The biosynthetic gene cluster for sophorolipids: a biotechnological interesting biosurfactant produced by Starmerella bombicola

Inge N. A. Van Bogaert,1* Kevin Holvoet,2,3 Sophie L. K. W. Roelants,1 Bing Li,4 Yao-Cheng Lin,5 Yves Van de Peer4,5 and Wim Soetaert1
1Laboratory of Industrial Biotechnology and Biocatalysis, Department of Biochemical and Microbial Technology, and 2Laboratory of Food Microbiology and Food Preservation, Department of Food Safety and Food Quality, Faculty of Bioscience Engineering, Ghent University, Coupure Links 653, B-9000 Ghent, Belgium.
3Research Group EnBiChem, Department of Industrial Engineering and Technology, University College West-Flanders (Howest), Graaf Karel de Goedelaan 5, B-8500 Kortrijk, Belgium.
4Department of Plant Biotechnology and Bioinformatics, Ghent University, Gent, Belgium.
5Department of Plant Systems Biology, VIB, Technologiepark 927, 9052 Gent, Belgium.

Summary

SOPHOROLIPIDS ARE PROMISING BIOLOGICAL DERIVED SURFACTANTS OR DETERGENTS WHICH FIND APPLICATION IN HOUSEHOLD CLEANING, PERSONAL CARE AND COSMETICS. THEY AREproduced by specific yeast species and among those, Starmerella bombicola (former Candida bombicola) is the most widely used and studied one. Despite the commercial interest in sophorolipids, the biosynthetic pathway of these secondary metabolites remained hitherto partially unsolved. In this manuscript we present the sophorolipid gene cluster consisting of five genes directly involved in sophorolipid synthesis: a cytochrome P450 monooxygenase, two glucosyltransferases, an acetyltransferase and a transporter. It was demonstrated that disabling the first step of the pathway – cytochrome P450 monooxygenase mediated terminal or subterminal hydroxylation of a common fatty acid – results in complete abolishment of sophorolipid production. This phenotype could be complemented by supplying the yeast with hydroxylated fatty acids. On the other hand, knocking out the transporter gene yields mutants still able to secrete sophorolipids, though only at levels of 10% as compared with the wild type, suggesting alternative routes for secretion. Finally, it was proved that hampering sophorolipid production does not affect cell growth or cell viability in laboratory conditions, as can be expected for secondary metabolites.

Introduction

The non-pathogenic yeast Starmerella bombicola has been known for decades as a potent producer of the economical relevant biosurfactant called sophorolipids (Spencer et al., 1970; Fig. 1). Although these glycolipids find applications in various sectors such as household cleaning, personal care and cosmetics and are moreover claimed to have some biological activities (reviewed in Van Bogaert et al., 2011b), detailed information on the core biochemical pathway and involved genes is incomplete. The hypothetical pathway has been reviewed before (Van Bogaert et al., 2007): a hydroxylated fatty acid is in a step-wise manner glucosylated to form a non-acetylated acidic sophorolipid which can be further modified by one or two acetylations of the sophorose disaccharide and/or internal esterification (lactonization). However, the order of these latter events is not clear; some responsible enzymes have not been identified nor is anything known on the transport or secretion process.

As sophorolipids are only produced in stationary phase and do not seem to have an essential function for cell viability, they can be considered as secondary metabolites. Indeed, also other types of glycolipids such as the celluloselipids flocculosin and ustilagic acid synthesized by respectively Pseudozyma flocculosa and Ustilago maydis, are regarded as secondary metabolites with antibiotic properties towards competing organisms (Teichmann et al., 2007; 2011). Other examples are mannosylerythritol lipids (MELs) produced by various Pseudozyma species. The genes for both celluloselipid and MEL synthesis are found in co-regulated clusters (Hewald et al., 2006), a feature quite typical for fungal genes involved in secondary metabolism. Although sophorolipids are claimed to have

Accepted 2 March, 2013. *For correspondence, E-mail inge.vanbogaert@ugent.be; Tel. (+32) 9 264 60 34; Fax (+32) 9 264 62 31.
© 2013 Blackwell Publishing Ltd
some mild antimicrobial activities, we believe that these secondary metabolites are mainly synthesized as extracellular carbon storage material. This is supported by the fact that sophorolipids, parallel to intracellular carbon storage for oleaginous yeasts, are only produced upon high carbon to nitrogen ratio’s; a condition linked to S. bombicola’s natural niche: flowers and bees (‘bombicola’ refers to Bombus terrestris). Conversion of excess carbon material to sophorolipids safeguards the consumption by other organisms (niche protection) which is strengthened by the mild antimicrobial effect. Finally, sophorolipids are indeed metabolized upon starvation conditions (Van Bogaert et al., 2009).

Various enzymes are involved in synthesis of the large pool secondary metabolites. However, tailoring of these structurally diverse molecules is often conducted by similar types of enzymes, e.g. acyl- or acetyltransferases, glycosidases and monoxygenases. Furthermore, transcriptional regulators and transporters take part in regulation. These characteristic tailoring and regulatory genes can be used to pinpoint secondary metabolite gene clusters (Park et al., 2010). The above mentioned cellobioselipid and MEL clusters indeed harbour cytochrome P450 monoxygenases, glycosyltransferases, acyl- and acetyltransferases, specific transporters and transcription factors. We previously reported on two glucosyl- and one acetyltransferase involved in sophorolipid production by S. bombicola (Saerens et al., 2011a,b,c). As there were strong indications that the responsible genes were organized in a cluster, we examined genome sequencing data to localize these genes and discovered the full sophorolipid biosynthetic cluster.

Results and discussion

Defining the sophorolipid cluster

We previously identified two UDP-glucose dependent transferases (UgtA1 and UGTB1); the first one catalyses the coupling of glucose to a hydroxylated fatty acid to give rise to a glucolipid, while the second one uses this product as an acceptor to generate a sophorolipid molecule (Saerens et al., 2011a,c). Furthermore, also the enzyme responsible for the optional acetylations was discovered (Saerens et al., 2011b). In the framework of the Belgian National IWT project ‘Biosurf’, the S. bombicola genome was de novo sequenced and annotated. Surprisingly, the above mentioned genes were found in very close proximity of each other: the acetyltransferase and first glucosyltransferase stop codons are 262 bp separated and the second glucosyltransferase is found 4899 bp upstream of this region (Fig. 2). Other genes found in this region are a putative multidrug resistance protein (MDR) and the cytochrome P450 monoxygenase CYP52M1 (Van Bogaert et al., 2009).

Fig. 1. Examples of sophorolipids produced by S. bombicola.
A. Diacetylated lactonic sophorolipid.
B. Non-acetylated open-chain sophorolipid.
Table 1. Overview of the genes found in the sophorolipid gene cluster, their function and homology with the cellobioselipid genes form U. maydis and P. flocculosa.

<table>
<thead>
<tr>
<th>Gene</th>
<th>Size (AA)</th>
<th>Strand</th>
<th>Function</th>
<th>U. maydis E value</th>
<th>P. flocculosa E value</th>
</tr>
</thead>
<tbody>
<tr>
<td>adh</td>
<td>344</td>
<td>+</td>
<td>Putative alcohol dehydrogenase</td>
<td>/</td>
<td>/</td>
</tr>
<tr>
<td>ugtB1</td>
<td>432</td>
<td>-</td>
<td>Second glucosyltransferase</td>
<td>3.7</td>
<td>1e-3</td>
</tr>
<tr>
<td>mdr</td>
<td>1299</td>
<td>+</td>
<td>Sophorolipid transporter</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>at</td>
<td>259</td>
<td>+</td>
<td>Acetyltransferase</td>
<td>not significant</td>
<td>not significant</td>
</tr>
<tr>
<td>ugtA1</td>
<td>463</td>
<td>-</td>
<td>First glucosyltransferase</td>
<td>3e-1</td>
<td>3e-2</td>
</tr>
<tr>
<td>cyp52m1</td>
<td>538</td>
<td>+</td>
<td>Fatty acid hydrolase</td>
<td>4e-41 (CYP2)</td>
<td>2e-35 (CYP1)</td>
</tr>
<tr>
<td>orf</td>
<td>147</td>
<td>+</td>
<td>Unknown</td>
<td>/</td>
<td>/</td>
</tr>
</tbody>
</table>

et al., 2009) which formerly had been suggested to take part in sophorolipid synthesis based on quantitative RT-PCR data. Yet, the function of this latter gene has never been confirmed. The importance of both genes is discussed further down in this manuscript.

All genes are intronless, which corresponds to the general trend observed for several S. bombicola genes. Similarity of the glucosyltransferases and the cytochrome P450 monoxygenase with the ones involved in cellobioselipid production is low, while there is some resemblance between the transporters and the cytochrome P450 monoxygenase genes (Table 1). As depicted in Fig. 2, the gene pairs ugtb1-mdr and ugtA1-cyp52m1 are transcribed divergently from a common intergenic region. Such arrangements make it very likely that these genes are co-regulated and share regulatory elements located in these intergenic regions. Yet, no conserved motive for transcription factor binding was suggested upon further analysis of these regions.

The five genes very likely taking part in sophorolipid synthesis are flanked by a putative alcohol dehydrogenase and a gene of unknown function at the 5′ and 3′ site respectively. In U. maydis, an alpha dehydrogenase located at the 3′ border of the cellobioselipid gene cluster is responsible for α-hydroxylation of the fatty acid (Teichmann et al., 2007). Based on these similarities, we decided to verify the role of this gene, albeit such reaction at first glance does not intervene in sophorolipid formation. Furthermore, we also examined the small open reading frame, as such small regions could encode for specific transcription factors or regulators. Single knockouts were created and nor the Δadh or Δorf strains displayed a loss in cell viability or a change in sophorolipid titres, composition or timing compared with the wild type when grown under conditions favouring sophorolipid production. These findings indicate that these genes are not essential for sophorolipid biosynthesis and do not effect cell survival in the tested conditions.

Further upstream of the adh gene, another putative cluster is found: six genes similar (42–64% AA ID) to the bacterial entA to -F genes responsible for formation of the siderophore enterobactin combined with a putative transporter and a transmembrane ferric reductase. Although no information is available on the activity of this cluster, it seems like all required elements are there and these are clearly attributed to siderophore formation. Hence, the sophorolipid cluster is not extended at the 5′ site beyond the ugtb1 gene. The 3′ boundary of the cluster is most likely defined by the cyp52m1 gene. Indeed, cyp52m1 is only 29.6 kb away from the telomeric end and in this region only six putative genes are retrieved, including the tested orf, with low similarities to known genes.

The fact that these gene clusters are found subtelomERICally is not exceptional. On the one hand, it is believed that this location is highly susceptible for variation and gene transfer; features associated with adaptation to specific environments and hence the necessity for secondary metabolites. On the other hand this telomere position effect seems to play an important role in transcription controlled by chromatin remodelling and epigenetics; a system suggested to be very suitable for co-regulated expression of large gene clusters (Palmer and Keller, 2010).

MDR: the missing sophorolipid transporter

Despite the industrial importance of S. bombicola and its sophorolipids, it is not known how the sophorolipids are excreted in such high amounts into the culture medium (up to 400 g l⁻¹). Vesicles could be involved, but the process might as well be mediated by active or passive transporters. Up to now, there was no indication that such transporter existed. Yet, the MDR found in the presumed cluster makes up a good candidate.

Translation of this large gene results in a protein of 1299 amino acids and assuming no post-translational modifications this corresponds with a molecular weight of 142 kDa and a isoelectric point (pI) of 6.38. The protein shows up to 49% identity with ABC multidrug resistance transporters of several Aspergillus species. Those transporters take part in the efflux of xenobiotics and/or the...
secretion of antibiotics. AtrDp from *Aspergillus nidulans* for instance, enhances resistance against cytotoxic components and is at the same time required for efficient penicillin secretion (Andrade *et al.*, 2000).

Being transporters, MDR proteins are membrane integrated. Analysis of the amino acid sequence suggested the presence of 12 transmembrane helices (TM; Krogh *et al.*, 2001) and two nucleotide binding domains (NBD; Zdobnov and Apweiler, 2001) arranged in the characteristic homodimer-like (TM6-NBD)2 MDR structure (Fig. 3). When comparing the two halves of the enzyme, there is a striking similarity between them; it is believed that the transporters emerged from a true homodimer after gene duplication and fusion. For example, the MDR Sav1866 from *Staphylococcus aureus* has a TM6-NBD structure and appears as a homodimer (Dawson and Locher, 2007). As presented in Fig. 3, the active part of the transporter is located in the cytosol and in agreement with this, the intracellular loops, including the NBD’s, are highly conserved when compared intra- or intermolecular whereas the TM regions and extracellular loops show higher diversity. Figure 4 shows the alignment of the two NBD’s of the *S. bombicola* sophorolipid transporter. The conserved amino acid sequences for ATP binding: the Walker A (GXXGXGKS/T) and B motifs (hhhhD) and the ABC signature (LSGGQQ/R/KQR), are clearly present (Walker *et al.*, 1982).

In order to investigate the potential role of this transporter in sophorolipid secretion, a knockout cassette was constructed as described in Experimental procedures. This linear fragment was used to transform the ura3-negative *S. bombicola* PT36 strain. The genotype of the transformants was checked by yeast colony PCR with two primerpairs. Five out of 31 colonies displayed the desired profile.

The mutants were first evaluated for their resistance towards several antibiotics. *S. bombicola* is known to be highly resistant towards several antibiotics commonly used in yeast research (Van Bogaert, 2008). Until now, only hygromycin can be used as a dominant drug selective

**Fig. 3.** Transmembrane helix prediction according to Krogh *et al.* (2001) corresponding to the (TM6-NBD)2 structure.

**Fig. 4.** Alignment of the first and second NBD of the sophorolipid transporter. The conserved motives for ATP binding are indicated.
marker, while the yeast keeps growing in the presence of high concentrations of G418, zeocin and phleomycin (e.g. > 1400 μg ml⁻¹ G418, whereas 200 μg ml⁻¹ is sufficient to kill S. cerevisiae). Different cell concentrations of all five mutant strains were put on solid media containing phleomycin, G418 or zeocin. No difference could be observed between the wild type and the mutants; growth was observed at the same time points and for the same cell concentrations. This finding strengthened the hypothesis that the sophorolipid transporter was not directly involved in the high resistance phenotype, but is assigned a specific role in sophorolipid transport.

If the transporter takes part in sophorolipid export, knocking out the gene should result in reduced sophorolipid production or even toxicity for the producing cell. Sophorolipid synthesis of three genetically identical mutants (MDR12, MDR21 and MDR31) was evaluated on rapeseed oil; the preferred hydrophobic carbon source for high sophorolipid yield. A first indication for reduced sophorolipid production is a decrease in glucose consumption. Whereas in the first part of the stationary phase glucose consumption of the wild type and the mutants is more or less the same, there is a clear difference in the later part; glucose is much faster consumed by the wild type. Indeed, quantification of the sophorolipid synthesis revealed a significant difference between the wild type and the mutants; although sophorolipids were still detected, they never reached more than 10% of the wild-type titre (Fig. 5).

It must be stressed that cell growth or viability of the mutants was not affected; colony-forming units (cfu), cell dry weight (CDW) and cell shape were similar to the wild type.

CYP52M1: the first enzyme of the sophorolipid biosynthetic pathway

In order to investigate whether cyp52m1 takes indeed part in the sophorolipid biosynthetic pathway, a S. bombicola strain with a disabled cyp52m1 gene was created as described in the Experimental procedures section. A linear knockout cassette was used to transform S. bombicola wild-type cells. The genotype of 10 transformants was checked by yeast colony PCR and all 10 transformants displayed the right genotype. The effect of the disrupted cyp52m1 gene was tested by evaluating three randomly selected transformants for their production of sophorolipids in liquid medium. CDW and cfu were similar as compared with the wild-type strain, indicating that the gene disruption did not affect cell-growth.

However, similar to the Δmdr strains, glucose consumption is not as effective as compared with the wild type. Furthermore, rapeseed oil stayed floating on the culture medium surface of the knockouts during the whole incubation period despite a normal growth, indicating that also the hydrophobic carbon source required for sophorolipid production is not consumed. Indeed, fatty acids derived from the oil are building blocks for sophorolipid synthesis and upon good sophorolipid production all oil is metabolized. Hence, bad or no consumption of oil indicates that sophorolipid production is negatively affected.

Finally, biosurfactant production was checked by HPLC (high-performance liquid chromatography)-ELSD (Evaporative Light Scattering Detection) analysis of samples taken during and at the end of the incubation period. Whereas there was a clear production of lactonic sophorolipids for the wild type, no sophorolipids of any type could be detected in the medium or the cells of all three knockouts (Fig. 6). The peaks observed for the transformant are degradation products of the not consumed rapeseed oil: oleic acid, the major constituent of the rapeseed oil fatty acids (59%) is detected at 36.2 min and co-elutes with palmitic acid, this latter fatty acid making up about 6% of the rapeseed oil fatty acids. Furthermore also linoleic acid (21%) and stearic acid (2%) are retrieved.

These findings demonstrate that disabling the gene completely inhibits sophorolipid synthesis. To verify that
CYP52M1 is indeed the first enzyme of the sophorolipid biosynthetic pathway, the Δcyp52m1 strain was supplemented with 16-hydroxy-palmitic acid with the aim to restore sophorolipid production. The fatty acid moiety of wild-type sophorolipids consists of terminal or subterminal hydroxylated C16 or C18 fatty acids, with 17-hydroxy-oleic acid being the most predominant (Davila et al., 1994). As 16-hydroxy-palmitic acid is the only commercially available sophorolipid fatty acid intermediate, tests were conducted with this molecule. HPLC-ELSD and LC-MS (Liquid Chromatography Mass Spectroscopy) analysis indeed confirmed the synthesis of C16-harbouring sophorolipids, in this way complementing the effect of the knockout. As expected, no sophorolipids with a fatty acid chain different from C16 were detected (Fig. 7).

The above mentioned results demonstrate that CYP52M1 is the cytochrome P450 monooxygenase responsible for the synthesis of hydroxylated fatty acids, which are essential for sophorolipid production. In contrast to mdr, disabling cyp52m1 fully inhibits sophorolipid synthesis.

Conclusion

In this manuscript we described the sophorolipid biosynthetic gene cluster of the yeast S. bombicola. Sophorolipid production by this organism is of industrial importance as companies from several continents produce these biosurfactants or include them in their products. The cluster consists of five genes and all of them are enzymes directly involved in the core biosynthetic pathway: a cytochrome P450 monooxygenase, two glucosyltransferases, an acetyltransferase and a transporter. No transcription factors or other clear regulatory elements were detected in this region, nor an enzyme taking care of the internal esterification or lactonization. However, this later process might happen spontaneously. Nevertheless, we are currently evaluating several candidate enzymes. The fact that the enzyme is not located in the cluster, might suggest that acetylation occurs prior to ring closure.

Disabling the sophorolipid transporter or the P450 does not affect cell growth and viability. Yet, the phenotype is slightly different as for the Δcyp52m1 strain no sophoroli-
was used instead of 100. Transformants were selected on synthetic dextrose (SD) plates [0.67% yeast nitrogen base without amino acids (DIFCO) and 2% glucose] for the creation of the Δmdr strain and on YPD plates (1% yeast extract, 2% peptone, 2% glucose and 2% agar) containing 500 µg ml⁻¹ hygromycin for creation of the Δcyp52M1 strain. E. coli cells were transformed as described by Inoue et al. (1990).

Creation of the knockout cassettes and strains

**Δmdr.** The coding region of 3900 bp and 521 bp upstream and downstream of the sophorolipid transporter gene were amplified with the primers MDRtotFor and MDRtotRev (Table S1), yielding a fragment of 4789 bp which was cloned into the pGEM-T® vector (Promega). The created vector was digested with BglII, cutting the coding sequence of the gene twice, in this way deleting 2498 bp of the transporter coding region.

The *S. bombicola* Ura3 auxotrophic marker (Van Bogaert et al., 2008b) was inserted by means of the In-Fusion™ 2.0 Dry-Down PCR Cloning Kit (Clontech). The primers uraInfMdrFor and uraInfMdrRev were designed according to the guidelines of the manual and used for integration of the ura3 cassette (2091 bp) into the sophorolipid transporter gene.

The primer pair MDRtotFor and MDRtotREV were used for the amplification of a 4356 bp fragment containing the ura3 marker with approximately 1 kb of the sophorolipid transporter sequence on each site. This linear fragment was used to transform *S. bombicola* PT36 (ura3 auxotrophic mutant). The correct genotype of the transformants was checked by yeast colony PCR with two primer pairs. The first combination, MDRinsertCheckUp and Ura3up.n, verifies the upstream recombination event; MDRinsertCheckUp binds the marker gene of the disruption cassette. The second pair checks the downstream part in the same way: MDRinsertCheckDown binds the genomic DNA preceding the integration region and Ura3up.n binds the marker gene of the disruption cassette. The second pair checks the downstream part in the same way: MDRinsertCheckDown binds the genomic region, whereas ura3OutEndRev binds the marker gene.

**Δcyp52M1.** The 1617 bp coding fragment and 218 and 1060 bp upstream and downstream of the CYP52M1 gene were amplified with the primers A21TotFor and A21TotRev (Table S1), yielding a fragment of 2869 bp which was cloned into the pGEM-T® vector (Promega). The created vector was digested with AvaI, cutting the coding sequence of CYP52M1 twice, in this way deleting 308 bp of the CYP52M1 sequence. The *E. coli* hygromycin resistance gene controlled by the *S. bombicola* GPD promoter (Van Bogaert et al., 2008a) was inserted by means of the In-Fusion™ 2.0 Dry-Down PCR Cloning Kit (Clontech). The primers GHInfA21For and HygroInfA21Rev were designed according to the guidelines of the manual and used for integration of the hygromycin resistance cassette (1968 bp) into CYP52M1. The primerpair A21KnockHygroCasFor and A21KnockHygroCasRev were used for the amplification of a 4003 bp fragment containing the hygromycin resistance cassette with approximately 1000 bp of the CYP52M1 sequence on each site. The fragments were used to transform the *S. bombicola* wild-type strain. The genotype of the transformants was checked by yeast colony PCR with the primer HygroInsertCheckFor, binding on the knockout

### Experimental procedures

#### Strains and culture conditions

*Starmerella bombicola* ATCC 22214 was used in all experiments. When sophorolipid production was intended, the medium described by Lang et al. (2000) was used. 37.5 g l⁻¹ rapeseed oil or 5 g l⁻¹ juniperic acid (Sigma) was added two days after inoculation. Yeast cultures were incubated at 30°C and 200 r.p.m. for a total time of 10 days.

Antibiotic resistance of the mutants was tested on yeast peptone dextrose (YPD) plates (1% yeast extract, 2% peptone, 2% glucose and 2% agar) containing 50 μg ml⁻¹ phleoymycin or 400 or 800 μg ml⁻¹ G418, or 300 μg ml⁻¹ zeocin at pH 6.5 or 7. The different yeast cultures were grown o/n and put at the same optical density before 10-fold dilutions from 10⁻¹ till 10⁻⁵ were made. The plates were incubated at 30°C during several days and growth was monitored daily.

*Escherichia coli* XL10-Gold cells were used in all cloning experiments and were grown in Luria–Bertani (LB) medium (1% trypton, 0.5% yeast extract and 0.5% sodium chloride) supplemented with 100 mg l⁻¹ ampicillin. Liquid *E. coli* cultures were incubated at 37°C and 200 r.p.m.

#### DNA isolation and sequencing

Yeast genomic DNA was isolated with the GenElute™ Bacterial Genomic DNA Kit (Sigma). Preceding protoplast formation was performed by incubation at 30°C for 90 min with zymolase (Sigma).

Bacterial plasmid DNA was isolated with the QIAprep Spin Miniprep Kit (Qiagen). All DNA sequences were determined at Agowa genomics (Berlin, Germany).

#### Transformation

*Starmerella bombicola* cells were transformed with the lithium acetate method (Gietz and Schiestl, 1995), but 50 mM LiAc was used instead of 100. Transformants were selected on synthetic dextrose (SD) plates [0.67% yeast nitrogen base without amino acids (DIFCO) and 2% glucose] for the creation of the Δmdr strain and on YPD plates (1% yeast extract, 2% peptone, 2% glucose and 2% agar) containing 500 µg ml⁻¹ hygromycin for creation of the Δcyp52M1 strain. E. coli cells were transformed as described by Inoue et al. (1990).
cassette and primer A21totRev, binding the genomic region (Table S1).

**Sampling**

Analytical sophorolipid samples were prepared as follows: 440 μl ethylacetate and 11 μl acetic acid were added to 1 ml culture broth and shaken vigorously for 5 min. After centrifugation at 9000 g for 5 min, the upper solvent layer was removed and put into a fresh eppendrof tube with 600 μl ethanol. At the end of the incubation period, 3 vols ethanol were added to the culture broth for total extraction of sophorolipids. Cell debris was removed by centrifugation at 1500 g during 10 min.

For further gravimetric analysis, the supernatants water-ethanol mixture was evaporated. Two volumes of ethanol were added to dissolve the sophorolipids and the residual hydrophobic carbon source. The mixture was filtered to remove the water-soluble compounds and was evaporated again. One volume of water was added and set at pH 7, then 1 vol. of hexane was added and after vigorous shaking, the mixture was allowed to separate. The different fractions were collected, evaporated and the mass was determined. The hexane phase will contain residual oil, while the water phase contains the sophorolipids.

Samples were analysed by HPLC and ELSD, and if required by LC-MS as described before (Van Bogaert et al., 2011a).

Cell dry weight was measured by centrifugation of 2 ml culture broth for 5 min at 9000 g. Pellets were washed two times with ethanol to remove sophorolipids and hydrophobic substrate and finally dissolved in distilled water. The suspension was transferred to a cellulose nitrate filter with a pore diameter of 0.45 μm (Sartorius) and the dry weight was determined in the XM60 automatic oven from Precisa Instruments.

Glucose concentration in the culture supernatans was determined by analysed with the 2700 Select Biochemistry Analyzer (YSI).

For further analyses, the supernatans water-ethanol mixture was evaporated. Two volumes of ethanol were added to dissolve the sophorolipids and the hydrophobic carbon source. The mixture was filtered to remove the water-soluble compounds and was evaporated again. One volume of water was added and set at pH 7, then 1 vol. of hexane was added and after vigorous shaking, the mixture was allowed to separate. The different fractions were collected, evaporated and the mass was determined. The hexane phase will contain residual oil, while the water phase contains the sophorolipids.

**GenBank accession numbers**

The GenBank accession numbers of the discussed genes are: second glucosyltransferase: HM440974; transporter: HQ660581; acetyltransferase: HQ670751; first glucosyltransferase: HM440973; cyp52M1: EU552419.

**Acknowledgements**

The authors would like to thank Lien Saey for technical support. This work was financially supported by the Flemish Agency for Innovation by Science and Technology (Grant IWT80050). The authors have no conflict of interest to declare.

**References**


**Supporting information**

Additional supporting information may be found in the online version of this article at the publisher’s web-site.