I   E   K   G   A   S   Y   E   E   I   K   A   V   V   K   A   A   S   E   266
799  AAC GGC GAG CTC AAG GGT ATC CTT GGC TAC ACT GAG GAC CAG GTT GTC TCC ACC GAC   855
267   N   G   E   L   K   G   I   L   G   Y   T   E   D   Q   V   V   S   T   D   285
856  CTT AAC GGT GAT GAG CGC TCC TCC ATC TTC GAT GCC GCC GCC GGT ATC GCC CTT AAC   912
286   L   N   G   D   E   R   S   S   I   F   D   A   A   A   G   I   A   L   N   304
913  TCC AAC TTC ATC AAG CTT GTC TCC TGG TAC GAC AAC GAG TGG GGT TAC TCC CGC CGT   969
305   S   N   F   I   K   L   V   S   W   Y   D   N   E   W   G   Y   S   R   R   323
970  GTT GTT GAC CTC ATC TCC TAC ATT GCC GGT GTC GAT GGC CAG TAG                   1014
324   V   V   D   L   I   S   Y   I   A   G   V   D   G   Q   *                    337
   gaatcaggacagcgtcctggggtcagaagtgttctttaaggtgctgggtaccttgtcgcacaacctacgaatatgggtgcataaggg
ctccgccttgtagacaaaagcaagatgataacgtcattaatgaaacaaaccatttccaattcatccttccacaaatcgc  +166
```

Figure 1. Nucleotide and predicted peptide sequence of the *PeGPD* gene. ORF is shown in upper-case letters; 3'- and 5'-UTRs are in lower-case letters. The predicted amino acids are displayed with single-letter symbols below their capitalized codon using DNAsmac2.0.2.6. Putative CAAT boxes, TATA box with flanking pyrimidine-rich stretch in 5'-UTR and polyadenylation sequence in 3'-UTR are underlined. This sequence is available in GenBank database under the accession No. JN389439.

©FUNPEC-RP www.funpecrp.com.br

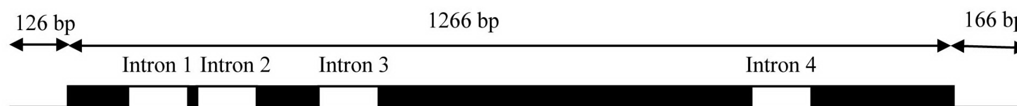

Figure 2. Gene structure of cloned *PeGPD* (positions 1-1558 bp) (GenBank accession No. JN389439.). UTRs, exons, and introns are indicated by bold lines, filled boxes, and open boxes, respectively.

Table 2. Codon usage in the *PeGPD* gene.

Codon	Amino acid	GPD1	GPD2	Codon	Amino acid	GPD1	GPD2
TTT	Phe	0	0	TAT	Tyr	0.09	1
TTC	Phe	1	11	TAC	Tyr	0.91	10
TTA	Leu	0	0	TAA	-	0	0
TTG	Leu	0.13	2	TAG	-	1	1
CTT	Leu	0.31	5	CAT	His	0	0
CTC	Leu	0.56	9	CAC	His	1	5
CTA	Leu	0	0	CAA	Gln	0.14	1
CTG	Leu	0	0	CAG	Gln	0.86	6
ATT	Ile	0.25	6	AAT	Asn	0.05	1
ATC	Ile	0.75	18	AAC	Asn	0.95	20
ATA	Ile	0	0	AAA	Lys	0	0
ATG	Met	1	7	AAG	Lys	1	22
GTT	Val	0.33	12	GAT	Asp	0.39	7
GTC	Val	0.67	24	GAC	Asp	0.61	11
GTA	Val	0	0	GAA	Glu	0	0
GTG	Val	0	0	GAG	Glu	1	18
TCT	Ser	0.13	3	TGT	Cys	0	0
TCC	Ser	0.87	20	TGC	Cys	1	3
TCA	Ser	0	0	TGA	-	0	0
TCG	Ser	0	0	TGG	Trp	1	4
CCT	Pro	0.09	1	CGT	Arg	0.55	6
CCC	Pro	0.91	10	CGC	Arg	0.45	5
CCA	Pro	0	0	CGA	Arg	0	0
CCG	Pro	0	0	CGG	Arg	0	0
ACT	Thr	0.19	4	AGT	Ser	0	0
ACC	Thr	0.81	17	AGC	Ser	0	0
ACA	Thr	0	0	AGA	Arg	0	0
ACG	Thr	0	0	AGG	Arg	0	0
GCT	Ala	0.35	13	GGT	Gly	0.68	21
GCC	Ala	0.62	23	GGC	Gly	0.32	10
GCA	Ala	0	0	GGA	Gly	0	0
GCG	Ala	0.03	1	GGG	Gly	0	0

Column GPD1 presents the fraction of synonymous codon. Column GPD2 gives the total number of specific codon in the *PeGPD* gene.

Characterization of the deduced PeGPD protein

Calculated by the pI/MW tool software (http://www.expasy.org), the pI and molecular weight of the PeGPD protein were presumed to be 6.01 and 35.97 kDa, respectively. The homology of the PeGPD protein was aligned with GPDs from animal, plant, bacterium, and fungi (Figure 3) and ranged from a minimum of 68% to a maximum of 87% with fungal GPD. Compared to GPDs from *Aspergillus niger* and *P. canescens*, PeGPD exhibited a high sequence identity of 87 and 86%, respectively. Aligned with members of other phyla, sequence identity was still higher than 68%, which displayed a considerable conservation among species. This result demonstrates that the protein obtained here belongs to that of GPD, which is well conserved, and shows a close evolutionary relationship in the central metabolic pathway.

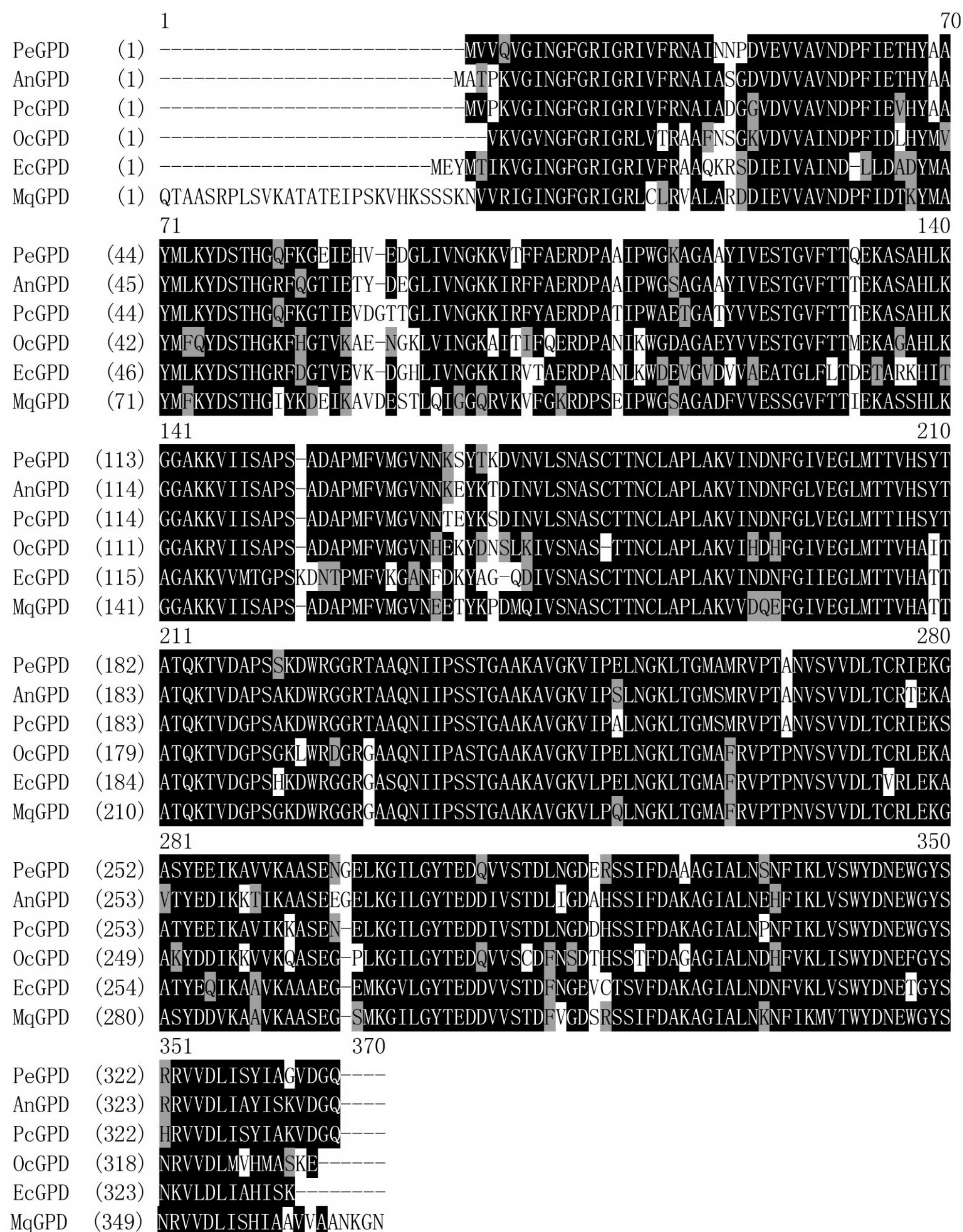

Figure 3. Mutiple sequence alignment of the *PeGPD* with other known *GPD* sequences using the CLUSTAL X software (1.81). The aligned GPDs are from *Aspergillus niger* (AnGPD, XP_001397496), *Penicillium canescens* (PcGPD, ACX54281), *Oryctolagus cuniculus* (OcGPD, 1J0X_R), *Escherichia coli* (EcGPD, NP_754078), and *Marsilea quadrifolia* (MqGPD, CAA06030). Highly conserved residues in all the sequences are indicated in white with black background and only partially conserved residues in the *GPD* sequences are shown in black with gray background.

 ©FUNPEC-RP www.funpecrp.com.br

Molecular evolution analysis

A phylogenetic tree of GPD proteins from different organisms was drawn using MEGA version 3.1, which demonstrated that GPD proteins originated from a common ancestor and later diverged into several groups (Figure 4). Interestingly, phylogenetic analysis based on the GPD proteins suggested that, in terms of evolutionary distance, *P. expansum* was closer to *Aspergillus* species but comparatively distant from *P. marneffei*, which is consistent with previous morphological observations (Malloch, 1985). Similarly, this closer relationship to *Aspergillus* was also reported in *P. chrysogenum* based on the whole genome alignment (van den Berg et al., 2008), which may indicate that the *Penicillium* species evolved from different ancestors.

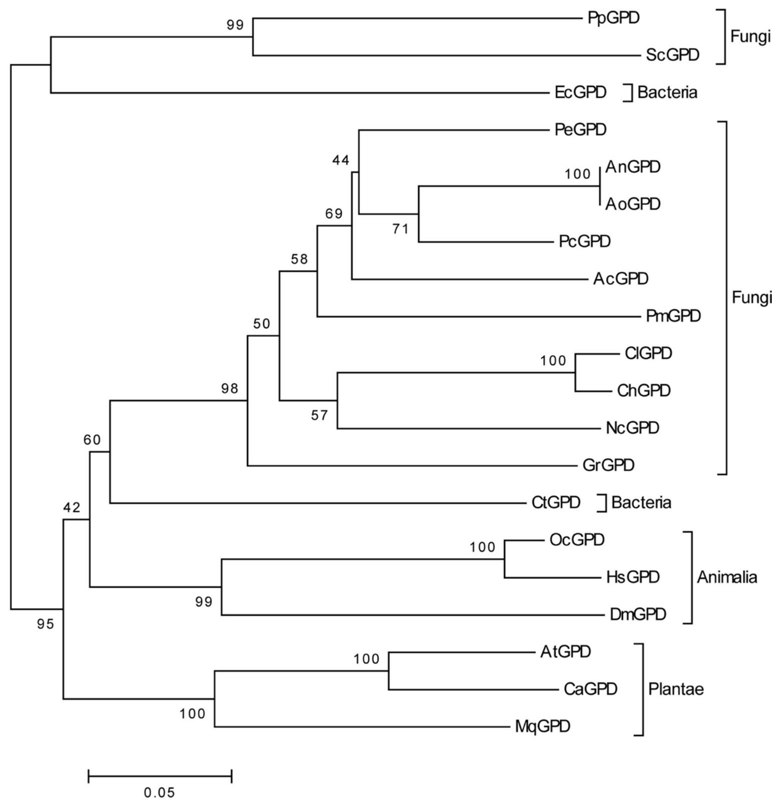

Figure 4. Phylogenetic tree showing the relationships between animal, plant, bacterial, and fungal GPD sequences. The MEGA version 3.1 software was used to construct the phylogenetic tree by neighbor-joining method. The numbers on the branches represent bootstrap support for 1000 replicates. The GPDs used in the phylogenetic tree analysis were from *Aspergillus niger* (AnGPD, XP_001397496), *A. oryzae* (AoGPD, XP_003189631), *Penicillium canescens* (PcGPD, ACX54281), *Ajellomyces capsulatus* (AcGPD, AF273703_1), *P. marneffei* (PmGPD, XP_002151104), *Curvularia lunata* (ClGPD, X58718), *Cochliobolus heterostrophus* (ChGPD, X63516), *Neurospora crassa* (NcGPD, U67457), *Galiella rufa* (GrGPD, AAT76626), *Pichia pastoris* (PpGPD, AAC49649), *Saccharomyces cerevisiae* (ScGPD, P00360), *Oryctolagus cuniculus* (OcGPD, 1J0X_R), *Homo sapiens* (HsGPD, NP_002037), *Drosophila melanogaster* (DmGPD, AAN09371), *Arabidopsis thaliana* (AtGPD, AAM67077), *Capsicum annuum* (CaGPD, CAC80377), *Clostridium thermocellum* (CtGPD, YP_001036571), *Escherichia coli* (EcGPD, NP_754078), and *Marsilea quadrifolia* (MqGPD, CAA06030).

Secondary structure prediction and homology modeling of PeGPD

Secondary structure of the PeGPD protein was predicted using the SOMPA program, which displayed α-helix (33.23%), extended strand (23.74%), β-turn (9.79%), and random coil (33.23%). The α-helix and random coil constituted interlaced domination of the main part of the secondary structure.

The homology model of the PeGPD (Figure 5A) was developed with the closest PDB model 1J0X_R of *O. cuniculus* with a 73% sequence identity to the query sequence. Usually, the sequence alignment algorithms, which are critical for the quality of 3-D models, do not make gross errors when the percentage of conserved residues is above 30% (Sengupta and Chandra, 2010). The 3-D structure, when analyzed with Ramachandran plot, showed 87.2% residues in the most favored region, 11.5% in the additionally allowed region, 1.0% in generously allowed region, and 0.3% in disallowed region, which indicated that the model is of good quality. Structural similarity of this model with 1J0X_R when analyzed using the FATCAT software showed that the 2 structures overlapped reasonably well (Figure 5B). The RMSD value of PeGPD against 1J0X_R is 0.15, which indicates that the structures are significantly similar, and hence, the structural positions of the catalytic residues can be accurately predicted from the model. When compared to the structure of 1J0X_R (Sirover, 1999; Cowan-Jacob et al., 2003), the following features have been found in PeGPD: Cys^{151} appears to be at the binding site of the substrate glyceraldehyde-3-phosphate and plays a role in creating the hemithioacetal intermediate; Lys^{185} is involved in phosphate binding to NAD^+; the residues Phe^{36} and Phe^{101} interact with the adenine ring, Asp^{34} with the adenosine ribose, Gly^{11}, Arg^{12} and Ile^{13} with the NAD^+ phosphate, and Gly^{99} and Ala^{125} with the nicotinamide ribose.

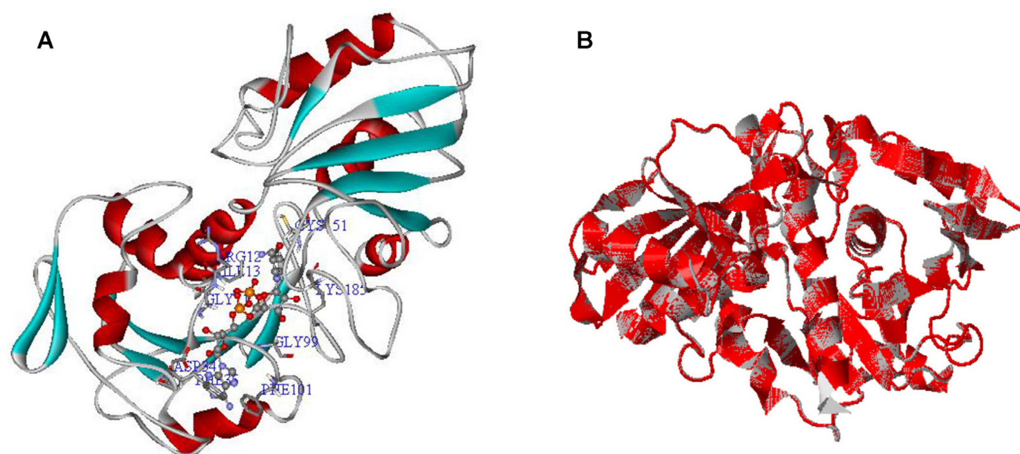

Figure 5. A. Homology model of the PeGPD (ribbon), using the crystal structure of GPD from *Oryctolagus cuniculus* (PDB: 1J0X_R) as the template. The conserved amino acids were labeled as stick and the substrate NAD^+ as scaled ball and stick. **B.** Comparison of the predicted structure of PeGPD by superposition of PeGPD (gray) onto the template structure of GPD from *O. cuniculus* (PDB: 1J0X_R) (red) using FATCAT.

 ©FUNPEC-RP www.funpecrp.com.br

Southern blot analysis

The number of genes encoding GPD varies among diverse fungal species. *GPD* have been cloned as a single-copy gene from *R. miehei* (Vastag et al., 2004) and *Candida bombicola* (Van Bogaert et al., 2008). However, *GPD* were also found to be multicopy genes in *M. circinelloides*, where 3 *GPD* copies were isolated (Wolff and Arnau, 2002). To investigate whether *PeGPD* was a single- or a multiple-copy gene, Southern blot analysis was performed under high-stringency conditions, and one clear band appeared in lanes loaded with genomic DNA samples digested with *Eco*RI, *Hin*dIII, *Kpn*I, *Dra*I, and *Bam*HI (Figure 6), which indicated a single copy of the *GPD* gene in the genome of *P. expansum* PE-12, and consequently, it is plausible that the cloned gene is functional.

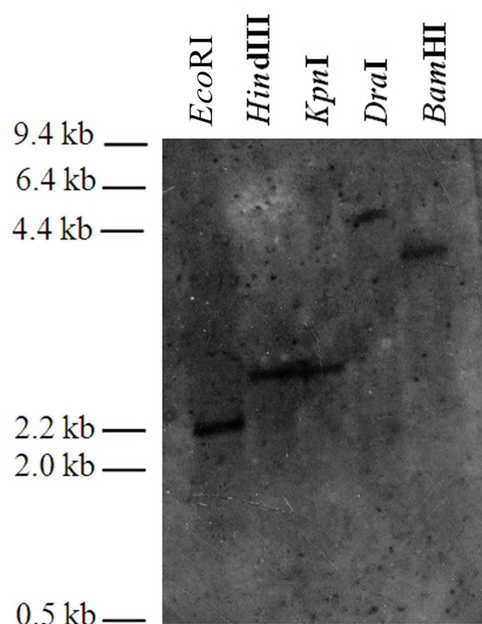

Figure 6. Southern blot analysis of *GPD* in *Penicillium expansum*. Total genomic DNA isolated from fresh mycelium of *P. expansum* PE-12 was digested overnight at 37°C with *Eco*RI, *Hin*dIII, *Kpn*I, *Dra*I, *Bam*HI, followed by hybridization with the alkaline phosphatase-labeled *PeGPD* probe.

Gene expression analysis

The transcriptional strength of *PeGPD* and *PEL* was evaluated by semi-quantitative RT-PCR. As a result (Figure 7), the expression level of *PeGPD* was much higher than that of *PEL*. Combined with the Southern blot results, we can conclude that the highly expressed *PeGPD* was not due to its gene dosage but to its own highly active promoter. Therefore, we will further investigate the corresponding cis-regulatory elements of *PeGPD*. Further cloning the promoter regions of *PeGPD* gene will promote the over-production of PEL through homologous expression in *P. expansum* PE-12.

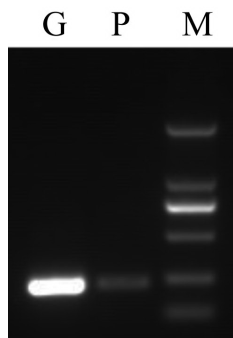

Figure 7. Expression analysis of *GPD* in *Penicillium expansum*. Total RNA (400 ng) was isolated from *P. expansum* PE-12 and subjected to RT-PCR amplification. *Lane G* = RT-PCR amplification for *GPD*; *lane P* = RT-PCR amplification for *PEL*; *lane M* = DL-2000 marker (Takara). Bands from top to bottom: 2000, 1000, 750, 500, 250, and 100 bp, respectively.

This paper reports the isolation and characterization of the gene coding for GPD of *P. expansum*. In expectation of a high and constitutive expression, the promoters of *GPD* genes have been used in various transformation vectors of different fungi (de Groot et al., 1998), among which heterologous promoters and regulatory elements are reported in several successful cases (Zhong et al., 2007; Zhang et al., 2008), but in most cases, homologous ones have been found to be more efficient (Ma et al., 2003; Kajita et al., 2004; Sugiura et al., 2009). The data presented here and the availability of the homologous expression signals of the *PeGPD* allow the construction of similar transformation vectors, which should extend our understanding of this industrially important fungus.

ACKNOWLEDGMENTS

Research supported by the China National High-Tech "863" Program, the China Ministry of Education and the Shanghai Leading Academic Discipline Project.

REFERENCES

Bian C, Yuan C, Lin L, Lin J, et al. (2005). Purification and preliminary crystallographic analysis of a *Penicillium expansum* lipase. *Biochim. Biophys. Acta* 1752: 99-102.

Cowan-Jacob SW, Kaufmann M, Anselmo AN, Stark W, et al. (2003). Structure of rabbit-muscle glyceraldehyde-3-phosphate dehydrogenase. *Acta Crystallogr. D Biol. Crystallogr.* 59: 2218-2227.

Dai D and Xia L (2005). Enhanced production of *Penicillium expansum* PED-03 lipase through control of culture conditions and application of the crude enzyme in kinetic resolution of racemic Allethrolone. *Biotechnol. Prog.* 21: 1165-1168.

de Groot MJ, Bundock P, Hooykaas PJ and Beijersbergen AG (1998). *Agrobacterium tumefaciens*-mediated transformation of filamentous fungi. *Nat. Biotechnol.* 16: 839-842.

Fei X, Zhao MW and Li YX (2006). Cloning and sequence analysis of a glyceraldehyde-3-phosphate dehydrogenase gene from *Ganoderma lucidum*. *J. Microbiol.* 44: 515-522.

Gustafsson C, Govindarajan S and Minshull J (2004). Codon bias and heterologous protein expression. *Trends Biotechnol.* 22: 346-353.

Holland MJ and Holland JP (1978). Isolation and identification of yeast messenger ribonucleic acids coding for enolase, glyceraldehyde-3-phosphate dehydrogenase, and phosphoglycerate kinase. *Biochemistry* 17: 4900-4907.

Kajita S, Sugawara S, Miyazaki Y, Nakamura M, et al. (2004). Overproduction of recombinant laccase using a homologous expression system in *Coriolus versicolor*. *Appl. Microbiol. Biotechnol.* 66: 194-199.

Kumar S, Tamura K and Nei M (2004). MEGA3: Integrated software for Molecular Evolutionary Genetics Analysis and sequence alignment. *Brief. Bioinform.* 5: 150-163.

Laskowski RA, MacArthur MW, Moss DS and Thornton JM (1993). PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Cryst.* 26: 283-291.

Li B, Rotsaert FA, Gold MH and Renganathan V (2000). Homologous expression of recombinant cellobiose dehydrogenase in *Phanerochaete chrysosporium*. *Biochem. Biophys. Res. Commun.* 270: 141-146.

Ma B, Mayfield MB and Gold MH (2003). Homologous expression of *Phanerochaete chrysosporium* manganese peroxidase, using bialaphos resistance as a dominant selectable marker. *Curr. Genet.* 43: 407-414.

Malloch D (1985). The Trichomaceae: Relationships With Other Ascomycetes. In: Advances in *Penicillium* and *Aspergillus* Systematics (Samson RA and Pitt JI, eds.). Plenum Press, New York, 365-382.

Piechaczyk M, Blanchard JM, Marty L, Dani C, et al. (1984). Post-transcriptional regulation of glyceraldehyde-3-phosphate-dehydrogenase gene expression in rat tissues. *Nucleic Acids Res.* 12: 6951-6963.

Saitou N and Nei M (1987). The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.* 4: 406-425.

Sambrook J and Russell DW (2001). Molecular Cloning. 3rd edn. Cold Spring Harbor Laboratory, New York.

Sengupta S and Chandra TS (2010). Molecular and structural characterization of GTP-cyclohydrolase II in *Eremothecium ashbyi* NRRL Y-1363: cDNA cloning, comparative sequence analysis and molecular modeling. *Fungal Biol.* 114: 731-738.

Sengupta S and Chandra TS (2011). Sequence analysis and structural characterization of a glyceraldehyde-3-phosphate dehydrogenase gene from the phytopathogenic fungus *Eremothecium ashbyi*. *Mycopathologia* 171: 123-131.

Sirover MA (1999). New insights into an old protein: the functional diversity of mammalian glyceraldehyde-3-phosphate dehydrogenase. *Biochim. Biophys. Acta* 1432: 159-184.

Stocklein W, Sztajer H, Menge U and Schmid RD (1993). Purification and properties of a lipase from *Penicillium expansum*. *Biochim. Biophys. Acta* 1168: 181-189.

Sugiura T, Yamagishi K, Kimura T, Nishida T, et al. (2009). Cloning and homologous expression of novel lignin peroxidase genes in the white-rot fungus *Phanerochaete sordida* YK-624. *Biosci. Biotechnol. Biochem.* 73: 1793-1798.

Thirach S, Cooper CR Jr and Vanittanakom N (2008). Molecular analysis of the *Penicillium marneffei* glyceraldehyde-3-phosphate dehydrogenase-encoding gene (*gpdA*) and differential expression of *gpdA* and the isocitrate lyase-encoding gene (*acuD*) upon internalization by murine macrophages. *J. Med. Microbiol.* 57: 1322-1328.

Van Bogaert IN, De Maeseneire SL, Develter D, Soetaert W, et al. (2008). Cloning and characterisation of the glyceraldehyde 3-phosphate dehydrogenase gene of *Candida bombicola* and use of its promoter. *J. Ind. Microbiol. Biotechnol.* 35: 1085-1092.

van den Berg MA, Albang R, Albermann K, Badger JH, et al. (2008). Genome sequencing and analysis of the filamentous fungus *Penicillium chrysogenum*. *Nat. Biotechnol.* 26: 1161-1168.

Vastag M, Kasza Z, Acs K, Papp T, et al. (2004). Cloning and sequence analysis of the glyceraldehyde-3-phosphate dehydrogenase gene from the zygomycetes fungus *Rhizomucor miehei*. *Antonie Van Leeuwenhoek* 86: 111-119.

Wolff AM and Arnau J (2002). Cloning of glyceraldehyde-3-phosphate dehydrogenase-encoding genes in *Mucor circinelloides* (Syn. racemosus) and use of the gpd1 promoter for recombinant protein production. *Fungal. Genet. Biol.* 35: 21-29.

Ye Y and Godzik A (2003). Flexible structure alignment by chaining aligned fragment pairs allowing twists. *Bioinformatics.* (Suppl 2) 19: ii246-ii255.

Yuan C, Lin L, Shi QQ and Wu SG (2003). Overexpression of *Penicillium expansum* lipase gene in *Pichia pastoris*. *Sheng Wu Gong Cheng Xue Bao* 19: 231-235.

Zhang WX, Jiang YM, Zou XL, Chen B, et al. (2001). Breeding of alkaline lipase overproducing strain by screening resistant mutant. *Chin. J. Ind. Microbiol.* 31: 14-16.

Zhang P, Xu B, Wang Y, Li Y, et al. (2008). *Agrobacterium tumefaciens*-mediated transformation as a tool for insertional mutagenesis in the fungus *Penicillium marneffei*. *Mycol. Res.* 112: 943-949.

Zhang Z, Yang J, Xu L, Liu Y, et al. (2010). Cloning, codon optimization and expression of mature lipase gene *Penicillium expansum*. *Wei Sheng Wu Xue Bao* 50: 228-235.

Zhong YH, Wang XL, Wang TH and Jiang Q (2007). *Agrobacterium*-mediated transformation (AMT) of *Trichoderma reesei* as an efficient tool for random insertional mutagenesis. *Appl. Microbiol. Biotechnol.* 73: 1348-1354.