J Ind Microbiol Biotechnol (2008) 35:1085–1092
DOI 10.1007/s10295-008-0386-x

ORIGINAL PAPER

# Cloning and characterisation of the glyceraldehyde 3-phosphate dehydrogenase gene of *Candida bombicola* and use of its promoter

**Inge N. A. Van Bogaert · Sofie L. De Maeseneire · Dirk Develter · Wim Soetaert · Erick J. Vandamme**

Received: 28 April 2008 / Accepted: 4 June 2008 / Published online: 2 July 2008


**Abstract** The glyceraldehyde-3-phosphate dehydrogenase gene (GPD) of the sophorolipid producing yeast *Candida bombicola* was isolated using degenerated PCR and genome walking. The obtained 3,740 bp contain the 1,008 bases of the coding sequence and 1,613 and 783 bp of the upstream and downstream regions, respectively. The corresponding protein shows high homology to the other known *GPD* genes and is 74% identical to the gyceraldehyde-3-phosphate dehydrogenase of *Yarrowia lipolytica*. The particular interest in the *C. bombicola GPD* gene sequence originates from the potential use of its promoter for high and constitutive expression of homologous and heterologous genes. Southern blot analysis did not give any indication for the presence of multiple *GPD* genes and it can therefore be expected that the promoter can be used for efficient and high expression. This hypothesis was further confirmed by the biased codon usage in the *GPD* gene. *GDP* promoter fragments of different lengths were used to construct hygromycin resistance cassettes. The constructs were used for the transformation of *C. bombicola* and all of them, even the ones with only 190 bp of the *GPD* promoter, were able to render the cells resistant to hygromycin. The efficacy of a short *GPD* promoter can be a convenient characteristic for the construction of compact expression cassettes or vectors for *C. bombicola*. The GenBank accession number of the sequence described in this article is *EU315245*.

**Keywords** *Candida bombicola* · Sophorolipids · Glyceraldehyde 3-phosphate dehydrogenase · Promoter · Hygromycin resistance

I. N. A. Van Bogaert (✉) · S. L. De Maeseneire · W. Soetaert · E. J. Vandamme
Laboratory of Industrial Microbiology and Biocatalysis, Department of Biochemical and Microbial Technology, Faculty of Bioscience Engineering, Ghent University, Coupure Links 653, 9000 Ghent, Belgium
e-mail: Inge.VanBogaert@UGent.be

D. Develter
Ecover Belgium NV, Industrieweg 3, 2390 Malle, Belgium

## Introduction

Since many years, the non-pathogenic yeast *Candida bombicola* is known to produce sophorolipids [1]. These surface-active compounds are synthesized at quite high levels and find applications in the food, pharmaceutical, cosmetic, and cleaning industries (reviewed in [2]). In order to get better insights into the sophorolipid synthesis pathway and to open up perspectives for improved or altered production, a transformation and selection system for *C. bombicola* was developed [3]. One of the logical next steps in genetic engineering of this yeast species is the construction of an expression system.

Glyceraldehyde-3-phosphate dehydrogenase (GPD, E.C. 1.2.1.12) is one of the key enzymes in the Embden–Meyerhof–Parnas or glycolysis pathway. It catalyzes the reversible oxidation and phosphorylation of glyceraldehyde-3-phosphate to produce 1,3-diphosphoglycerate. The enzyme contributes in this way to the formation of ATP and provides additional energy to the cell by reducing $NAD^+$ to NADH and $H^+$ upon its action. GPD is a tetrameric enzyme and typically does not follow simple Michaelis–Menten kinetics [4]. Since GPD is a crucial enzyme functioning in a pathway which is essential in every living cell, its amino acid sequence is highly conserved. Meanwhile, *GPD* sequences of a wide range of organisms have been determined, and are an important tool in phylogenetic analysis and protein evolution.

Due to its key function in glycolysis, GPD can represent up to 5% of the soluble cellular protein content of *Saccharomyces cerevisiae*, *Aspergillus nidulans* and other (higher) eukaryotic organisms [5, 6]. The abundance of the GPD protein suggests that the *GPD* gene is regulated by a constitutively and highly active promoter. The promoter sequences of native *GPD* genes have been successfully applied for the expression of heterologous genes in several yeasts and filamentous fungi such as *S. cerevisiae* [7], *Pichia pastoris* [8], *Lentinula edodes* [9], *Mucor circinelloides* [10] and *Flammulina velutipes* [11].

Because of the high conservation degree of the coding sequence of the *GPD* gene, it should be possible to clone part of the glyceraldehyde-3-phosphate dehydrogenase gene of *C. bombicola* by degenerate PCR. Upstream genome walking should then result in achievement of the promoter sequence.

## Materials and methods

### Strains, plasmids and culture conditions

*Candida bombicola* ATCC 22214 was used for the preparation of genomic DNA. *Escherichia coli* DH5α was used to maintain plasmids.

All PCR products were cloned into the pGEM-T® vector (Promega). The plasmid pREP4FLAG (LMBP 4343) was obtained from the Laboratory of Molecular Biology, Ghent University (LMBP).

*Candida bombicola* was cultured in medium containing 10% glucose, 1% yeast extract and 0.1% urea. Liquid yeast cultures were incubated at 30 °C and 200 rpm. *E. coli* was grown in Luria–Bertani (LB) medium (1% tryptone, 0.5% yeast extract and 0.5% sodium chloride) supplemented with 100 mg/L ampicillin and 40 mg/L X-gal if necessary. Liquid *E. coli* cultures were incubated at 37 °C and 200 rpm.

### DNA isolation and sequencing

Yeast genomic DNA was isolated with the GenElute™ Bacterial Genomic DNA Kit (Sigma), but cell lysis was performed by incubation at 30 °C during 90 min with zymolyase (Sigma). Plasmid DNA was isolated with the QIAprep Spin Miniprep Kit (Qiagen). All DNA sequences were determined at the VIB Genetic Service Facility (Belgium).

### Transformation

*Candida bombicola* cells were transformed with the lithium acetate method [12], but 50 mM LiAc was used instead of 100. Transformants were selected on YPD plates containing 500 mg/mL hygromycin.

*E. coli* cells were transformed as described by Sambrook and Russell [13].

### Degenerate PCR

Part of the *GPD* gene of *C. bombicola* was amplified by the use of the degenerate primers GpdDegFor and GpdDegRev (Table 1) based on the conserved sequences of the *GPD* genes of the *Saccharomycotina*. PCR amplification was carried out with an initial denaturation of 94 °C for 4 min, a 40-fold repeat of 94 °C for 30 s, 50 °C for 1 min, 72 °C for 2 min with 5 s increasing time and a final 7 min elongation at 72 °C.

### Genome walking

The unknown genomic DNA sequences upstream and downstream of the degenerate PCR fragment were identified by genome walking. Those walks were carried out according to the user's manual of the BD GenomeWalker™ Universal Kit (BD Biosciences). For both the upstream and downstream sequence, two PCR amplifications—a primary followed by a nested PCR—were performed. Four gene-specific primers (GSP) were designed and are listed in Table 1. The PCR reaction mixture and cycles were optimized for use with the Expand Long Template PCR System (Roche Diagnostics), as described by De Maeseneire et al. [14].

### Yeast colony PCR

Integration of the hygromycin resistance cassette was tested by yeast colony PCR with the primers GpdHygro190For and HygroInsertCheckRev (Table 1). The temperature program consists of an initial denaturation of 94 °C for 7 min, a 30-fold repeat of 94 °C for 30 s, an annealing temperature of 64 °C for 30 s, 72 °C for 2.5 min with 5 s increasing time and a final 7 min elongation at 72 °C.

### Sequence analysis

Sequences were analysed with the Clone Manager Professional Suite Software (Version 6.0). The BLAST program [15] was used for similarity searches in databases available on the NCBI website (http://www.ncbi.nlm.nih.gov). Multiple sequence alignments were made with the CLUSTAL W program [16]. Subsequently, a phylogenetic tree was constructed in BioEdit using the Protein Maximum Likelihood (ProML) algorithm.

The Effective Number of Codons (Nc) was determined with the EMBOSS-program CHIPS at http://emboss.bioinformatics.nl/cgi-bin/emboss/chips.

### Southern blot

A total of 20 μg digested gDNA was blotted onto a positively charged nylon filter (Hybond N+, GE Healthcare)

**Table 1** Primers used for the isolation of the *Candida bombicola GPD* gene, creation of the Southern blot probe and the expression cassette

| Name | Feature | Sequence |
|---|---|---|
| GpdDegFor | Degenerated primer | GTTYAARTAYGAYTCYACYCAYGG |
| GpdDegRev | Degenerated primer | GAGTARCCRWAYTCRTTRTCRTACC |
| GpdUp1 | Primary upstream GPS | ATCAGCGGAAGGAGCAGAGATCACAA |
| GpdUpN | Nested upstream GPS | CCTTAGCCTTCTCGGTGGTGGTGAAGAC |
| GpdDown1 | Primary downstream GPS | TCTGCGGTGACACTCACTCCTCTATCTAC |
| GpdDownN | Nested downstream GPS | ACGGCAACTTCGTCAAGCTCATCTCC |
| GpdProbeFor | Creation Southern blot probe | GACTATGCTGCCTACATGTTCAAG |
| GpdProbeRev | Creation Southern blot probe | GAGTAGCCGTACTCGTTGTCGTACCA |
| HygroCdsFor | Amplification hygromycin gene | TCTAGATTCTCGAGATGAAAAAGCCTGAACTCAC |
| HygroCdsRev | Amplification hygromycin gene | TGAACAAACGACCCAACAC |
| GpdPromInfFor | Amplification *GPD* promoter | CCGCGGGATTTCTAGGACATCCGATGTGTAGTTAATCA |
| GpdPromInfRev | Amplification *GPD* promoter | GCTTTTTCATCTCGATTGTGTAGAGTTGTTTTTGTTG |
| GpdHygro488For | Amplification expression cassette | TCAAGCACTGCGGCACTCCTA |
| GpdHygro394For | Amplification expression cassette | GCTGGAGTCTCATCTGCAAGGTT |
| GpdHygro352For | Amplification expression cassette | AATACGAGCAATCGAAGCCTTGG |
| GpdHygro190For | Amplification expression cassette | CCGTCATTGCAGGGTGTGCGTCTA |
| GpdHygroRev | Amplification expression cassette | ACCCAACACCCGTGCGTTT |
| HygroInsertCheckRev | Presence expression cassette | GCGATTTGTGTACGCCCGACAGT |

All primers were obtained from Sigma Genosys

with the Trans-Blot® SD DNA/RNA Blotting Kit (Bio-Rad) according to the manufacturer's instructions. DNA was hybridised to a DIG labelled probe which was constructed with the DIG High Prime Labelling and Detection Kit (Roche Diagnostics). The applied hybridisation conditions allowed 18% mismatch between probe and target. The initial unlabelled probe of 842 bp was obtained by means of the High Fidelity PCR Master kit (Roche Diagnostics) and the primers GpdProbeFor and GpdProbeRev (Table 1). PCR products were purified by using the GenElute™ PCR CleanUp Kit (Sigma). Restriction enzymes were obtained from Roche Diagnostics or New England Biolabs and were used as indicated by the manufacturer.

### Construction of the GPD-hygromycin resistance cassette

The potency of the *GPD* promoter was tested by fusing it to a hygromycin B phosphotransferase gene. The hygromycin B phosphotransferase gene and the TK terminator were obtained by high fidelity PCR amplification with BD Advantage 2 Polymerase Mix (Clontech) and the primers HygroCdsFor and HygroCdsRev and pREP4FLAG as template. The forward primer HygroCdsFor was extended with the *Xba*I (TCTAGA) and *Xho*I (CTCGAG) restriction sites for convenient cloning of the *GPD* promoter fragment (Table 1). The obtained fragment of the hygromycin B phosphotransferase gene and the TK terminator (1,154 bp) was cloned into pGEM-T® (Promega). A vector with a T7 oriented insert was chosen to continue cloning. This vector was cut with the unique cutters *Xba*I and *Xho*I, located just upstream the hygromycin resistance gene, and the *GPD* promoter of *C. bombicola* was inserted by means of the In-Fusion™ 2.0 Dry-Down PCR Cloning Kit (Clontech). This *GPD* promoter was obtained by high fidelity PCR amplification with the BD Advantage 2 Polymerase Mix (Clontech) and the primers GpdPromInfFor and GpdPromInfRev, amplifying 1,560 bp of the *GPD* promoter. The obtained vector harbours the hygromycin resistance gene controlled by the *C. bombicola GPD* promoter.

For obtaining linear hygromycin resistance cassettes under the control of a *GPD* promoter with a length of 488, 394, 352 and 190 bp, respectively, the primers GpdHygro488For, GpdHygro394For, GpdHygro352For and GpdHygro190For were combined with GpdHygroRev (Table 1).

### Nucleotide sequence accession number

The nucleotide sequence described in this paper has been deposited at the GenBank nucleotide database under the accession number *EU315245*.

## Results and discussion

### Isolation of the *C. bombicola GPD* gene

Since *GPD* genes are known to be highly conserved among species, it is quite easy to create good degenerate primers

for the amplification of uncharacterized *GPD* genes. The degenerate primers GpdDegFor and GpdDegRev (Table 1) were designed based on the conserved regions of the *GPD* genes of the *Saccharomycotina* and generated a single fragment of approximately 820 bp, which was in the expected range of size. The nucleotide sequence of the fragment was determined and subsequent blasting against the NCBI databases revealed a 74% amino acid identity with the glyceraldehyde-3-phosphate dehydrogenase of *Y. lipolytica*.

The regions adjacent to the obtained fragment were cloned using the Universal GenomeWalker kit™. As a result, a DNA sequence of 3,740 bp was obtained, carrying the total coding sequence of 1,008 bp and an upstream and downstream region of 1,613 and 783 bp, respectively.

## Characterisation of the *GPD* sequence

Since the *GPD* gene is a highly expressed gene, it can be expected that it has at least some promoter elements in its 5′ region. Two possible CAAT boxes can be found 366 and 354 bp upstream the ATG start codon (Fig. 1). No G-residues were found in the seven bases preceding the start

**Fig. 1** Nucleotide and deduced amino acid sequence of the *GPD* gene of *C. bombicola*. Possible promoter elements are *underlined* and sequences presumed to be involved in polyadenylation are marked in *bold cases*. The conserved active site amino acids are marked as *bold* and *underlined* cases. The beginning of the putative ARL1 gene is also indicated

```
   1  tttttcaggt gccacacgcg cattaatcga gcaagaaggt acccttggta gacgtgcata
  61  catgcgagtt gacgtcgtaa cgtcggcacg ctcgtaacgc tcgcgccccc tctcaagcac
 121  tgcggcactc ctaccgttag gacttaggct tagggctatg cttgtgcgcg ccgcgaaacg
 181  gttagtataa gcagtacgac gttgttgctg gagtctcatc tgcaaggttg agtaccaatc
 241  cctgcccccaa tacgagcaat cgaagccttg gggaaagatg cggcgggcta gcttcagcaa
 301  taaatagcag gcgacacaca aaaattaggc ggcaagcgca cgctcagcat gccatctacc
 361  agggcaaaaa gcaaggcaac ctctttttcg catcccgatt tagagcctac ccgtcattgc
 421  agggtgtgcg tctacgatat aacacgatcg acatcgcgct gggtatgctt ctgggtaagg
 481  ggtcgcaacg tgtgagttgt cagcactggc cgatacccaa agtatataat gcgccgttga
 541  acggttatag tcggtcaagc tcttaaagaa agacttaaca acaaaaacaa ctctacacaa
 601  atggctggat tcactgtcgg tatcaacggt ttcggacgca ttggtcgtct cgttcttcgc
   1    M  A  G   F  T  V   G  I  N  G   F  G  R   I  G  R   L  V  L  R
 661  aacgctctcg agaacagcaa cgtcaacgtc gttgccatca acgatccgtt catcgctccg
  21    N  A  L   E  N  S   N  V  N  V   V  A  I   N  D  P   F  I  A  P
 721  gactatgctg cctacatgtt caagtacgac tctactcacg gccgcttcaa gggcgatgtc
  41    D  Y  A   A  Y  M   F  K  Y  D   S  T  H   G  R  F   K  G  D  V
 781  caggctgcct ctgacggtat tgtcatcaac ggcaagaagg ttcttgtctt caacgagaag
  61    Q  A  A   S  D  G   I  V  I  N   G  K  K   V  L  V   F  N  E  K
 841  gacccggccc agatcccctg gggcaagagc ggcgtttctt acgttgttga gtccactggt
  81    D  P  A   Q  I  P   W  G  K  S   G  V  S   Y  V  V   E  S  T  G
 901  gtcttcacca ccaccgagaa ggctaaggct cacctcgctg gcggcgccaa gcgtgttgtg
 101    V  F  T   T  T  E   K  A  K  A   H  L  A   G  G  A   K  R  V  V
 961  atctctgctc cttccgctga tgcccctatg tacgttgttg gtgtcaacct tgagaagtac
 121    I  S  A   P  S  A   D  A  P  M   Y  V  V   G  V  N   L  E  K  Y
1021  tctcccaagg accagatcgt ctccaacgct tcgtgcacca ccaactgcct tgctcccctc
 141    S  P  K   D  Q  I   V  S  N  A   S  C  T   T  N  C   L  A  P  L
1081  gccaaggtca tccacgacga gttcggtctc gttgagggtc tcatgactac tgtccactcc
 161    A  K  V   I  H  D   E  F  G  L   V  E  G   L  M  T   T  V  H  S
1141  tacactgcca cccagaagac cgttgacgga ccttcccaca aggactggcg tggaggccgt
 181    Y  T  A   T  Q  K   T  V  D  G   P  S  H   K  D  W   R  G  G  R
1201  actgccgctg ccaacatcat cccgtcttcg actggtgctg ccaaggctgt cggcaaggtc
 201    T  A  A   A  N  I   I  P  S  S   T  G  A   A  K  A   V  G  K  V
1261  atccctgagc tcaacggcaa gctcaccggt atctctgtcc gtgtccctac ccctgacgtc
 221    I  P  E   L  N  G   K  L  T  G   I  S  V   R  V  P   T  P  D  V
1321  tctgttgtcg acctcactgc tcgtctccag aagagcgcca ccattgagga gatcaacgct
 241    S  V  V   D  L  T   A  R  L  Q   K  S  A   T  I  E   E  I  N  A
1381  accatcaaga agtactctga gggccagatg aagggtgtgc tcggctacac tgacgaggac
 261    T  I  K   K  Y  S   E  G  Q  M   K  G  V   L  G  Y   T  D  E  D
1441  gtcgtctcca gcgacttctg cggtgacact cactcctcta tctacgacgc caaggcttcc
 281    V  V  S   S  D  F   C  G  D  T   H  S  S   I  Y  D   A  K  A  S
1501  ctcgccctca acggcaactt cgtcaagctc atctcctggt acgacaacga gtacggctac
 301  L  A  L   N  G  N   F  V  K  L   I  S  W   Y  D  N   E  Y  G  Y
1561  tctgcccgtg tcgtcgactt gcttgtcgct atcgctaaga aggactaaat ttctcctaat
 321    S  A  R   V  V  D   L  L  V  A   I  A  K   K  D  *  ← GPD
1621  aggctgtcag cgcatatgtg aggcgctcat ataaaacaat ataaatcaaa acccatgtta
1681  aaaacttgtt gatcccagca cttttgagaa gcgcactccg aactaaatct aaaaacactt
1741  cagcttaagc tattattgcc tgattctcgt catatcgctg gggcccgagc gacgcacgat
1801  ccaataccgc gttgggatta ctcagcacct tgcaagatcg acgctcgctc gtgctacact
1861  gttcatctat ctgtagctgt gagtagccat gggaaacacc ttcagctcgc ttctaggcaa
                  putative ARL1 →  M  G  N  T   F  S  S   L  L  G
```

codon, which is a typical feature for highly expressed genes of *S. cerevisiae* [17]. Furthermore, a conventional TATA-box, TATATAAT, is present at position −78.

The polyadenylation signal in yeasts can differ from the signal found in eukaryotic genes (AATAAA), as described by Graber et al. [18]. The 3′ region of the *C. bombicola GPD* gene contained two possible sequences: ATAAAA 43 and AATATA 50 bp downstream the TAA stop codon.

One feature often seen for highly expressed genes is a biased codon usage. This phenomenon is also observed for the *GPD* gene of *C. bombicola*: there is a preference for use of a pyrimidine at the third position (80.35%) and when the choice is allowed between a pyrimidine and a purine, a pyrimidine is chosen in 94.40% of the cases. Twenty-four of the 61 possible sense codons are unused (Table 2).

The codon bias can be determined more accurate by means of the Effective Number of Codons (Nc). Nc indicates to which extent a certain codon is used in a gene. The Nc varies from 20 for a highly biased gene to 61 for a gene without bias. The calculated value for the discussed *GPD* is 31.09. To compare, the Nc-values of all other *C. bombicola* genes isolated by our research group are situated between 47.70 and 58.36, which confirms the high bias of the *GPD* gene and suggests a strong promoter activity and high expression level.

As to the other known genes of *C. bombicola*, there were no indications for the presence of an intron [19, 20]. The *GPD* open reading frame is translated in a protein of 335 amino acids with a calculated size of 35.71 kDa and an estimated *PI* value of 7.67. The protein is most identical to the glyceraldehyde-3-phosphate dehydrogenase of *Y. lipolytica* (74% amino acid identity) and furthermore shows high homologies with GPD's from both ascomycetous, basidiomycetous and zygomycetous fungi. The identities between the *C. bombicola* GPD and those of lower and higher *eukaryota* are in general above 65%, illustrating the high conservation grade of the GPD sequence among organisms. Also for bacteria, identities between 59 and 68% were observed. Figure 2 shows an alignment between the *C. bombicola* GPD and representatives of different taxonomic classes. Figure 2 reveals that the enzyme contains the GPD consensus substrate binding region found in all GPD enzymes (PROSITE): [ASV]-S-C-[NT]-T-x(2)-(LIM). The cysteine residue is essential for the enzymatic activity since it functions as the binding site in the catalytic region [21]. This pattern is present in the *C. bombicola* protein as ASCTTNCL. Other conserved amino acids are a histidine, and two phosphate binding residues: lysine and arginine, in the *C. bombicola* sequence found at positions 179, 190 and 234 respectively, and indicated in Fig. 2 [22].

## Other putative genes flanking *GPD*

An open reading frame (ORF) of 549 bp is located 281 bp downstream the *GPD* stop codon (Fig. 1). There is even a possible CAAT-box present 80 bp upstream the initiation codon of this ORF. The putative protein shows a 69% amino acid identity to the ADP-ribosylation Factor-like 1 Protein (ARL1) from *S. cerevisiae*. ARL1 is located in the Golgi complex, where it is believed to play a role in regulation of membrane traffic.

**Table 2** Codon usage for the *C. bombicola GPD* gene (336 codons)

| AA | Codon | # | % | AA | Codon | # | % | AA | Codon | # | % | AA | Codon | # | % |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Ala | GCA | 0 | 0 | His | CAC | 6 | 100 | Pro | CCA | 0 | 0 | Ser | AGC | 4 | 15.4 |
| | GCC | 15 | 42.9 | | CAT | 0 | 0 | | CCC | 3 | 23.1 | | AGT | 0 | 0 |
| | GCG | 0 | 0 | Ile | ATA | 0 | 0 | | CCG | 4 | 30.8 | | TCA | 0 | 0 |
| | GCT | 20 | 57.1 | | ATC | 17 | 85 | | CCT | 6 | 46.2 | | TCC | 9 | 34.6 |
| Cys | TGC | 3 | 100 | | ATT | 3 | 15 | Gln | CAA | 0 | 0 | | TCG | 2 | 7.7 |
| | TGT | 0 | 0 | Lys | AAA | 0 | 0 | | CAG | 6 | 100 | | TCT | 11 | 42.3 |
| Asp | GAC | 18 | 85.7 | | AAG | 25 | 100 | Arg | AGA | 0 | 0 | Val | GTA | 0 | 0 |
| | GAT | 3 | 14.3 | Leu | CTA | 0 | 0 | | AGG | 0 | 0 | | GTC | 23 | 62.2 |
| Glu | GAA | 0 | 0 | | CTC | 14 | 70 | | CGA | 0 | 0 | | GTG | 2 | 5.4 |
| | GAG | 13 | 100 | | CTG | 0 | 0 | | CGC | 3 | 30 | | GTT | 12 | 32.4 |
| Phe | TTC | 10 | 100 | | CTT | 5 | 25 | | CGG | 0 | 0 | Trp | TGG | 3 | 100 |
| | TTT | 0 | 0 | | TTA | 0 | 0 | | CGT | 7 | 70 | Tyr | TAC | 12 | 92.3 |
| Gly | GGA | 4 | 13.3 | | TTG | 1 | 5 | Thr | ACA | 0 | 0 | | TAT | 1 | 7.7 |
| | GGC | 14 | 46.7 | et | ATG | 5 | 100 | | ACC | 11 | 50 | Stop | TAA | 1 | 100 |
| | GGG | 0 | 0 | Asn | AAC | 17 | 100 | | ACG | 0 | 0 | | TAG | 0 | 0 |
| | GGT | 12 | 40 | | AAT | 0 | 0 | | ACT | 11 | 50 | | TGA | 0 | 0 |

The absolute (#) and relative (%) numbers for each synonymous codon are given

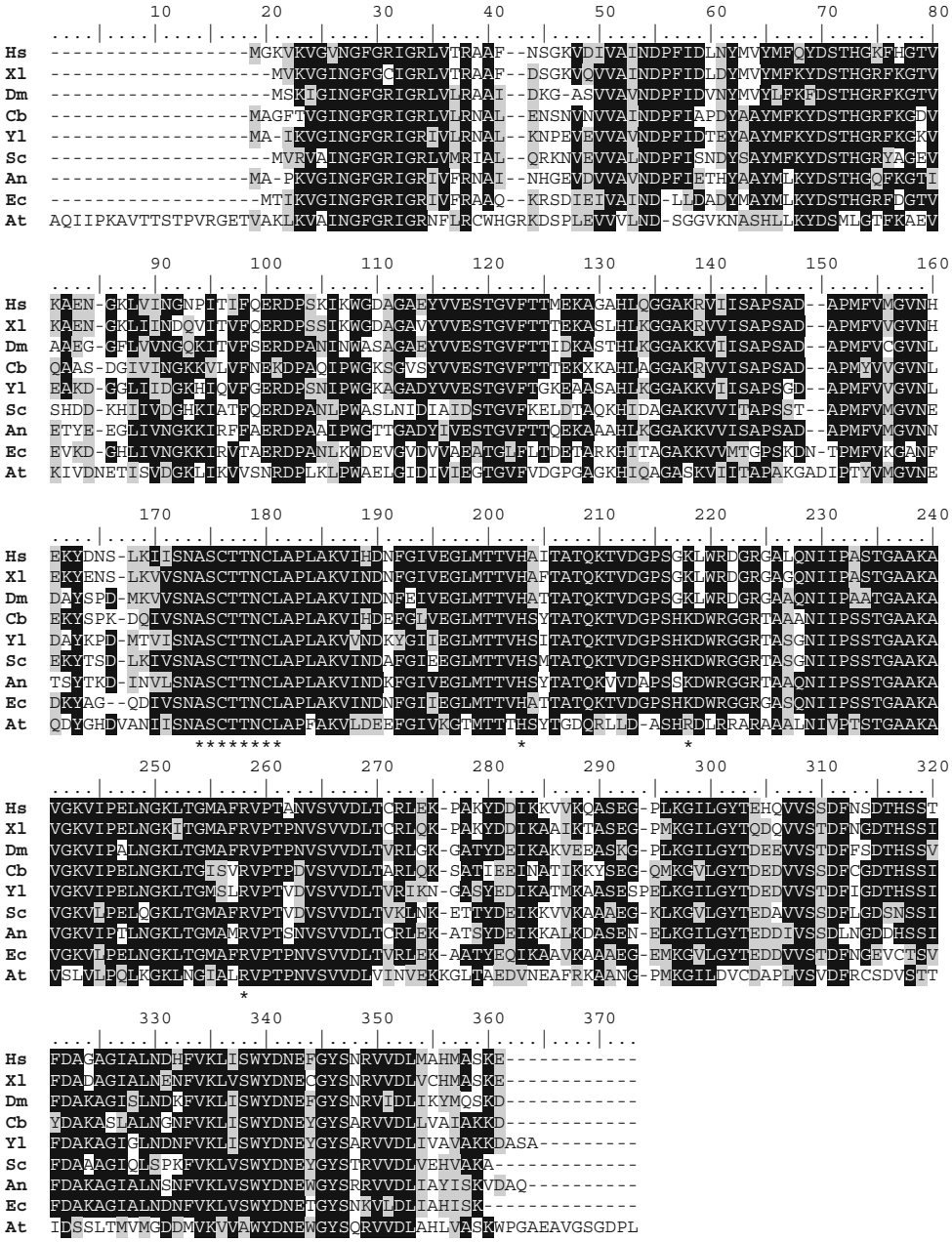


**Fig. 2** Alignment of GPD sequences of organisms from different taxonomic classes: *Homo sapiens* (*Hs*), *Xenopus laevis* (*Xl*), *Drosophila melanogaster* (*Dm*), *C. bombicola* (*Cb*), *Y. lipolytica* (*Yl*), *S. cerevisiae* (*Sc*), *Aspergillus niger* (*An*), *E. coli* (*Ec*) and *Arabidopsis thaliana* (*At*). The substrate binding region and other conserved amino acids are marked with stars underneath the sequence alignment

Determination of the *GPD* copy number

In most yeast genomes, only one *GPD* gene is present. For some others, such as *S. cerevisiae*, two or three slightly different *GPD* genes are differentially expressed, suggesting a distinct cellular role [23]. In order to verify this for *C. bombicola*, a Southern blot was performed. *C. bombicola* genomic DNA was cut with three different restriction enzymes: *Ava*I, *Eco*RI and *Xho*I. Only *Ava*I cuts twice into the known sequence, and binding of the probe should visualize a fragment of 1,117 bp. All other enzymes do not cut the known sequence and accordingly should yield fragments larger than 3,740 bp. The pattern of the Southern blot matched these predictions and no other bands were observed (Fig. 3). These findings indicate that only a single copy of the *GPD* gene is present in the genome of *C. bombicola* and therefore it is plausible that the cloned gene is a functional one.

Evaluation of the *C. bombicola GPD* promoter

The usefulness of the *GPD* promoter for heterologous expression in *C. bombicola* was tested with the hygromycin resistance gene. An expression cassette containing the *GPD* 5′ region, the coding sequence of the hygromycin resistance gene (hygromycin phosphotransferase) and the *TK* terminator was constructed as described in the "Materials and methods". Different promoter lengths were tested to

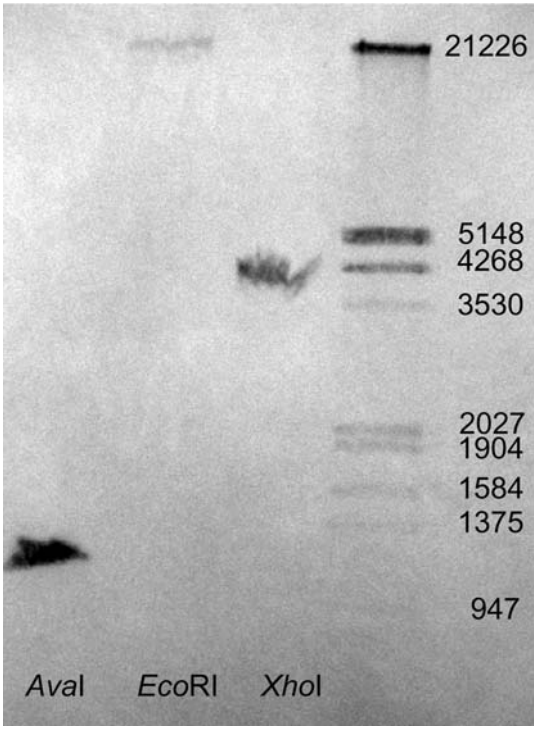


**Fig. 3** Southern blot with a *GPD* probe based on the obtained sequence. *C. bombicola* gDNA was cut with *Ava*I, *Eco*RI and *Xho*I. The DIG-labelled DNA Moleclar-Weight Marker III from Roche was run alongside the gDNA. The size of its fragments are indicated on the *right*

find out whether a certain length should be respected in order to get sufficient expression. Promoter parts of 488, 394, 352 and 190 bp were tested. All fragments contained the TATA-box found at −78, and the fragments larger than 352 bp also possessed both CAAT-sequences located at positions −366 and −354. Consequently, the fragment of 352 bp harboured only one CAAT-sequence, while the 190 fragment had none.

All used expression cassettes gave rise to hygromycin resistant transformants. The presence of the transformation cassette into the genome was verified by yeast colony PCR with the primers GpdHygro190For and hygroInsertCheck-Rev (Table 1). All PCR reactions yielded the expected fragment of 1,124 bp.

However, the activity of the promoter fragment of 190 bp is somewhat surprising. For example, it has been shown that further upstream regions of the fungal *Lentinula edodes GPD* promoter are essential for expression [9]. Also for *S. cerevisiae*, larger promoter sequences are used for *GPD* controlled expression. Yet, the most essential sequences of the human *GPD* promoter are located in the region −181 to −144, and a promoter fragment of 181 bp gave an almost identical $\beta$-galactosidase activity as compared to a 489 bp fragment [24]. The efficacy of a short *GPD* promoter can be a convenient characteristic for the construction of compact expression cassettes or vectors for *C. bombicola*. However, the exact expression level of the different *GPD* promoter fragments should first be investigated by e.g. expressing $\beta$-galactosidase and quantifying its activity.

**Acknowledgments** The authors wish to thank Ecover Belgium NV and the Bijzonder Onderzoekfonds of Ghent University for financial support (grants A05/003 and 01D18604).

## References

1. Gorin PAJ, Spencer JFT, Tulloch AP (1961) Hydroxy fatty acid glycosides of sophorose from *Torulopsis magnoliae*. Can J Chem 39:846–855. doi:10.1139/v61-104
2. Van Bogaert INA, Saerens K, De Muynck C, Develter D, Soetaert W, Vandamme EJ (2007) Microbial production and application of sophorolipids. Appl Microbiol Biotechnol 76:23–34. doi:10.1007/s00253-007-0988-7
3. Van Bogaert INA, De Maeseneire SLD, Develter D, Soetaert W, Vandamme EJ (2008) Development of a transformation and selection system for the glycolipid producing yeast *Candida bombicola*. Yeast 25:273–278. doi:10.1002/yea.1586
4. Harris JI, Waters M (1976) Glyceraldehyde-3-posphate dehydrogenase. In: Boyer PD (ed) The enzymes. Academic Press, New York, pp 1–49
5. Piechaczyk M, Blanchard JM, Marty L, Dani C, Panabieres F, El Sabouty S, Fort P, Jeanteur P (1984) Post-transcriptional regulation of glyceraldehyde-3-phosphate dehydrogenase gene expression in rat tissues. Nucleic Acids Res 12:6951–6963. doi:10.1093/nar/12.18.6951
6. Punt PJ, Dingemanse MA, Kuyvenhoven A, Soede RD, Pouwels PH, Van den Hondel CA (1990) Functional elements in the promoter region of the *Aspergillus nidulans gpdA* gene encoding glyceraldehyde-3-phosphate dehydrogenase. Gene 93:101–109. doi:10.1016/0378-1119(90)90142-E
7. Bitter GA, Egan KM (1984) Expression of heterologous genes in *Saccharomyces cerevisiae* from vectors utilizing the glyceraldehyde-3-phorsphate dehydrogenase gene promoter. Gene 32:263–274. doi:10.1016/0378-1119(84)90002-7
8. Doring F, Klapper M, Theis S, Daniel H (1998) Use of the glyceraldehyde-3-phosphate dehydrogenase promotor for production of functional mammalian membrane transport proteins in the yeast *Pichia pastoris*. Biochem Biophys Res Commun 250:531–535. doi:10.1006/bbrc.1998.9342
9. Hirano T, Sato T, Yaegashi K, Enei H (2000) Efficient transformation of the edible basidiomycete *Lentinus edodes* with a vector using a glyceraldehyde-3-phosphate dehydrogenase promoter to hygromycin B resistance. Mol Gen Genet 263:1047–1052. doi:10.1007/s004380050033
10. Wolff AM, 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. doi:10.1006/fgbi.2001.1313
11. Kuo CY, Chou SY, Huang CT (2004) Cloning of glyceraldehyde-3-phosphate dehydrogenase gene and use of the *gpd* promoter for transformation in *Flammulina velutipes*. Appl Microbiol Biotechnol 65:593–599. doi:10.1007/s00253-004-1635-1
12. Gietz RD, Schiestl RH (1995) Transforming yeast with DNA. Methods Mol Cell Biol 5:255–269
13. Sambrook J, Russell DW (2001) Molecular cloning: a laboratory manual. Cold Spring Harbour, New York
14. De Maeseneire SL, De Groeve MRM, Dauvrin T, De Mey M, Soetaert W, Vandamme EJ (2006) Cloning, sequence analysis and

heterologous expression of the *Myrothecium gramineum* orotidine-5′-monophosphate decarboxylase gene. FEMS Microbiol Lett 261:262–271. doi:10.1111/j.1574-6968.2006.00373.x
15. Altschul SF, Madden TL, Schaffer AA, Zhang JH, Zhang Z, Miller W, Lipman DJ (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res 25:3389–3402. doi:10.1093/nar/25.17.3389
16. Higgins DG, Bleasby AJ, Fuchs R (1992) Clustal-V—improved software for multiple sequence alignment. Comput Appl Biosci 8:189–191
17. Hamilton R, Watanabe CK, Deboer HA (1987) Compliation and comparison of the sequence context around the aug startcodons in *Saccharomyces cerevisiae* messenger-RNAs. Nucleic Acids Res 15:3581–3593. doi:10.1093/nar/15.8.3581
18. Graber JH, Cantor CR, Mohr SC, Smith TF (1999) Genomic detection of new yeast pre-mRNA 3′-end-processing signals. Nucleic Acids Res 27:888–894. doi:10.1093/nar/27.3.888
19. Van Bogaert INA, De Maeseneire SL, De Schamphelaire W, Develter D, Soetaert W, Vandamme EJ (2007) Cloning, characterization and functionality of the orotidine-5′-phosphate decarboxylase gene (*URA3*) of the glycolipid-producing yeast *Candida bombicola*. Yeast 24:201–208. doi:10.1002/yea.1448
20. Van Bogaert INA, Develter D, Soetaert W, Vandamme EJ (2007) Cloning and characterisation of the NADPH cytochrome P450 reductase gene (*CPR*) from *Candida bombicola*. FEMS Yeast Res 7:922–928. doi:10.1111/j.1567-1364.2007.00262.x
21. Fei X, Zhao MW, Li YX (2006) Cloning and sequence analysis of a glyceraldehyde-3-phosphate dehydrogenase gene from *Ganoderma lucidum*. J Microbiol 44:515–522
22. Van Wert SL, Yoder OC (1992) Structure of the *Cochliobolus heterostrophus* glyceraldehyde-3-phospate dehydrogenase gene. Curr Genet 22:29–35. doi:10.1007/BF00351738
23. McAlister L, Holland MJ (1985) Differential expression of the three yeast glyceraldehyde-3-phosphate dehydrogenase genes. J Biol Chem 260:15019–15027
24. Aki T, Yanagisawa S, Akanuma H (1997) Identification and characterization of positive regulatory elements in the human glyceraldehyde 3-phosphate dehydrogenase gene promoter. J Biochem 122:271–278