(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
21 June 2012 (21.06.2012)

WIPO | PCT

(10) International Publication Number
WO 2012/080116 A1

(51) International Patent Classification:
C12N 1/16 (2006.01) *C12N 9/10* (2006.01)
C12P 19/44 (2006.01)

(21) International Application Number:
PCT/EP2011/072304

(22) International Filing Date:
9 December 2011 (09.12.2011)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
10195091.3 15 December 2010 (15.12.2010) EP

(71) Applicant *(for all designated States except US)*: **UNIVERSITEIT GENT** [BE/BE]; Sint-Pietersnieuwstraat 25, B-9000 Gent (BE).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: **SAERENS, Karen** [BE/BE]; Asselstraat 58, B-9031 Drongen (BE). **VAN BOGAERT, Inge** [BE/BE]; Haeghensgoed 33, B-9270 Laarne (BE). **SOETAERT, Wim** [BE/BE]; Tempeliersdreef 8, B-Lovendegem 9920 (BE).

(74) Agents: **VANHALST, Koen Victor Rachel** et al.; E. Gevaertdreef 10a, B-9830 Sint-Martens-Latem (BE).

(81) Designated States *(unless otherwise indicated, for every kind of national protection available)*: AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

(84) Designated States *(unless otherwise indicated, for every kind of regional protection available)*: ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Declarations under Rule 4.17:

— *of inventorship (Rule 4.17(iv))*

Published:

— *with international search report (Art. 21(3))*

— *before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments (Rule 48.2(h))*

— *with sequence listing part of description (Rule 5.2(a))*

WO 2012/080116 A1

(54) Title: PRODUCING UNACETYLATED SOPHOROLIPIDS BY FERMENTATION

(57) Abstract: Unacetylated sophorolipids show interesting applications in several industrial fields but cannot be produced in a pure and straight way by conventional fermentation. The present invention discloses the production of entirely unacetylated sophorolipids without acetylated variants by fermentation. More specifically, the present invention discloses a yeast strain which is mutated in a gene encoding for an acetyltransferase and which is capable of producing a mixture of entirely unacetylated sophorolipids. Preferably, the majority within said mixture are unacetylated lactonic sophorolipids. In addition, the acetyltransferase of the present invention can be used to acetylate carbohydrates or carbohydrate-containing compounds.

Producing unacetylated sophorolipids by fermentation

Technical field of invention

Unacetylated sophorolipids show interesting applications in several industrial fields but cannot be produced in a pure and straight way by conventional fermentation. The present invention discloses the production of entirely unacetylated sophorolipids without acetylated variants by fermentation. More specifically, the present invention discloses a yeast strain which is mutated in a gene encoding for an acetyltransferase and which is capable of producing a mixture of entirely unacetylated sophorolipids. Preferably, the majority within said mixture are unacetylated lactonic sophorolipids. In addition, the acetyltransferase of the present invention can be used to acetylate carbohydrates or carbohydrate-containing compounds.

Background art

The non-pathogenic yeast *Candida (Starmerella) bombicola* ATCC 22214 (CBS 6009) is commercially applied for the production of sophorolipids. These glycolipid biosurfactants are constituted of a sophorose head group (2-O-β-D-glucopyranosyl-D-glucopyranose) attached to a (sub)terminal hydroxylated C_{18} or C_{16} fatty acid and this by a glycosidic linkage between the anomeric C-atom of the sugar and the hydroxylgroup of the fatty acid. Sophorolipids are typically produced by fermentation in presence of a hydrophobic carbon source and are always constituted of a mixture of structurally related molecules with variation in 1) degree of fatty acid saturation (saturated, mono-unsaturated or di-unsaturated), 2) presence of acetylgroups at C6' and/or C6'' atoms, 3) lactonization between the carboxyl end of the fatty acid and either the C4'', C6' or C6'' atom of the sophorose group, 4) fatty acid chain length and 5) (ω) or (ω-1) hydroxylation of the fatty acid (Asmer *et al.*, 1988).

Due to this structural variation, sophorolipids show many interesting applications in a wide range of industrial fields (Banat *et al.*, 2010; Franzetti *et al.*, 2010; Kralova and Sjoblom, 2009; Mulligan, 2009). Since structural composition is reflected in the physico-chemical properties, several industries are particularly interested in specific structural variants. Acetylation is one structural feature that gains a lot of attention especially because of its contribution to biological activity in addition to its influence on water solubility and foaming properties. The decreased cytotoxicity of unacetylated sophorolipids as compared to acetylated variants for example has attracted attention to use these molecules as new antiviral drugs (Shah *et al.*, 2005). On the other hand, the presence of acetylgroups

strongly increases antibacterial (Gross and Shah, 2004), antifungal (Gross and Shah, 2005) and antiviral (Shah *et al.*, 2005) properties with diacetylated sophorolipids being better antimicrobial agents as compared to mono-acetylated variants. Furthermore, unacetylated sophorolipids have been used as starting molecules for the synthesis of e.g. dispersible nanoparticles (Kasture *et al.*, 2007) and glycolipid derivatives (Azim *et al.*, 2006; Zerkowski *et al.*, 2006) or have served as source molecules for the production of glucolipids and specialty fatty acids (Rau *et al.*, 2001; Saerens *et al.*, 2009), which are on their turn used for synthesis of polymers (Zerkowski *et al.*, 2007) or precursors for plastics and flavours (Rau *et al.*, 2001).

To date, it is impossible to produce unacetylated sophorolipids in a pure way by fermentation in a concentration of more than 50% of unacetylated sophorolipids within a sophorolipid mixture. The best result obtained so far by fermentation technologies was published by Otto *et al.* (1999), who obtained a sophorolipid mixture with 45% being unacetylated lactonic forms after a single-step cultivation of *C. bombicola* ATCC 22214 on whey concentrate and rapeseed oil. The only way to obtain unacetylated sophorolipids in high purity is via the chemical conversion of natural sophorolipid mixtures to unacetylated products by alkaline hydrolysis. By this conversion however, the lactone ring is always hydrolysed and only acidic (= open-ring) unacetylated sophorolipids can be obtained. Since lactonic sophorolipids generally are considered to be industrially more relevant (Hu and Lu, 2001a), a technology that leads to unacetylated sophorolipids herewith leaving the lactone ring intact especially attracts attention.

An alternative way to obtain entirely unacetylated variants would be to prevent acetylation within the sophorolipid-producing organism, e.g. through mutations in or deletions of their acetylating enzymes. Several fungal genomes contain multiple (putative) O-acetyltransferase genes and the specific activity of an O-acetyltransferase is hard to predict based on the amino acid sequence, resulting in a lot of these sequences only being marked as “putative” O-acetyltransferase without any information on the specific action. For instance, in the genome of *C. guilliermondii* ATCC 6260, 2 putative O-acetyltransferase genes with activity towards sugars can be found. In the genome of *Debaryomyces hansenii* CBS767, 5 O-acetyltransferases genes can be retrieved with 2 of them possibly involved in sugar acetylation (DEHA2D01166g, DEHA2F02970g). In the genome of *Yarrowia lipolytica* CLIB122 these numbers are 5 and 3 respectively (YALI0B11176g, YALI0A15081g, YALI0C00187g) and for *Aspergillus nidulans* FGSC A4, 8 O-acetyltransferase genes can be found with 5 of them being putative acetyltransferases with activity to sugar molecules (AN6393.2, AN6836.2, AN3419.2, AN1402.2, AN8386.2).

Summary of the invention

The present invention provides a nucleic acid molecule consisting of the sequence as depicted by SEQ ID N° 1 encoding for an acetyltransferase, or a fragment thereof encoding for a protein retaining said acetyltransferase activity, or a variant thereof encoding for a protein having at least 50% sequence identity with SEQ ID N°2 and having said acetyltransferase activity.

In addition, the invention provides a polypeptide consisting of the amino acid sequence as depicted by SEQ ID N° 2 and having acetyltransferase activity, or a fragment thereof retaining said acetyltransferase activity, or a variant thereof having at least 50% sequence identity with SEQ ID N°2 and having said acetyltransferase activity.

The invention further provides the use of a nucleic acid molecule as defined above having lost its capability to encode for a functional acetyltransferase, or, the use of a polypeptide as defined above having lost its acetyltransferase activity to produce a mixture comprising entirely unacetylated sophorolipids.

In a further aspect, the present invention provides a the use of a fungal species which is capable of producing sophorolipids to produce a mixture comprising entirely unacetylated sophorolipids wherein said fungal species has at least one mutation in a nucleic acid molecule defined above and wherein said mixture comprises at least 50% of entirely unacetylated sophorolipids. In particular, said nucleic acid molecule is depicted by SEQ ID N° 1 encoding for an acetyltransferase, or is a fragment thereof encoding for a protein retaining said acetyltransferase activity, or is a variant thereof encoding for a protein having at least 50% sequence identity with SEQ ID N°2 and having said acetyltransferase activity.

The invention further provides methods of producing a mixture comprising entirely unacetylated sophorolipids, using a nucleic acid molecule as defined above having lost its capability to encode for a functional acetyltransferase, or, the use of a polypeptide as defined above having lost its acetyltransferase activity. In particular, said method comprises the steps of: a) providing a host cell comprising the nucleic acid molecule or polypeptide as defined above, and b) allowing said host cell to produce sophorolipids using standard culturing techniques. Optionally, said sophorolipids can be isolated or purified from the cell culture. Said host cell can be a bacterium, a fungus, a yeast cell, an insect cell, a plant cell or an animal cell.

In a preferred embodiment, said entirely unacetylated sophorolipids comprises at least or equal to 70% of lactonic sophorolipids.

In a further embodiment, said yeast species is selected from the group consisting of *Candida bombicola, Candida apicola, Candida batistae, Candida floricola, Candida riodocensis, Candida stellata, Candida* sp. NRRL Y-27208, *Rhodotorula bogoriensis*, *Wickerhamiella domericqiae* and sophorolipid-producing species of the *Starmerella* clade. Preferably, said *Candida bombicola* is the strain *Candida (Starmerella) bombicola* ATCC 22214.

In a preferred embodiment, said mutation is a deletion and/or insertion and said deletion and/or insertion results in a non-functional polypeptide.

The invention further provides a modified yeast strain belonging to a fungal species capable of producing sophorolipids, characterized in that said fungal strain, compared to an unmodified wild type strain: a) has at least one mutation in a nucleic acid molecule as defined above, and b) produces a mixture of entirely unacetylated sophorolipids comprising at least 50% of entirely unacetylated sophorolipids. In particular, said nucleic acid molecule is depicted by SEQ ID N° 1 encoding for an acetyltransferase, or is a fragment thereof encoding for a protein retaining said acetyltransferase activity, or is a variant thereof encoding for a protein having at least 50% sequence identity with SEQ ID N°2 and having said acetyltransferase activity.

The invention further provides for the use of a polypeptide having acetyltransferase activity as defined above to acetylate carbohydrates or carbohydrate-containing compounds. In particular, said polypeptide consists of the amino acid sequence as depicted by SEQ ID N° 2 and has acetyltransferase activity, or is a fragment thereof retaining said acetyltransferase activity, or is a variant thereof having at least 50% sequence identity with SEQ ID N°2 and having said acetyltransferase activity. The invention further provides for methods for acetylating carbohydrates or carbohydrate-containing compounds, using a polypeptide having acetyltransferase activity as defined above.

The invention further provides for the use of a modified host strain expressing a polypeptide having acetyltransferase activity as defined above to acetylate carbohydrates or carbohydrate-containing compounds. In particular, said polypeptide consists of the amino acid sequence as depicted by SEQ ID N° 2 and has acetyltransferase activity, or is a fragment thereof retaining said acetyltransferase activity, or is a variant thereof having at least 50% sequence identity with SEQ ID N°2 and having said acetyltransferase activity. The invention further provides for methods of acetylating carbohydrates or carbohydrate-containing compounds, using the modified host strain expressing a polypeptide having acetyltransferase activity as defined above.

In a particular embodiment, said modified host strain is transformed with an exogenous nucleic acid molecule as defined above or said modified host strain over-expresses an endogenous nucleic acid molecule as defined above. In particular, said nucleic acid molecule is depicted by SEQ ID N° 1 encoding for an acetyltransferase, or a fragment thereof encoding for a protein retaining said acetyltransferase activity, or a variant thereof encoding for a protein having at least 50% sequence identity with SEQ ID N°2 and having said acetyltransferase activity.

Preferably, said modified host strain is a bacterium, a fungus, a yeast cell, an insect cell, a plant cell or an animal cell.

More preferably, said yeast is *Candida (Starmerella) bombicola* ATCC 22214.

Brief description of figures

Fig. 1: The major-but not the sole- components of the new sophorolipid mixture of the present invention. Under the fermentation conditions tested here, entirely unacetylated lactonic 17- and 18-O-sophorosyl-octadecenoic acid were the most predominant structures (a,b), in addition to entirely unacetylated lactonic 17-O-sophorosyl-octadecanoic acid (c) and acidic 17-O-sophorosyl-octadecenoic acid (d).

Fig. 2: Sophorolipids typically produced during fermentation are considered to be a mixture of compounds represented by formulas a) acidic form and b) lactonic form.

Fig. 3: Complete sequence of the *C. bombicola AT* coding sequence (=SEQ ID N° 1) (GenBank accession number HQ670751) and encoded protein (= SEQ ID N° 2). Conserved amino acids of the active site are written in shaded capitals.

Fig. 4: Alignment of the *C. bombicola* acetyltransferase sequence with five model sequences that define the conserved domain of Lbh_MAT-GAT subfamily proteins. The 30 conserved amino acids from the active site are indicated by arrows. gi|731964 = putative acetyltransferase YJL218W *Saccharomyces cerevisiae*, gi|158333766 = putative maltose O-acetyltransferase *Acaryochloris marina* MBIC 11017, gi|27552460 = putative O-acetyltransferase *Physcomitrella patens*, gi|76803978 = putative maltose O-acetyltransferase *Vibrio sp.* DAT 722, gi|23466264 = thiogalactoside acetyltransferase *Bifidobacterium longum* NCC 2705.

Fig. 5: Schematic representation of the *C. bombicola* wild type *AT* locus (A), the knock-out construct (B) and the Δ*at* mutant allele after homologous recombination between the genome and the construct (C). All primers used for generation of the knock-out construct

and genotype control of the mutants are indicated. Sites of homologous recombination are marked by dotted lines. Coding sequences are indicated by arrows: AT = Acetyltransferase, hygroR = hygromycin B resistance marker.

Fig. 6: Colony forming units for *C. bombicola* ATCC 22214 (filled circles) and the Δ*at* deletion mutant (filled squares). Rapeseed oil was added after 48 hours of incubation (arrow).

Fig. 7: HPLC-ELSD chromatogram from the culture extract of a *C. bombicola* Δ*at* deletion mutant, eleven days after addition of rapeseed oil. Wild type sophorolipids usually elute between 25 and 30 minutes. Corresponding molecular masses were obtained after mass spectrometric analysis of the same sample: unacetylated lactonic 17- and 18-O-sophorosyl-octadecenoic acid (m/z = 604), unacetylated lactonic 17-O-sophorosyl-octadecanoic acid (m/z = 606) and acidic 17-O-sophorosyl-octadecenoic acid (m/z = 622).

Fig 8. Optical density (OD) and glucose consumption for the Δ*at* deletion mutant (AT3) and wild type *C. bombicola* ATCC22214 (WT) cultivated in a 2L Biostat reactor. Additional glucose (30 g/L) was added to the wild type cultivation after 140 hours. Rapeseed oil (approximately 30 g/L) was added stepwise at 28h, 52h and 117h to AT3 and at 46h, 65h and 140h to WT.

Fig 9. Ratio of lactonic to acidic sophorolipids as a function of incubation time for the Δ*at* deletion mutant (AT3) and wild type *C. bombicola* ATCC22214 (WT). Total peak area of acidic and lactonic sophorolipids were calculated from HPLC-ELSD chromatograms and expressed relatively to the total peak area of the sample. Cultivation of the AT3 mutant was prolonged to 383h after additional feeding of glucose and rapeseed oil at 309h.

Fig 10. Ratio of lactonic to acidic sophorolipids in the sophorolipid mixture of the Δ*at* deletion mutant AT3 as a function of incubation time and medium composition. Panel A: "Reference" = standard sophorolipid production medium as described by Lang et al. (2000) with 120 g/L glucose, 4 g/L yeast extract, 5 g/L citrate and 40 g/L rapeseed oil; Panel B: "Excess oil" = 50 g/L glucose; Panel C: "YE 15 g/L" = 15 g/L yeast extract; Panel D: "Excess glucose" = 200 g/L glucose; Panel E: "YE 1 g/L" = 1 g/L yeast extract; Panel F: "No citrate" = 0 g/L Na.citrate.$2H_2O$.

Fig 11. *AT* overexpression construct (A), mutant Δ*ura3* locus of the *ura3*-negative *C. bombicola* (B) and resulting genotype of the *AT* overexpressing mutant after homologous recombination and double cross-over between the mutated *ura3* locus and the construct (C). Primer sites are indicated and regions used for homologous recombination are highlighted by dotted lines.

Fig 12. Growth (CFU) and glucose consumption for the *AT*-overexpressing *C. bombicola* (AT+780_1) and *C. bombicola* ATCC2214 (WT) cultivated in a 2L Biostat reactor.

Fig 13. Mass abundance of diacetylated acidic sophorolipids for the *AT*-overexpressing *C. bombicola* (AT+780_1) and *C. bombicola* ATCC2214 (WT) as a function of incubation time. Mass abundances of compounds with molecular masses 704 (m/z), 706 (m/z) and 708 (m/z) were summed and resulting abundances expressed relatively to the total abundance calculated from the total ion chromatogram.

Description of the invention

The present invention discloses the identification of a single acetyltransferase gene *AT* from *Candida (Starmerella) bombicola* which is fully responsible for the acetylation of sophorolipids. Deletion of the gene surprisingly results in a yeast species producing only unacetylated sophorolipids (see Figure 1). Moreover, under the standard fermentation conditions the lactonic unacetylated sophorolipids are the predominant molecules in the mutant mixture (see Figure 1). With this new structural composition, the created mutant offers a one-step production technology for the fermentative synthesis of industrially important molecules making use of cheap, renewable substrates. Up to date, it was not possible to produce a sophorolipid mixture with this structural composition. Due to the higher foaming capacity and better water solubility of the mutant mixture in addition to the decreased cytotoxicity generally ascribed to unacetylated sophorolipids, these compounds have unique properties and show better performances for several applications such as use as detergent, pharmaceutical applications, cosmetic applications, etc. It should be pointed-out that during the research leading to the present invention two other putative acetyltransferases that belong to the same subfamily as the *AT* gene, were detected in the genome of *C. bombicola*. These respectively showed 43% and 38% sequence identity with the *AT* sequence. Therefore it was not unlikely to think that one of these could replace the *AT* gene function or that the acetylation is carried out by several independent enzymes (e.g. one takes care of the 6' position, and another of the 6'' position). Surprisingly, when creating a single knock-out of the *AT* gene described in this invention, acetylation of sophorolipid molecules was completely blocked.

In addition, the identification of the *AT* gene as the single gene responsible for acetylation of sophorolipids, enables the creation of an overexpression mutant where diacetylated sophorolipids are remarkably enriched thus with the mutant mixture being deprived from mono- and unacetylated sophorolipids. Because the significant increase in antibacterial (Gross and Shah, 2004), antifungal (Gross and Shah, 2005) and antiviral (Shah *et al.*,

2005) activity of diacetylated sophorolipids as compared to mono- or unacetylated variants, such mutant mixture attracts attention of pharmaceutical and medical industries.

The present invention thus relates to the usage of a fungal species which is capable of producing sophorolipids to produce a mixture of entirely unacetylated sophorolipids wherein said fungal species has at least one mutation in a nucleic acid molecule of the present invention encoding for an acetyltransferase of the present invention and wherein said mixture comprises at least 50% of unacetylated sophorolipids. The present invention further preferably relates to the usage as indicated above wherein the majority (i.e. more or equal than 70%) within said mixture are lactonic, unacetylated sophorolipids.

The term 'fungal species capable of producing sophorolipids' refers to a phylogenetically diverse group of yeasts (predominantly Ascomycetes and few Basidiomycetes) which spontaneously synthesize sophorolipids constituted of the sugar sophorose attached to a hydroxylated fatty acid (see Figure 2). Said phylogenetically diverse group of yeasts comprises the species *Candida apicola* (Gorin *et al.*, 1961) which was initially identified as C. *magnolia*, *C. bombicola* (Spencer *et al.*, 1970), *Wickerhamiella domericqiae* (Chen *et al.*, 2006), *Rhodotorula bogoriensis* (Tulloch *et al.*, 1968), *Candida batistae* (Konishi *et al.*, 2008), *Candida floricola* (Imura *et al.*, 2010), *Candida riodocensis*, *Candida stellata* and *Candida sp.* NRRL Y-27208 (Kurzman *et al.*, 2010) and other species of the so-called *Starmerella* clade which encompasses over 40 species.

The term 'wherein said fungal species has at least one mutation in a nucleic acid molecule of the present invention encoding for an acetyltransferase of the present invention' refers to a modified yeast species or yeast strain characterized in having at least one mutation in a nucleic acid molecule of the present invention encoding for an acetyltransferase of the present invention. The term 'mutation' refers to a spontaneous mutation and/or to an induced mutation in the genome of said yeast strain. Said mutation can be a point mutation, deletion, insertion or any other type of mutation. The term most specifically refers to knock outs (KO) via insertion of a KO cassette. Inducing a mutation in the genome of a yeast strain can be undertaken by any method in the art known by a skilled person such as the insertion of a KO cassette into a gene of interest. Similarly, tracing or detecting whether there is a mutation in the genome of a modified strain –compared to a wild type strain- can also be done by any method known in the art.

The term 'sophorolipids' (see Figure 2) refers to carbohydrate-based, amphiphilic biosurfactants that are constituted of the sugar sophorose attached to a hydroxylated fatty acid/alkyl chain, i.e. hydroxylated fatty acid/alkyl chains wherein the fatty acid/alkyl chain contains 5 to 26 carbon atoms. Preferably -and especially with regard to *C. bombicola*- said fatty acid chain is composed of 16 or 18 C-atoms. More specifically the term refers to

glycolipid biosurfactants that are constituted of a sophorose head group (2-O-β-D-glucopyranosyl-β-D-glucopyranose) from which the anomeric C-atom is attached to an (ω) or (ω-1) hydroxylated C_{10}, C_{12}, C_{14}, C_{16}, C_{18}, C_{22} or C_{24} fatty acid. They occur either as open-ring structures (acidic form) or as lactones (closed-ring structures or lactonic form or lactonised form) with an intra-esterification between the fatty acid carboxyl group and the 4'', 6' or 6'' carbon atom of the sophorose head group. In addition, acetyl groups can be attached at the 6' and/or 6'' positions (Asmer *et al.*, 1988).

The term 'unacetylated sophorolipids' refers to sophorolipids without acetylgroups at the C6' and C6'' atoms.

The term 'a mixture comprising at least 50% of unacetylated sophorolipids' refers to a mixture which is less complex (as shown in Figure 1) as compared to the mixture which one obtains in a typical wild type *Candida bombicola* fermentation (Asmer *et al.*, 1988). Indeed the mixture of the present invention is deprived from acetylated forms. In other words, no acetylated forms can be detected –using well-known methods- in the mixture of the present invention.

Hence, the mixture of the present invention comprises at least 50%, i.e. 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 96, 97, 98, 99 or 100% of unacetylated sophorolipids. Preferably, said mixture of the present invention comprises -or can consist of- a majority of unacetylated lactonic sophorolipids (i.e. more or equal than 70%) and a minority of unacetylated acidic sophorolipids (less than 30%). This lactonic:acidic ratio of unacetylated sophorolipids (i.e. 70%/30%, 71%/29%, 72%/28%, 73%/27%, 74%/26%, 75%/25%, 76%/24%, 77%/23%, 78%/22%, 79%/21%, 80%/20%, 81%/19%, 82%/18%, 83%/17%, 84%/16%, 85%/15%, 86%/14%, 87%/13%, 88%/12%, 89%/11%, 90%/10%, 91%/9%, 92%/8%, 93%/7%, 94%/6%, 95%/5%, 96%/4%, 97%/3%, 98%/2%, 99%/1% or 100%/0%) is surprisingly high and can be influenced by changing the incubation time or the medium composition. . In a typical *C. bombicola* fermentation on glucose and oleic acid , 62% is composed of diacetylated lactonic forms, 4% is composed of monoacetylated lactonic forms and 4% is composed of unacetylated lactonic forms while the other compounds are constituted of 1',6' lactones and 1',6'' lactones (4%), acidic sophorolipids (8%) and other lipids at the end of the cultivation period (Asmer *et al.*, 1988). Hu and Ju (2001b) observed a maximum relative percentage of lactonic forms of 50% using soybean oil and 80% using hexadecane. In another certain experiment where palm esters were used, 79.1 % occurred in the lactonic form, but when sunflower oil was applied, only 55.6 % lactonic sophorolipids were retrieved (Davila *et al.*, 1994). When in a mixed carbon source fed-batch fermentation only oil was added in stationary phase, 13% of sophorolipids were in lactonic form (87% acidic) but when a mixture of oil and glucose was fed in stationary

phase, 65% were in lactonic form (Davila *et al.*, 1997). Yeast extract concentration and presence of citric acid influence the ratio of lactonic to acidic sophorolipids too: when yeast extract concentration was 1 g/L, 65% of sophorolipids were in lactonic forms but when the concentration was increased to 20 g/L, all sophorolipids were in acidic form (Casas and Garcia-Ochoa, 1999). Addition of 5 g/L of citric acid to the medium increased the percentage of lactonic forms in the sophorolipid mixture of *Candida apicola* (Hommel *et al.,* 1994).

The exact ratio of lactonic to acidic sophorolipids further changes during cultivation with typically more lactonic forms after prolonged incubation times (Casas and Garcia-Ochoa, 1999; Hu and Ju, 2001b). Hence, the present invention further specifically and preferably relates to the usage as described above wherein said mixture comprises at least or equal to 70% of lactonic, unacetylated sophorolipids.

Furthermore, the present invention relates to the usages as described above wherein said yeast species is selected from the group consisting of *Candida bombicola, Candida apicola, Candida batistae, Candida floricola, Candida riodocensis, Candida stellata, Candida sp.* NRRL Y-27208*, Rhodotorula bogoriensis*, *Wickerhamiella domericqiae* and sophorolipid-producing species of the *Starmerella* clade.

More specifically, the present invention relates to the usage as described above, wherein said *Candida bombicola* is the strain *Candida bombicola* ATCC 22214 (CBS 6009).

The present invention also relates to a modified fungal strain belonging to a fungal species capable of producing sophorolipids as described above, characterized in that said fungal strain, compared to an unmodified wild type strain : a) has at least one mutation in the gene encoding for the acetyltransferase of the present invention , and b) produces a mixture comprising at least 50% entirely unacetylated sophorolipids.

The present invention relates to the usages as described above, wherein said mutation is a deletion in the acetyltransferase of the present invention..

The present invention thus relates to a nucleic acid sequence as depicted by SEQ ID N° 1 encoding for an acetyltransferase, or a fragment thereof encoding for a protein retaining said acetyltransferase activity, or a variant thereof encoding for a protein having at least 50% sequence identity with SEQ ID N°2 and having said acetyltransferase activity. SEQ ID N° 1 corresponds to the following open reading frame of 780 base pairs:

```
1    atggttgtaa actcctcgaa ggaccctcaa aacaaaggaa tgactcctag aaaagaaatt
61   gaccaggaaa tggtctcttg ggccaaaaaa aacctcaaaa acacccctgg caatgaaaac
```

```
121  tatgagaaga tggtctcagg agttccttac aatccatacg atccagatct tatgtttaga
181  gccctggcta ctagtgagaa agttagggag ttcaatacca ttgcaagtga aagtcgtact
241  tttgagtcaa atcacgctgc ttatatcaag aaggtcgaga ttctcaaaga cacttttggt
301  caaacaaagg atattgtctg gctgaccgct ccattctcag ttgattttgg attcaacatc
361  agcgtaggcg agcactttta cgccaacttc aacgtttgct tcttggactc ggctccaata
421  atctttggtg atgaggtgat tgtagggccc aatacaacgt tcgtgactgc gactcatcct
481  attagccccg agaaacgtgc gaggagaatt gtgtatgctc ttcctatcaa ggtggggaat
541  aatgtatgga ttggtgcgaa tgtgactgtc ctgccgggtg ttacgattgg agatggctca
601  acaattgcgg ctggtgctgt cgttcgagaa gatgttcctc ctcgtactgt ggtgggagga
661  gtccctgcgc gaatcctcaa gcatattcca gaggaggatc ccgacgaggc tgaaggagag
721  gaactggaat tccttcttcc agttgaaatg aacgtcaata ccgctaacca gaaggtctag
```

The term 'nucleic acid' and a 'fragment' or a 'variant' thereof corresponds for example to DNA, cDNA, RNA, sense and anti-sense nucleic acids and the like.

The term 'fragment' specifically refers to a nucleic acid sequence containing fewer nucleotides than the nucleic acid sequence as depicted by SEQ ID N° 1 and that encodes for a protein retaining said acetyltransferase activity. The term "variant" specifically refers to a nucleic acid encoding for a protein having at least 50 % sequence identity, preferably having at least 51-70 % sequence identity, more preferably having at least 71-90% sequence identity or most preferably having at least 91, 92, 93, 94, 95, 96, 97, 98 or 99 % sequence identity with SEQ ID N° 2 (see further) or with a fragment thereof, and, that encodes for a protein retaining said acetyltransferase activity. A specific, but non-limiting example of a fragment of the present invention is a fragment that encodes for the amino acid sequence having SEQ ID N° 13 which corresponds to the following sequence:

```
  1  mtprkeidqe mvswakknlk ntpgnenyek mvsgvpynpy dpdlmfrala tsekvrefnt
 61  iasesrtfes nhaayikkve ilkdtfgqtk divwltapfs vdfgfnisvg ehfyanfnvc
121  fldsapiifg devivgpntt fvtathpisp ekrarrivya lpikvgnnvw iganvtvlpg
181  vtigdgstia agavvredvp prtvvggvpa rilkhipeed pdeaegeele fllpvemnvn
241  tanqkv
```

A specific, but non-limiting example of a fragment of the present invention encoding the amino acid sequence having SEQ ID N°13 is the following nucleic acid sequence (SEQ ID N°14):

```
   1  atgactccta gaaaagaaat tgaccaggaa atggtctctt gggccaaaaa aaacctcaaa
  61  aacacccctg gcaatgaaaa ctatgagaag atggtctcag gagttcctta caatccatac
 121  gatccagatc ttatgtttag agccctggct actagtgaga aagttaggga gttcaatacc
 181  attgcaagtg aaagtcgtac ttttgagtca aatcacgctg cttatatcaa gaaggtcgag
 241  attctcaaag acacttttgg tcaaacaaag gatattgtct ggctgaccgc tccattctca
 301  gttgattttg gattcaacat cagcgtaggc gagcactttt acgccaactt caacgtttgc
 361  ttcttggact cggctccaat aatctttggt gatgaggtga ttgtagggcc caatacaacg
 421  ttcgtgactg cgactcatcc tattagcccc gagaaacgtg cgaggagaat tgtgtatgct
 481  cttcctatca aggtggggaa taatgtatgg attggtgcga atgtgactgt cctgccgggt
 541  gttacgattg gagatggctc aacaattgcg gctggtgctg tcgttcgaga agatgttcct
 601  cctcgtactg tggtgggagg agtccctgcg cgaatcctca agcatattcc agaggaggat
 661  cccgacgagg ctgaaggaga ggaactggaa ttccttcttc cagttgaaat gaacgtcaat
 721  accgctaacc agaaggtcta g
```

The present invention further relates to an amino acid sequence as depicted by SEQ ID N° 2 having acetyltransferase activity, or a fragment thereof retaining said acetyltransferase activity or a variant thereof having at least 50% sequence identity with SEQ ID N°2 and having said acetyltransferase activity.

SEQ ID N° 2 corresponds to the following 259 amino acids (given by the one-letter code of amino acids):

```
   1  mvvnsskdpq nkgmtprkei dqemvswakk nlkntpgnen yekmvsgvpy npydpdlmfr
  61  alatsekvre fntiasesrt fesnhaayik kveilkdtfg qtkdivwlta pfsvdfgfni
 121  svgehfyanf nvcfldsapi ifgdevivgp nttfvtathp ispekrarri vyalpikvgn
 181  nvwiganvtv lpgvtigdgs tiaagavvre dvpprtvvgg vparilkhip eedpdeaege
 241  elefllpvem nvntanqkv
```

The term ‘fragment’ refers to a protein (or peptide or polypeptide) containing fewer amino acids than the amino acid sequence as depicted by SEQ ID N° 2 and that retains said acetyltransferase activity. Such fragment can -for example- be a protein with a deletion of 10% or less of the total number of amino acids at the C- and/or N-terminus. A specific example is indicated by SEQ ID N° 13. The term “variant” refers to a protein having at least 50 % sequence identity, preferably having at least 51-70 % sequence identity, more preferably having at least 71-90% sequence identity or most preferably having at least 91,

92, 93, 94, 95, 96, 97, 98 or 99 % sequence identity with SEQ ID N° 2 or with a fragment thereof, and, that encodes for a protein retaining said acetyltransferase activity.

Hence, orthologues, or genes in other genera and species (than the strain *Candida bombicola* ATCC 22214 from which SEQ ID N° 1 and 2 are derived*)* with at least 50% identity at amino acid level, and having the described function are part of the present invention. The percentage of amino acid sequence identity is determined by alignment of the two sequences and identification of the number of positions with identical amino acids divided by the number of amino acids in the shorter of the sequences x 100. The latter 'variant' may also differ from the protein as depicted by SEQ ID N° 2 only in conservative substitutions and/or modifications, such that the ability of the protein to have acetyltransferase activity is retained. A "conservative substitution" is one in which an amino acid is substituted for another amino acid that has similar properties, such that one skilled in the art of protein chemistry would expect the nature of the protein to be substantially unchanged. In general, the following groups of amino acids represent conservative changes: (1) ala, pro, gly, glu, asp, gln, asn, ser, thr; (2) cys, ser, tyr, thr; (3) val, ile, leu, met, ala, phe; (4) lys, arg, his; and (5) phe, tyr, trp, his.

Variants may also (or alternatively) be proteins as described herein modified by, for example, the deletion or addition of amino acids that have minimal influence on the acetyltransferase activity as defined above, secondary structure and hydropathic nature of the enzyme.

The present invention further relates to the usage of a nucleic acid molecule as defined above having lost its capability to encode for a functional acetyltransferase, or, to the usage of a polypeptide as defined above having lost its acetyltransferase activity to produce a mixture comprising entirely unacetylated sophorolipids. A nucleic acid molecule having lost its capability to encode for a functional acetyltransferase as defined above can be obtained by mutation or by any known means to silence the transcription or translation of said nucleic acid such as the insertion of a nucleic acid fragment, a marker gene or others in the functional coding or non-coding part of the acetyltransferase gene, a mutation or removal of the functional coding or non-coding part of the acetyltransferase gene, the usage of specific siRNAs, miRNAs, combinations hereof or any other way. Similarly, a polypeptide as defined above having lost its acetyltransferase activity can be obtained by any (small) compound or other means to disrupt the function of the acetyltransferases of the present invention. Means to silence the transcription or translation or means to disrupt the function of the acetyltransferase of the present invention or means to disrupt the function of a necessary regulator/activator protein of the acetyltransferase thus comprise the usage of any molecule such as –but not limited to- an

antibody, an amino acid, a peptide, a small molecule, an aptamer, a ribozyme, an oligoribonucleotide sequence such a dsRNA used to initiate RNA interference (RNAi) or an anti-sense nucleic acid. Such a molecule is thus capable to bind on a acetyltransferase protein or an activator/regulator protein thereof or is capable to interfere with the cellular synthesis of acetyltransferase or of an activator/regulator thereof by -for example- binding and degrading mRNA's encoding for an acetyltransferase protein or an activator/regulator thereof.

The present invention further relates to the usage of the proteins/polypeptides/peptides having acetyltransferase activity as described above to acetylate carbohydrates or carbohydrate-containing compounds such as sophorolipids, cellobioselipids, alkylglucosides, etc..

Hence, the present invention relates to the usage of a modified host strain expressing a protein having acetyltransferase activity as described above to acetylate carbohydrates or carbohydrate-containing compounds. In this regard, the present invention relates to the usage as described above wherein said modified host strain is transformed with an exogenous nucleic acid sequence as described above or wherein said modified host strain over-expresses an endogenous nucleic acid sequence as described above. As such the present invention relates to the usage as described above, wherein said modified host strain is a bacterium, a fungus, a yeast cell, an insect cell, a plant cell or an animal cell.

The present invention and the above-indicated usages will be illustrated by the following non-limiting examples.

Examples

Example 1: Production of unacetylated sophorolipids

1.1 Materials and Methods

1.1.1 Strains and culture media

C. bombicola ATCC 22214 (CBS 6009) was maintained on YPD medium containing 1% yeast extract, 2% peptone, 2% dextrose and 2% agarose and incubated at 30°C. For production of sophorolipids, *C. bombicola* was shifted to 3C medium containing 10% glucose, 1% yeast extract, 0.1% urea and 2% agarose before inoculation to liquid production medium described by Lang *et al.* (2000). Liquid media were incubated at 30°C and 200 rpm for two days before addition of rapeseed oil (37.5 g/L). *Escherichia coli* DH5α F' was used for plasmid maintenance and grown on Luria Bertani medium (0.5%

Bacto yeast extract, 1% Bacto Trypton, 0.5% NaCl) containing 0.01% ampicillin and incubated at 37°C and 200 rpm. Plasmids were isolated by means of the MiniPrep Plasmid Isolation kit from Qiagen and sequenced at AGOWA (Germany).

1.1.2 Creation of the *AT knock-out* construct

Primer design, sequence analysis and strategy design were performed with the Clone Manager Professional Suite software (Version 8.0). Genome sequencing experiments yielded the pGEM-T® (Promega) derived plasmid pGATtot harbouring the complete acetyltransferase encoding gene sequence in addition to 43 bp and 2426 bp of its up- and downstream sequences respectively. A suitable *knock-out* construct was created from this plasmid by mutation of the *AT* gene followed by integration of a selectable marker (hygromycin B). A frameshift was induced in the *AT* gene by digesting plasmid pGATtot overnight with the single cutter *Bgl*II (New England Biolabs) according to Sambrook (Sambrook and Russell, 2001) resulting in linearization after bp 169 of the *AT* gene. The linearized plasmid was purified by means of the QIAquick® PCR purification kit from Qiagen and 1.77 µg was subjected to 2U Mung Bean Exonuclease (New England Biolabs) to remove the 5' overhangs. Incubation occured at 30°C for 30 minutes. After purification, 100 ng of the linearized plasmid was self-circularized by overnight incubation at 22°C with 5U T4 DNA ligase (Fermentas) in presence of 5% PEG4000 in T4 DNA ligase buffer (Fermentas). The resulting plasmid was called pGATtot mutAT since it contained a mutated acetyltransferase gene and used for transformation of *E. coli* DH5αF' according to Sambrook (Sambrook and Russell, 2001). Plasmids of 10 randomly selected transformants were prepared and deletion of the *Bgl*II restriction site was checked by double digestion with *Bgl*II and *Nco*I. The deletion in the *AT* gene on two positive plasmids was further confirmed by sequencing. Plasmid 10, referred to as pGATtot mutAT10, was chosen for further consideration.

For integration of the hygromycin resistance marker into pGATtot mutAT10, the In-Fusion Dry Down PCR cloning kit (Clontech) was used. Linearisation of pGATtot mutAT10 occurred by PCR with PfuUltra High Fidelity DNA polymerase (Stratagene) and primers KOATmut10linear inf F and KOATmut10 linear inf R (Table 1). Both primers anneal at the end of the mutated *AT* gene. A plasmid containing the hygromycin resistance marker in between tyrosine kinase (TK) terminator and *C. bombicola* GAPD promoter (Van Bogaert *et al.*, 2008a) was used for amplification of the marker by means of primers KOATgapdhygro inf F and KOATgapdhygro inf R (Table 1). These primers each contain 15 bp of homology to the ends of the linearized plasmid pGATtot mutAT allowing subsequent fusion of both fragments. The fusion product was called pGKOAT and cloned

in *E. Coli* Fusion Blue competent cells (Invitrogen), insertion of the marker was verified by colony PCR and correct transformants were used for plasmid isolation as described above. Correctness of the insert was confirmed by sequencing.

1.1.3 Creation of a Δ*at* deletion mutant

A linear knock-out cassette was created from plasmid pGKOAT by PfuUltra High Fidelity DNA polymerase (Stratagene) making use of primers KOAT F and KOAT R (Table 1), the obtained 3742 bp fragment was purified by means of the QIAquick® PCR purification kit (Qiagen) and 688 ng was used for transformation of *C. bombicola* ATCC 22214 by electroporation. For that, the yeast was grown overnight in 100 ml YPD and when OD reached 1, cells were harvested from 50 ml by centrifugation during 5 min at 4°C and 200 rpm. After washing twice with cold and sterile mQ water, cells were resuspended in 2 ml sterile sorbitol solution (1M). After centrifugation at 4°C and 4000 rpm, cells were resuspended in 2 ml sterile lithiumacetate (0.1 M) in presence of 2.5 mM DTT and left to rest at room temperature for 10 to 15 min. Cells were then pelleted again and washed with 2 ml sorbitol (1M) before resuspending in 250 µl sorbitol (1M). From this suspension, 50 µl was carried over into a sterile eppendorf tube, 688 ng of the linear *knock-out* construct was added and the mixture was incubated on ice for 2 min before transfer to a 2 mm electroporation cuvette. A pulse of 1.5 kV (200 Ohm) was given during 5 milliseconds after which 1ml of cold and sterile YPD was added immediately. The cells were incubated for 1h at 30°C and 200 rpm and centrifuged at room temperature during 5 min at 4000 rpm. Cell pellet was then resuspended in 1ml sorbitol (1M) and aliquots of 200 µl were plated on selective YPD plates containing 500 µg/L hygromycin B. Plates were incubated at 30°C until transformant colonies appeared.

1.1.4 Characterisation of the Δ*at* deletion mutant

The genotype of obtained mutants was checked by means of two yeast colony PCR reactions checking left- and right-sided homologous recombination with primer pairs KOATCtrl F/GAPD -775R and HygroInsertCheck F/KOATCtrl R respectively (Table 1). The reaction mix was composed of 1.25 U Standard Taq polymerase, 2.5 µl reaction buffer and 200 µM dNTP's (all from New England Biolabs) in addition to 400 µM of each primer and milliQ water to a total volume of 25 µl. A single colony was picked by means of a sterile needle and added to the reaction mixture. PCR cycling program was set at 94°C for 7 minutes followed by 30 cycli of subsequently 94°C for 30 s, 54°C for 30 s and

72°C for 2.5 min, the latter having time increments of 5 s per cycle. Final elongation was performed at 72°C during 7 minutes.

The mutants were checked for sophorolipid production by growing them in Lang medium as described above. *C. bombicola* ATCC 22214 served as a reference. Colony forming units and pH of the cultures were followed at regular time intervals throughout the whole experiment. In accordance, glucose concentrations were determined in the supernatant of 1 ml culture samples by means of HPLC-RI. Sophorolipid formation was checked 11 days after addition of rapeseed oil by means of HPLC-ELSD.

1.1.5 Sophorolipid extraction

To determine sophorolipid yields, both *C. bombicola* ATCC 22214 (wild type) and Δ*at* mutant were grown in 50 ml Lang medium. During incubation of the cultures, sophorolipid production was followed by extracting 1 ml culture sample with 400 µl technical ethylacetate in presence of 10 µl acetic acid. After vortexing for 5 min and centrifugation at 9000 g for 5 min, 300 µl of the solvent phase was diluted in 1.7 ml absolute ethanol and analysed on HPLC-ELSD. Quantitative sophorolipid yields were obtained 19 days after addition of rapeseed oil by addition of 500 µl acetic acid to each 50 ml culture and extracting three times with one volume of technical ethylacetate. Solvent phases were poured together and evaporated to dryness. Residual oil was removed and quantified by washing the crude extract twice with 10 ml technical n-hexane and evaporating combined hexane phases to dryness. Purity of the extracts was checked by means of HPLC-ELSD.

1.1.6 High Performance Liquid Chromatography

Glucose samples were analysed on a Varian ProStart HPLC system equipped with an Aminex® HPX-87H ion exclusion column (300 mm x 7.8 mm) (Biorad) coupled to a LaChrom RI detector (Merck). Column temperature was at 65°C and compounds were eluted in 25 minutes by means of a 5 mM H_2SO_4 solution at a constant flow rate of 0.6 ml/min. Culture extracts were analysed for sophorolipid content on a Varian ProStar HPLC system equipped with a Chromolith® Performance RP-18e 100-4.6-mm column (Merck KGaA) and connected to an Evaporative Light Scattering Detector (Alltech). Compounds were eluted by means of an acetonitrile/acetic acid (0.5% in water) gradient (5/95 to 95/5 in 40 min) under constant flow of 1 mL/min. Column temperature was set at 30°C. To check molecular masses of the produced sophorolipids, the same samples were analysed under the same conditions on a Shimadzu LC-10-AD HPLC system

connected to a quadrupole mass spectrometer (Waters). Molecules were identified by their native molecular masses after ESI (electron spray ionisation) without collision.

1.1.7 Biostat 1L Fed-Batch fermentation

To study sophorolipid production in more detail, a 1L fermentation was set up with the *C. bombicola* Δ*at* mutant AT3 and compared to the wild type yeast ATCC22214. A 100 ml preculture in medium of Lang *et al.* (2000) was inoculated from a 3C agar plate (1.1.1) and grown overnight at 30°C and 200 rpm. This preculture was used as inoculum for 1L fresh medium in a 2L benchtop BioStat fermentor (Sartorius Stedim Biotech, Germany). At the time of inoculation, 10% (v/v) sterile rapeseed oil was added to the preculture. Aeration rate was 1 L/min, agitation speed 800 rpm and temperature set at 30°C. pH was allowed to drop freely to 3.5 and was then buffered at this value by means of NaOH (5N). Rapeseed oil was added in fed-batch manner (in a total concentration of 37 g/L for the AT3 mutant and 30 g/L for the wild type yeast) generally in steps of 10 g/L at times were foaming became significant or when no residual oil was observed in daily taken samples. If necessary, also glucose was added in fed-batch manner. At regular time points, samples were taken to follow OD, glucose concentration and sophorolipid production (see 1.1.5 and 1.1.6). Glucose samples were quantified by means of a YSI Select 2700 Biochemistry Analyzer.

1.1.8 Lactonic:Acidic ratio as a function of medium composition

The effect of medium composition on the ratio of lactonic to acidic sophorolipids in the mixture of the AT3 mutant was evaluated in erlenmeyer flask experiments. One preculture of 100 ml production medium was prepared (1.1.1) and used to inoculate six 100 ml cultures with startOD equal to 0.2. All cultures were based on the medium described by Lang et al. (2000), but in 5 of them some slight modifications were incorporated as summarized in Table 3. To all cultures, 40 g/L rapeseed oil was added in 4 steps of 10 g/L after 48h, 96h, 144h and 192h. In the culture “excess glucose” stepwise addition of rapeseed oil was accompanied by additional feeding of 20 g/L glucose. At regular time points, colony forming units, glucose concentration, pH, optical density and sophorolipid production were determined as described above.

1.2 Results

1.2.1 Analysis of the *C. bombicola* acetyltransferase

Genome sequencing experiments with *C. bombicola* ATCC 22214 revealed a putative open reading frame of 780 bp with high homology to putative bacterial and fungal carbohydrate transacetylases, as shown by a blastx homology search to all non-redundant protein sequences available at the NCBI database (Altschul *et al.*, 1997). Table 2 gives an overview of the ten best blastx-hits. Highest homology of 65% (47% sequence identity) is found to a putative protein from *Clavispora lusitaniae* ATCC 42720 (gb/EEQ36094). The putative *C. bombicola* acetyltransferase of 259 amino acids had an estimated molecular mass of 28.6 kDa and estimated pI of 4.96. The complete coding sequence is given in Figure 3. The complete *AT* gene was submitted to GenBank under accession number HQ670751.

Analysis of the acetyltransferase (At) protein sequence via the Conserved Domain Database (CDD) (Marchler-Bauer *et al.*, 2009) at the NCBI website, shows that the protein belongs to the LbetaH superfamily (CDD accession number cd00208) of proteins, characterised by a left-handed parallel β-helix. Most of these proteins show acetyltransferase activity though some are involved in iron transport and translation initiation. The presence of a LbH_MAT_GAT conserved domain (CDD accession number cd003357) in the *C. bombicola* At sequence, confirms that this protein belongs to the Maltose-O-Acetyltransferase (MAT) and Galactoside-O-acetyltransferase (GAT) subfamily of the LbHMAT_like protein family (CDD accession number cd 04647). All putative proteins with high homology (Table 2) show this same conserved domain in their sequence. Figure 4 illustrates the alignment of the *C. bombicola* At protein sequence with five model sequences used to define the conserved domain. Both MAT and GAT catalyze the CoenzymeA-dependent acetylation of the C6 carbon atom of their respective carbohydrates. For MAT however, activity was observed to substrate analogues which have a hydrophobic side chain attached to the carbohydrate molecule. An active acetyltransferase is a trimer with three active sites being formed by the LbH interfaces. Each active site is a coenzyme A binding site, represented by 15 conserved amino acid residues.

1.2.2 Creation of the Δ*at* deletion mutant

To find out whether this gene is involved in sophorolipid production, a deletion mutant was created from *C. bombicola* ATCC 22214. Recombination efficiency in *C. bombicola* is increased when double-sided homologous recombination can occur between the targetted genome locus and a linear DNA fragment of which both left- and right-sided flanks show around 1000 bp homology with the genomic sequence (Van Bogaert *et al.*, 2008b). Thus, a suitable *knock-out* cassette is usually based on the integration of a selectable marker between 1000 bp of the 5' and 3' ends of the gene that needs to be deleted. The creation of the Δ*at* deletion construct was started from plasmid pGATtot, a pGEM-T® derived plasmid harbouring the complete AT coding sequence (780 bp) in addition to 43 bp of the upstream and 2426 bp of the downstream sequences respectively. For that, first a frame-shift was introduced in the *AT* coding sequence to assure that homologous recombination between wild type *AT* locus and the construct would result in the replacement of the *AT* gene by a disfunctional *at* copy. Figure 5 shows a schematic drawing of the wild type *AT* locus, the created *knock-out* construct and the Δ*at* mutant locus after homologous recombination between construct and genome.

After transformation of *C. bombicola* ATCC 22214 with the *knock-out* construct, many transformants were observed on the selective YPD plates of which 72 were subjected to yeast colony PCR to check their genotype. Of these, 88% resulted from double cross-over events, 8% from single cross-over events and 4% from illegitimate recombination. From the double cross-over mutants, three yeast colonies were randomly selected for further study of the phenotype.

1.2.3 Sophorolipid production by the Δ*at* deletion mutant

All three mutants showed equal growth as compared to the wild type yeast when grown in Lang medium with addition of rapeseed oil. For one randomly selected mutant, colony forming units were followed during cultivation and compared to the values obtained for the wild type yeast (Figure 6). After 9 days (7 days after addition of rapeseed oil), glucose was completely depleted in both cultures. Sophorolipid production was checked for all three mutants and the wild type yeast eleven days after addition of rapeseed oil, and the HPLC-ELSD chromatograms obtained showed that all three Δ*at* deletion mutants produced identical sophorolipids which were clearly more polar as compared to those from the wild type yeast. To verify if this was due to the absence of acetylation, the culture extract from one randomly selected mutant was analysed by mass-spectrometry. Figure 7 shows the HPLC-ELSD result from a culture extract from the Δ*at* mutant AT4 and the corresponding

masses as obtained by mass-spectrometric analysis. These masses confirm that only unacetylated sophorolipids were produced. Entirely unacetylated lactonic 17- and 18-O-sophorosyl-octadecenoic acid (m/z = 604) were the most predominant structures in the mixture, in addition to minor amounts of open-ring 17-O-sophorosyl-octadecenoic acid (m/z = 622) and lactonic 17-O-sophorosyl-octadecanoic acid (m/z = 606). No acetylated variant was observed in the mixture.

To determine total sophorolipid yield, the experiment was repeated and a total extraction was done 19 days after addition of rapeseed oil. While the wild type yeast produced 33 ± 4 g/L sophorolipids, only 5 ± 0.7 g/L were obtained from the mutant. Accordingly, 15.4 ± 0 g/L residual oil was found for the mutant while no residues could be detected for the wild type yeast. By this time, glucose was almost completely consumed by both cultures (0.07 ± 0.03 g/L residual glucose for the wild type and 0.13 ± 0.03 g/L for the mutant). Although the extraction procedure might need some optimization for the new sophorolipid mixture (the culture supernatant still contained minor amounts of sophorolipids), the significant amount of residual oil indicates that conversion of rapeseed oil to unacetylated sophorolipids by the Δ*at* deletion mutant occurs with lower efficiency as compared to native sophorolipid production by the wild type yeast. This might be explained by less efficient export of the unacetylated sophorolipids herewith delaying further synthesis of the compounds. Though the ABC transporter involved in secretion of cellobioselipids by *U. maydis* appeared not to be very selective (Teichmann *et al.*, 2007), the secretion of mannosylerythritol lipids by the same fungus seemed to require a double acylation at the 2' and 3' carbon of the mannosyl residue (Hewald *et al.*, 2006). Sort of selectivity of a transporter involved in sophorolipid secretion is thus not impossible. However, yields might be increased by optimization of process parameters e.g. fed-batch addition of rapeseed oil instead of batch.

The created Δ*at* deletion mutant produces mainly lactonic unacetylated sophorolipids. These are only produced in very minor amounts during conventional fermentation with *C. bombicola* or any other sophorolipid producing yeast species. The created mutant thus produces a new and inventive sophorolipid mixture which is mostly constituted of unacetylated lactonic glycolipids. Although sophorolipids are classified as low-foaming detergents, the unacetylated sophorolipid mixture of the mutant clearly showed increased foaming as was observed during extraction from the culture broth. The lack of acetylgroups tremendously increases the water solubility of the product so that the unacetylated sophorolipids have unique properties that are of interest to several industries. Since both acetylation and lactonization have important contributions to the antimicrobial, antiviral and cytotoxic properties, the unacetylated lactonic sophorolipids are

of interest to pharmaceutical industry as well (Gross and Shah, 2004; Gross and Shah, 2005, Shah *et al.*, 2005).

1.2.4 Evaluation of sophorolipid production in a BioStat 1L bioreactor

In order to evaluate growth and sophorolipid production in a more controlled manner the *C. bombicola* AT3 mutant was cultivated in a BioStat 2L fermentor and compared with the wild type yeast grown under the same conditions. Figure 8 shows the growth curve and glucose consumption for the AT3 mutant as compared to the wild type yeast. In this well-controlled bioreactor experiment, glucose consumption and optical density in the exponential phase is comparable between wild type yeast and AT3 mutant illustrating equal growth rates. In late exponential phase (at the onset of sophorolipid production) and throughout the whole stationary phase however, glucose is clearly faster consumed by the wild type yeast as compared to the AT3 mutant indicating that more glucose is directed to sophorolipid production in the cultivation with the wild type yeast. The steeper increase in optical density, typically related to sophorolipid production in stationary phase, further illustrates that this happens more efficiently in the wild type cultivation as compared to the AT3 cultivation. HPLC-ELSD analysis of the sophorolipids confirmed that total sophorolipid peak area was at the maximum factor 50 higher for the wild type sample extracts as compared to those of the mutant (results not shown). To investigate the structural composition of the sophorolipid mixture produced, and more specifically the ratio of lactonic to acidic sophorolipids, the peaks observed in the HPLC-ELSD chromatograms were first structurally identified after analysis of the same samples on HPLC-MS (1.1.6). Subsequently, the compounds were divided in two main groups: the acidic sophorolipids typically eluting between 19 and 25 min and the lactonic forms typically found between 25.5 and 30.5 min. Peak areas were calculated for both groups and expressed relatively to the total peak area obtained. Figure 9 illustrates the lactonic to acidic ratios in the sophorolipid mixtures of both strains. Generally the most predominant sophorolipids produced under conventional fermentation conditions (excess of glucose and in addition of a hydrophobic carbon source) are lactonic ones, and the ratio of lactonic to acidic sophorolipids further increases upon incubation. As clear from Figure 9, this is not only true for the wild type yeast but also for the AT3 mutant. While the ratio lactonic:acidic is initially a bit lower as compared to the wild type (70:30 as compared to 92:8), the final ratio obtained becomes comparable. Because after 212h there was still glucose present in the culture supernatant of the AT3 mutant, this cultivation was prolonged to 309h and finally to 383h after additional feeding of glucose (46 g/L) and rapeseed oil (10 g/L) after 309h. This prolonged incubation finally yielded a lactonic:acidic ratio of 99.9:0.1. It is surprising that

lactonic forms are the most predominant compounds in the mixture of the Δ*at* deletion mutant and the fact that the ratio of lactonic to acidic sophorolipids in the mutant mixture is comparable to the wild type yeast indicates that the defectiveness of the acetylation function does not influence the lactonization pattern of the sophorolipids proving that acetylation is not a prerequisite for lactonization.

1.2.5 Ratio of lactonic to acidic sophorolipids in function of medium composition

Unacetylated sophorolipids are of interest to several industries due to their different biological properties, foaming properties and better water solubility. For some applications, lactonic sophorolipids are preferable above acidic ones while for other industries the opposite might be true. Indeed, lactonization has a strong contribution to the overall polarity and chemical versatility of the sophorolipid. For wild type *C. bombicola* as well as for *C. apicola*, many papers have illustrated the effect of medium composition on the amount of acidic and lactonic sophorolipids (Casas and Garcia-Ochoa, 1999; Davila *et al*., 1992; Davila *et a*l., 1994; Hommel *et al*., 1994; Hu and Ju, 2001b). To investigate these effects on the mixture of unacetylated sophorolipids of the Δ*at* deletion mutant and in that way explore the flexibility of the mutant in view of tailored production of unacetylated sophorolipids in each of both configurations, 5 cultures with adapted composition were compared to the standard sophorolipid production medium as described by Lang *et al*., (2000). The ratio of glucose to oil, the concentration of yeast extract and the presence or absence of citric acid were the investigated parameters. Figure 10 shows the results regarding the sophorolipid lactonic:acidic ratios under the different conditions. The amount of glucose to oil, the concentration of yeast extract as well as the presence of citrate are clearly affecting the sophorolipid composition of the AT3 mutant. Not only sophorolipid lactonic:acidic ratio, but also cell growth is affected by a change in yeast extract concentration or the omittment of citrate. Since yeast extract is rich in growth factors, a decrease in its concentration to 1 g/L results in an early delay of cell growth (and corresponding glucose consumption) as compared to the standard production medium with 4 g/L yeast extract. Surprisingly, sophorolipid production is comparable to the standard conditions indicating higher productivity of the cells and the same lactonic:acidic ratio is found with a final ratio of 87:13 after 9 days (Figure 10). Probably the excess of glucose accumulated because of the limited growth is promoting sophorolipid synthesis. Accordingly, increasing the concentration to 15 g/L results in a prolongation of the growth phase, early glucose depletion and limited sophorolipid production. Total sophorolipid production in this case is comparable to the culture where initial glucose concentration was lower (“Excess oil”) but the pattern in lactonic:acid ratio during the incubation period is

different (Figure 10). In high yeast extract concentrations ("YE 15 g/L"), acidic sophorolipids are the most predominant compounds formed but from the moment glucose becomes limiting, the relative amount of lactonic sophorolipids increases and the mixture slowly shifts to predominantly lactonic forms after 140h with a final ratio of 65:35 after 9 days. Though the final ratio of the culture with limited initial glucose concentration ("Excess oil") is comparable (57:43 after 9 days, see Figure 10) as is the moment of glucose depletion, the pattern shift in lactonic:acidic ratio of the latter culture is different with predominantly lactonic forms in early stage of sophorolipid production. In contrast to the culture with high yeast extract concentration (YE 15g/L), the importance of lactonic sophorolipids here declines in function of incubation time. This suggests that yeast extract is somehow promoting the synthesis of acidic sophorolipids or limitates the production of lactonic sophorolipids as was also reported by Casas and Garcia-Ochoa (1999). Indeed, under initially high yeast extract concentrations, it takes some time before this compound becomes limiting inducing a late shift to more lactonic forms at the end of the cultivation. When yeast extract concentration is at the standard value of 4 g/L this compound becomes limiting after 48h, resulting in predominantly lactonic forms at early stage of sophorolipid production. The percentage of lactonic forms decreases without going to minority probably due to the supplemental feeding and high excess of oil as was reported by (Davila et al. 1997) (Figure 10). Citrate has a strong buffering effect and omitting it from the medium results in strong pH decrease with values below 2.5 at the end of exponential phase. As a result of this, cell growth slows down earlier as compared to the reference culture and sophorolipid production is extremely low. Under such conditions, acidic sophorolipids are the main compounds produced with a stable lactonic:acid ratio of about 10:90 (Figure 10). It has been reported by Hommel et al. (1994) that presence of citrate induces production of lactonic sophorolipids and it has been suggested that this is due to an activation of the lactonesterase responsible for lactonization of the molecules. However, the fact that the pH drops to extremely low values rises the idea that this pH effect is the direct parameter affecting sophorolipid structure, either by inhibiting an extracellular lactonesterase or by spontaneous cleavage of the ester bounds in the lactonic sophorolipids. Finally, the culture where addition of oil was accompanied by supplemental feeding of glucose ("Excess glucose") results in the same growth characteristics and sophorolipid production as compared to the standard production medium. In this standard medium, glucose is added already in excess and additional feeding of glucose in stationary phase just prevents limitation thereof. The final lactonic:acid ratio is comparable with the reference culture (92:8 for "Excess glucose" culture as compared to 87:13 for "Reference" culture, Figure 10). It is possible that supplemental addition of glucose keeps the glucose:oil ratio quite high in comparison with

the standard production medium where this ratio becomes small under glucose depletion at the end of the cultivation period. In agreement with the report of Davila et al. (1997), this might result in a small relative increase in acidic forms for the reference culture which is not observed for the culture "Excess glucose" though cultivation needs to be followed up for longer incubation times to judge if this difference is significant or not.

To conclude, the Δ*at C. bombicola* deletion mutant surprisingly behaves comparable to the wild type *C. bombicola* ATCC22214 regarding the influence of medium composition on the lactonic:acidic ratio of its produced sophorolipids. This creates the opportunity to further direct the production of these new unacetylated sophorolipids to more acidic or more lactonic forms, depending on the needs of a specific application field. It confirms that lactonization is completely independent from acetylation and shows that modification of the lactonization pattern of the new unacetylated sophorolipid mixture is as modifiable as it is for the wild type sophorolipid mixture.

Example 2: Production of sophorolipids enriched in diacetylated forms

2.1 Materials and Methods

2.1.1 Creation of the *AT* overexpression construct

An expression construct was created by fusing the *C. bombicola AT* gene to the constitutive GAPD promotor instead of its native inducable promotor. For that the GAPD promotor was amplified from plasmid GAPDpromHygro (Van Bogaert *et al.*, 2008a) using the PfuUltra High Fidelity DNA polymerase (Stratagene) and primers GAPDHygro1560 For and GAPDAT FusionB (Table 1). The *AT* coding region in addition to its terminator sequence were amplified the same way from plasmid pGATtot using primers GAPDAT FusionC and GAPDAT FusionD (Table 1). Both fragments were purified and fused to each other by PCR (Table 4 and 5) : in a first primerless PCR both fragments are allowed to anneal to each other by means of their overlapping sequences and in a second PCR round, the fused product is amplified after adding 0.6mM of each outer primer. The fusion product GAPDATterm was purified from the mix by gelextraction using the Qiaex II gel extraction kit (Qiagen). The 3' downstream sequence of the *C. bombicola URA3* locus was coupled to the 3' end of this GAPDATterm fragment. For that, the 3' downstream sequence of the *URA3* gene (145ORF3) was amplified from *C. bombicola* ATCC 22214 genomic DNA (Saerens *et al.*, 2011) by PfuUltra High Fidelity polymerase using primers GAPDATterm FusionE and GAPDAT FusionF (Table 1). The GAPDATterm fragment was prepared for fusion by amplification with the same polymerase using primers

GAPDHygro1560 For and GAPDATterm RevFus (Table 1). After purification, both fragments were fused as described above giving rise to the fragment GAPDATterm_145ORF3. This fragment was purified from the mix by gelextraction. The selection marker *URA3* in addition to its 5' upstream sequence (ORF4URA3) was amplified from *C. bombicola* ATCC 22214 genomic DNA by PfuUltra High Fidelity PCR using primers P1 and URA3TK Rev (Table 1). The TK terminator was amplified from plasmid GAPDpromHygro (Van Bogaert *et al.*, 2008a) using primers URA3TK For and HygroCDS&T Rev (Table 1) and fused to the ORF4URA3 fragment as described above to create the fragment ORF4URA3TK. After gelextraction, the latter fragment was fused to the GAPDATterm_145ORF3 fragment. For that GAPDATterm_145ORF3 was amplified by means of primers GAPDAT ForFus short and GAPDAT FusionF while ORF4URA3TK was amplified using primers ORF4URA3TK bp21F and URA3TK RevFus short. After purification, both fragments were fused to give rise to the final overexpression construct ORF4URA3TK_GAPDATterm145ORF3 (= AT780overex). The final cassette was cloned in pJET using the CloneJET™ PCR Cloning kit (Fermentas) and the resulting plasmid pJET_AT780overex was transferred to *E. coli* XL10 Gold cells for plasmid maintenance following the protocol of Inoue (Sambrook and Russell, 2001) with addition of 2 µl β-mercaptoethanol (Stratagene). The correctness of the sequences on plasmid pJET_AT780overex was checked by sequencing. A second overexpression construct, harbouring part of the *AT* coding sequence under control of the GAPD promotor was generated the same way. This partial *AT* sequence lacks the first 39 bp leading to a 5' truncated sequence referred to as AT741. The corresponding fragment GAPDAT741term was generated by PCR using fragment GAPDATterm as a template. The GAPD promotor was amplified using primers GAPDHygro 1560For and GAPDAT FusionBbis and the truncated *AT* sequence by means of primers GAPDAT FusionCbis and GAPDAT FusionF (Table 1). After purification of both fragments, they were fused by PCR to generate fragment GAPDAT741term. Further generation of the complete overexpression construct AT741overex was done as described for the AT780overex construct.

2.1.2 Creation and characterization of an *AT*-overexpressing *C. bombicola*

A linear cassette was amplified from plasmid pJET_AT780overex and from pJET_AT741overex using PfuUltra High Fidelity Polymerase and primers ORF4URA3TK bp114F and GAPDAT FusionF (Table 1). Cassettes were purified and 140 ng (AT780overex) resp. 126.5 ng (AT741overex) were used for transformation (1.1.3) of an *ura3* deficient *C. bombicola* derived from wild type *C. bombicola* ATCC22214.

Transformant yeasts were selected on SD medium (1.1.1) and checked by colony PCR for their genotype. Left- and right-sided cross-over was checked by primer pairs P33 and URA3Un respectively P35 and ATR (Table 1). Selected mutants arising from a double cross-over event were grown in sophorolipid production medium (1.1.1) to check their phenotype (growth and glucose consumption) and sophorolipid production. Wild type *C. bombicola* ATCC22214 was taken as a reference. In order to follow up sophorolipid production under well-controlled conditions, one of those mutants was used in a 1L Fed-Batch fermentation and growth, glucose consumption and sophorolipid production were compared to the wild type yeast (1.1.7). Daily samples were analysed for sophorolipids as described in 1.1.5 and 1.1.6. Final sophorolipid yields were quantified by a total extraction of the fermentation broth. For this, 3 volumes of technical ethanol were added to the fermentation broth and cellular debris was removed by centrifugation at 8000 rpm for 10 min. The supernatant was evaporated to dryness and the residue redissolved in 2.5 volumes technical ethanol before filtration. Non-dissolved particles were additionally rinsed with half a volume of technical ethylacetate to assure complete solubilization of all sophorolipids. The filtrate was evaporated to dryness and rinsed with half a volume of technical hexane. The hexane phase was removed and evaporated to dryness. Sophorolipid yields and residual oil were determined by weighing the residues obtained from the ethanol/ethylacetate and hexane phase respectively.

2.2 Results

Based on homologies between the 5' and 3' ends of the *AT*-overexpression construct and the mutated *ura3* locus of the *ura3* deficient host, integration of the *AT*-overexpression cassette occurs at this specific locus on the genome, resulting in one additional copy of the *AT* gene under control of the constitutive GAPD promotor (Figure 11). Indeed, the Δ*ura3*-deficient host still harbours its native copy of the *AT* gene under its own promotor elsewhere in its genome. In this way, acetyltransferase synthesis from this GAPD-controlled gene results in protein during the exponential growth phase of the mutant. This will result in a cytoplasmatic pool of acetyltransferase protein by the time sophorolipid production is initiated by nitrogen limiting conditions at the end of the exponential phase. In that way, more acetylation of freshly synthesised sophorolipids can occur leading to an enrichment of diacetylated forms and deprivation of mono- and unacetylated forms. Several mutants were obtained after transformation of the *ura3*-deficient *C. bombicola* with the overexpression constructs AT780overex (the complete *AT* sequence) and AT741overex (the truncated *AT* sequence). Two mutants that had resulted from a double cross-over event for each construct were subsequently grown in sophorolipid production

medium and compared to the wild type *C. bombicola* ATCC22214. All mutants showed comparable growth, glucose consumption and sophorolipid production as compared to this wild type yeast. In order to investigate the effect of the constitutively expressed *AT* copy on the structural variation of sophorolipids, one mutant referred to as AT+780_1 was grown in a 2L bioreactor and compared to the wild type yeast. Figure 12 shows the growth and glucose consumption for the *AT* overexpressing mutant AT+780_1 as compared to wild type *C. bombicola* ATCC22214. From this figure it is clear that the *AT* overexpressing mutant shows equal growth characteristics as compared to the wild type yeast. Also sophorolipid yields, determined as total peak area in the obtained HPLC-ELSD chromatograms, and the ratio of lactonic to acidic sophorolipids were comparable between both yeast strains. More of interest is the structural variation in terms of acetylation pattern. Under the standard fermentation conditions we have applied here, diacetylated lactonic forms typically are the most predominant compounds produced by *C. bombicola*. In addition, a significant amount of diacetylated acidic forms are produced while mono- and unacetylated sophorolipids (in both lactonic and acidic configuration) count for minor amounts (Asmer *et al.*, 1988). The additional *AT* copy in the genome of the mutant AT+780_1 is under control of the GAPD promotor and will thus lead to synthesis of acetyltransferase during the exponential growth phase. As a result of this, there will be a pool of acetyltransferase protein by the time that sophorolipid production is initiated at early stationary phase leading to more efficient acetylation of the freshly synthesized compounds and ideally bringing the number of mono- and unacetylated sophorolipids to zero. In all the samples we prepared from both wild type and *AT* overexpressing yeast, diacetylated lactonic sophorolipids were as expected the most important structural variants in addition to significant amounts of diacetylated acidic forms. To investigate the effect of the additional *AT* copy on the acetylation pattern of the sophorolipids, the relative abundancy of the diacetylated acidic sophorolipids was calculated from the mass spectra obtained. Indeed, the pool of acetyltransferase present in the AT+780_1 mutant at the onset of sophorolipid production will lead to more efficient acetylation at this early stage of production which is characterised by a higher relative amount of acidic forms as compared to later stages of production. The three most abundant diacetylated acidic sophorolipids had a molecular mass of 704 (m/z), 706 (m/z) and 708 (m/z) corresponding to sophorolipids with $C_{18:2}$, $C_{18:1}$ and $C_{18:0}$ lipid moiety respectively. In all samples obtained during the whole cultivation period, the relative abundancy of these three peaks was remarkably higher for the AT+780_1 mutant as compared to the wild type yeast (Figure 13).These results demonstrate that under standard fermentation conditions, one additional copy of the *AT* gene under control of the constitutive GAPD promotor results in a remarkable increase in diacetylated acidic

sophorolipids as compared to the wild type *C. bombicola* ATCC22214 and this counts for the whole incubation period.

Example 3: Acetylation of other carbohydrate containing compounds

For acetylation of other glucopyranose containing glycolipids such as cellobioselipids or glucolipids, or alkylglucosides the *C. bombicola* acetyltransferase is expressed in a suitable host. This host can either be *E. coli* or *S. cerevisiae*. Activity of the acetyltransferase towards alkylglucosides having a decyl aliphatic chain has been shown using the soluble protein fraction of a *C. bombicola* ATCC22214 cell lysate. Such protein source however contains minor amounts of *de novo* synthezised sophorolipids that make purification of the acetylated product more complex. For that, a heterologous host unable to produce sophorolipids can simplify the production of acetylated carbohydrate containing compounds in a pure way. A heterologous *E. coli* BL21 strain harbouring a suitable expression construct has been created. This construct is based on the pTrc99A plasmid (NCCB, The Netherlands) where the acetyltransferase is under control of the strong inducible *trc* promotor. Induction of protein expression occurs by addition of 0.1 mM IPTG (Carbosynth) to the culture medium, and cells are harvested for lysis 6 hours later. Enzymatic lysis occurs by means of the EasyLyse™ Bacterial Protein Extraction Solution (Epicentre Biotechnologies) where 200 µl of EasyLyse solution is added to each pellet from a 1ml culture sample with OD around 4. After lysis, the soluble fraction is used as crude acetyltransferase preparation for subsequent acetylation reactions. These occur in a suitable buffer system and acetylated products can be analysed by HPLC-ELSD after solvent extraction of the reaction mixture. On the other side, a heterologous *S. cerevisiae* FY 1679-01B is created using both the pYES2.1 TOPO®TA Expression Kit from Invitrogen and the overexpression constructs created for *C. bombicola* (see example 2).

Cellobioselipids are produced in nature by several fungal species of which the plant pathogenic dimorphic fungus *Ustilago maydis* is the best known (Hewald *et al.*, 2006; Spoeckner *et al.*, 1999). These compounds have attracted attention because of their high antifungal activity (Kulakovskaya et al, 2009 ; Kulakovskaya *et al.*, 2004 ; Kulakovskaya *et al.*, 2007). In accordance with sophorolipids, cellobioselipids also occur as mixtures of structurally related molecules where variations occur in acylation pattern and fatty acid composition (Lemieux and Charanduk, 1951). Since acetylation has been shown to strongly influence the antifungal spectrum of sophorolipids (Gross and Shah, 2005), this is also expected for cellobioselipids. Therefore, it is interesting to acetylate cellobioselipids with acetylgroups instead of the naturally occuring acylgroups. For this, unacylated cellobioselipids are used as substrate for the above mentioned *in vitro*

acetyltransferase reaction. Natural cellobioselipids are obtained after cultivation of haploid *Ustilago maydis* DSM17146 (Hewald *et al.*, 2005) on Potato Dextrose Broth (PDB) for one week. After cultivation, total cellobioselipid extraction occurs with ethylacetate and the solvent is removed by evaporation. The cellobioselipids are subsequently converted to the acidic deacylated form by alkaline hydrolysis following the protocol as described for sophorolipids (Saerens *et al.*, 2009). Deacylation of the cellobioselipids was confirmed by mass spectrometry. These deacylated cellobioselipids will be used in an acetyltransferase assay to give rise to acetylated cellobioselipids instead of the natural acylated cellobioselipids. Presence of acetylated cellobioselipids will be verified by solvent extraction of the reaction mixture and subsequent analysis on HPLC-ELSD.

Table 1. Primers used for creation of the Δ*at knock-out* construct, AT overexpression construct and genotype control of the obtained transformants. All primers were ordered at Sigma Genosys.

Primer name	**Primer function**	**Sequence**
KOATmut10linear inf F	linearisation pGATtotmutAT	5' GGATCCTCCTCTGGAATATG 3' (SEQ ID N° 3)
KOATmut10linear inf R	linearisation pGATtotmutAT	5' CGAGGCTGAAGGAGAGGAAC 3' (SEQ ID N° 4)
KOATgapdhygro inf F	amplification selection marker	5' TCTCCTTCAGCCTCGGAGAGTGGATCACGAGTAAG 3' (SEQ ID N° 5)
KOATgapdhygro inf R	amplification selection marker	5' TCCAGAGGAGGATCCGAACAAACGACCCAACACC 3' (SEQ ID N° 6)
KOAT F	amplification KO cassette	5' CAACGCCCAAGCACCGAACTCAATTCAC 3' (SEQ ID N° 7)
KOAT R	amplificati	5' TTCCTCCTTCCTTGCCTCATTCC 3'(SEQ ID N° 8)

Primer name	Primer function	Sequence
	on KO cassette	
KOATCtrl F	colony PCR	5' CAGCAGAGACCATCTGCCTAGCAACTTC 3' (SEQ ID N° 9)
GAPD-775 R	colony PCR	5' GCCACTGCCATTGGAGATTG 3' (SEQ ID N° 10)
HygroInsertCheck F	colony PCR	5' TTCGACAGCGTCTCCGACCTGAT 3' (SEQ ID N° 11)
KOATCtrl R	colony PCR	5' TGGTCTGGCCCTGAGTCTGAAG 3' (SEQ ID N° 12)
GAPDHygro1560 For	high fidelity PCR	5' GACATCCGATGTGTAGTTAATCA 3' (SEQ ID N° 15)
GAPDAT FusionB	fusion PCR	CTTCGAGGAGTTTACAACCATTTGTGTAGAGTTGTTTTTG (SEQ ID N°16)
GAPDAT FusionC	fusion PCR	CAAAAACAACTCTACACAAATGGTTGTAAACTCCTCGAAGG (SEQ ID N° 17)
GAPDAT FusionD	fusion PCR	AAATGACTGACAACAATGGATTGGAAAGCCACAACTCC (SEQ ID N° 18)
GAPDATterm FusionE	fusion PCR	GAGTTGTGGCTTTCCAATCCCTCAAACAGTTCCTTCAATGC(SEQ ID N° 19)
GAPDAT FusionF	high fidelity PCR	GCCTCGTCAACCATCTTATC (SEQ ID N° 20)
GAPDATterm RevFus	fusion PCR	GCATTGAAGGAACTGTTTGAGGGATTGGAAAGCCACAACTC (SEQ ID N° 21)
P1	high fidelity PCR	AGAACAAGGCCGAGTATGTC (SEQ ID N° 22)
URA3TK Rev	fusion	GTTAGCCTCCCCCATCTCCCTCATCTTGACTGAACTTT

Primer name	Primer function	Sequence
	PCR	TCTC (SEQ ID N° 23)
URA3TK For	fusion PCR	GAAAAGTTCAGTCAAGATGAGGGAGATGGGGGAGGCTAAC (SEQ ID N° 24)
HygroCDS&T Rev	high fidelity PCR	TGAACAAACGACCCAACAC (SEQ ID N° 25)
GAPDAT ForFus short	fusion PCR	GGTGTTGGGTCGTTTGTTCACTCCAATGGCAGTGGCTTACC (SEQ ID N° 26)
ORF4URA3TK bp21F	high fidelity PCR	GACGTTTCCAGGGACAACAG (SEQ ID N° 27)
URA3TK RevFus short	fusion PCR	GGTAAGCCACTGCCATTGGAGTGAACAAACGACCCAACACCC (SEQ ID N° 28)
GAPDAT FusionBbis	fusion PCR	CAATTTCTTTTCTAGGAGTCATTTGTGTAGAGTTGTTTTTG (SEQ ID N° 29)
GAPDAT FusionCbis	fusion PCR	CAAAAACAACTCTACACAAATGACTCCTAGAAAAGAAATTG (SEQ ID N° 30)
ORF4URA3TK bp114F	high fidelity PCR	GCCTTGGCGCTTGAGAAGTTCCC (SEQ ID N° 31)
P33	colony PCR	CCATACTCAAGCGCGAACAC (SEQ ID N° 32)
Ura3Un	colony PCR	CAGCCTCTTCTAGTCCGCTCACAATTCC (SEQ ID N° 33)
P35	colony PCR	GAGCTCAAGACGCGTTTACTCAATGC (SEQ ID N° 34)
ATR	colony PCR	CATGACAGCCTTTTCTTCTT (SEQ ID N° 35)

Table 2. Ten best homology scores for the *C. bombicola AT* gene after a blastx homology search against all non-redundant protein sequences available at NCBI website.

Protein	Organism	Accession n°[a]	% Hm[b]	% Id[c]
Hypothetical protein	*Clavispora lusitaniae* ATCC 42720	XP_002619058	65	47
Hypothetical protein	*Pichia pastoris* GS115	XP_002490805	63	40
Hypothetical protein	*Meyerozyma guilliermondii* ATCC 6260	XP_001486565	62	47
Hypothetical protein	*Debaromyces hansenii* CBS 767	XP_460495	61	46
Hypothetical protein	*Debaromyces hansenii* 767	XP_458519	60	44
Hypothetical maltose transacetylase	*Bacillus pseudofirmus* OF4	YP_OO3428191	60	41
Hypothetical maltose transacetylase	*Prevotella melaninogenica*	YP_003813638	59	42
Hypothetical protein	*Candida tropicalis* MYA-3404	XP_002551418	59	39
Hypothetical protein	*Gibberella zeae* PH-1	XP_388002	58	43
Hypothetical maltose transacetylase	*Lodderomyces elongisporus* NRRL YB-4239	XP_001526138	57	37

[a] GenBank Accession Number, [b] % Homology, [c] % Identity

5

Table 3. Medium composition for the six different cultures used to evaluate the effect on the ratio of lactonic to acidic sophorolipids.

	Reference	Excess oil	Excess glucose	YE 15 g/L	YE 1 g/L	No citrate
Component	**Concentration (g/L)**					
Glucose	120	**50**	120	120	120	120
Yeast extract	4	4	4	**15**	**1**	4

	Reference	Excess oil	Excess glucose	YE 15 g/L	YE 1 g/L	No citrate
Component	**Concentration (g/L)**					
Na.citrate.$2H_2O$	5	5	5	5	5	**0**
NH_4Cl	1.5	1.5	1.5	1.5	1.5	1.5
KH_2PO_4	1	1	1	1	1	1
K_2HPO_4	0.16	0.16	0.16	0.16	0.16	0.16
$MgSO_4.2H_2O$	0.7	0.7	0.7	0.7	0.7	0.7
NaCl	0.5	0.5	0.5	0.5	0.5	0.5
$CaCl_2.2H_2O$	0.27	0.27	0.27	0.27	0.27	0.27

Table 4 Composition of the PCR mix for fusion of two fragments. Both fragments are added in a molar ratio of 1 to 1.

Components	Final concentration	Volume
Milli-Q water		Up to 50 µl
10x *PfuUltra* AD buffer	1x	5 µl
2 mM dNTP mix	0.25 mM	6.25
Fragment 1	max 25 ng/kb	x µl
Fragment 2	max 25 ng/kb	x µl
PfuUltra DNA polymerase	2.5 U	1 µl

Table 5 Fusion PCR cycling program.

	Cycle	Time	Temperature (°C)
	1x	2 min	94
	15x	30 s	94
PCR1		1 min	Depending on Tm

	Cycle	Time	Temperature (°C)
		1.5 min for first kb, 1 min for additional kb	72
	1x	7 min	72
	1x	2 min	94
	30x	30 s	94
PCR2		1 min	Depending on Tm
		1.5 min for first kb, 1 min for additional kb	72
	1x	7 min	72

References

Altschul SF, Madden TL, Schaffer AA, Zhang JH, Zhang Z, Miller W, Lipman D (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* **25:** 3389-3402.

Asmer HJ, Lang S, Wagner F, Wray V (1988) Microbial production, structure elucidation and bioconversion of sophorose lipids. *J Am Oil Chem Soc* **65:** 1460-1466.

Azim A, Shah V, Doncel GF, Peterson N, Gao W, Gross R (2006) Amino acid conjugated sophorolipids: a new family of biologically active functionalized glycolipids. *Bioconjugate Chem* **17**: 1523-1529.

Banat IM, Franzetti A, Gandolfi I, Bestetti G, Martinotti MG, Fracchia L, Smyth TJ, Marchant R (2010) Microbial biosurfactants production, applications and future potential. *Appl Microbiol Biotechnol* **87**: 427-444.

Casas JA, Garcia-Ochoa F (1999) Sophorolipid production by *Candida bombicola*: Medium composition and culture methods. J Biosci Bioeng **88**:488-494

Chen J, Song X, Zhang H, Qu YB, Miao JY (2006) Production, structure elucidation and anticancer properties of sophorolipid from *Wickerhamiella domercqiae*. *Enzyme Microb Technol* **39**: 501-506

Davila AM, Marchal R, Vandecasteele JP (1992) Kinetics and balance of a fermentation free from product inhibition - sophorose lipid production by *Candida bombicola*. *Appl Microbiol Biotechnol* **38**: 6-11.

Davila AM, Marchal R, Vandecasteele JP (1994) Sophorose lipid production from lipidic precursors - Predictive evaluation of industrial substrates. *J Indust Microbiol* **13**: 249-257.

Davila AM, Marchal R, Vandecasteele JP (1997) Sophorose lipid fermentation with differentiated substrate supply for growth and production phases. *Appl Microbiol Biotechnol* **47**:496-501.

Franzetti A, Tamburini E, Banat IM (2010) Applications of biological surface active compounds in remediation technologies. In *Advances in Experimental medicine and biology* Volume 672. Ramkrishna S (ed). Springer-Verlag Berlin: Germany; 121-134.

Gorin PAJ, Spencer JFT, Tulloch AP (1961) Hydroxy fatty acid glycosides of sophorose from *Torulopsis magnoliae*. *Can J Chem* **39**: 846-855.

Gross R, Shah V (2004) Antimicrobial properties of various forms of sophorolipids. International Patent WO 2004/044216 A1.

Gross R, Shah V (2005) Antifungal properties of various forms of sophorolipids US Patent US 2005/0164955 A1.

Hewald S, Josephs K, Bölker M (2005) Genetic analysis of biosurfactant production in *Ustilago maydis*. *Appl Environ Microb* **71:** 3033-3040.

Hewald S, Linne U, Scherer M, Marahiel MA, Kämper J, Bölker M (2006) Identification of a Gene Cluster for Biosynthesis of Mannosylerythritol Lipids in the Basidiomycetous Fungus *Ustilago maydis. Appl Environ Micr* **72** : 5469-5477.

Hommel RK, Weber L, Weiss A, Himmelreich U, Rilke O, Kleber HP (1994) Production of sophorose lipid by *Candida (Torulopsis) apicola* grown on glucose. *J Biotechnol* **33**:147-155.

Hu YM, Ju LK (2001a) Purification of lactonic sophorolipids by crystallization. *J Biotechnol* **87**: 263-272.

Hu YM, Ju LK (2001b) Sophorolipid production from different lipid precursors observed with LC-MS. *Enzyme Microbial Technol* **29**: 593-601.

Imura T, Masuda Y, Minamikawa H, Fukuoka T, Konishi M, Morita T, Sakai H, Abe M and Kitamoto D (2010) Enzymatic conversion of *unacetylated* sophoroselipid into acetylated glucoselipid: surface-active properties of novel bolaform biosurfactants. *J Oleo Sci* **59**: 495-501.

Kasture M, Singh S, Patel P, Joy PA, Prabhun AA, Ramana CB, Prasad BLV (2007) Multiutility sophorolipids as nanoparticle capping agents: synthesis of stable and water dispersible co nanoparticles. *Langmuir* **23** : 11409-11412.

Konishi M, Fukuoka T, Morita T, Imura T, Kitamoto D (2008) Production of new types of sophorolipids by *Candida batistae. J Oleo Sci* **57**: 359-369.

Kralova I, Sjoblom J (2009) Surfactants used in food industry: a review. *J Disper Sci Technol* **30**: 1363-1383.

Kulakovskaya T, Shashkov A, Kulakovskaya E, Golubev W, Zinin A, Tsvetkov Y, Grachev A, NIfantiev N (2009) Extracellular cellobiose lipid from yeast and their analogues: structures and fungicidal activities. *J Oleo Sci* **58**: 133-140.

Kulakovskaya TV, Shashkov AS, Kulakovskaya EV, Golubev WI (2004) Characterization of an antifungal glycolipid secreted by the yeast *Sympodiomycopsis paphiopedili*. *FEMS Yeast Res* **5:** 247-252.

Kulakovskaya EV, Kulakovskaya TV, Golubev CI, Shashkov AS, Grachev AA, Nifantiev NE (2007) Fungicidal activity of cellobiose lipids from culture broth of yeast *Cryptococcus humicola* and *Pseudozyma fusiformata*. *Russ J Bioorg Chem* **33**: 156-160.

Kurtzman CP, Price NPJ, Ray KJ, Kuo TM (2010) Production of sophorolipid biosurfactants by multiple species of the *Starmerella* (*Candida*) *bombicola* yeast clade. *FEMS Microbiol Lett* **311**: 140-146.

Lang S, Brakemeier A, Heckmann R, Spöckner S, Rau U (2000) Production of native and modified sophorose lipids. *Chim Oggi* **18:** 76-79.

Lemieux RU, Charanduk (1951) Biochemistry of the Ustilaginales. VI. The acyl groups of ustilagic acid. *Can J Chem* **29**: 409-414.

Marchler-Bauer A, Anderson JB, Chitsaz F, Derbyshire MK, DeWeese-Scott C, Fong JH, Geer LY, Geer RC, Gonzales NR, Gwadz M, He S, Hurwitz DI, Jackson JD, Ke ZK, Lanczycki CJ, Liebert CA, Liu C, Lu F, Lu S, Marchler GH, Mullokandov M, Song JS, Tasneem AK, Thanki N, Yamashita RA, Zhang D, Zhang N, Bryant SH (2009) CDD: Specific functional annotation with the Conserved Domain Database. *Nucleic Acids Res* **37**: 205-210.

Mulligan CN (2009) Recent advances in the environmental applications of biosurfactants. *Curr Opin Colloid In* **14**: 372-378.

Otto RT, Daniel HJ, Pekin G, Müller-Decker K, Fürstenberger G, Reuss M and Syldatk C (1999) Production of sophorolipids from whey. II. Product composition, surface active properties, cytotoxicity and stability against hydrolases by enzymatic treatment. *Appl Microbiol Biotechnol* **52**:495-501.

Rau U, Hammen S, Heckmann R, Wray V, Lang S (2001) Sophorolipids: a source for novel compounds. *Ind Crop Prod* **13**: 85-92.

Saerens K, Van Bogaert I, Soetaert W, Vandamme EJ (2009) Production of glucolipids and specialty fatty acids from sophorolipids by *Penicillium decumbens* naringinase : Optimization and kinetics. *Biotechnol J* **4 :** 517-524.

Saerens KMJ, Roelants S, Van Bogaert INA, Soetaert W (2011) Identification of the UDP-glucosyltransferase gene *UGTA1*, responsible for the first gluvcosylation step in the sophorolipid biosynthetic pathway of *Candida bombicola* ATCC22214. *FEMS Yeast Res* **11**:123-132.

Sambrook J, Russell DW (2001) *Molecular Cloning: A Laboratory Manual*, *3rd edition*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York.

Shah V, Doncel GF, Seyoum T, Eaton KM, Zalenskaya I, Hagver R, Azim A, Gross R (2005) Sophorolipids, microbial glycolipids with anti-human immunodeficiency virus and sperm-immobilizing activities. *Antimicrob Agents Chemother* **49**: 4093-4100.

Spencer JFT, Gorin PAJ, Tulloch AP (1970) *Torulopsis bombicola* sp. n. *A Van Leeuw J Microb Ser* **36**: 129-133.

Spoeckner S, Wray V, Nimtz M, Lang S (1999) Glycolipids of the smut fungus *Ustilago maydis* from cultivation on renewable resources. *Appl Microbiol Biotechnol* **51**: 33-39.

Teichmann B, Linne U, Hewald S, Marahiel MA, Bölker M (2007) A biosynthetic gene cluster for a secreted cellobiose lipid with antifungal activity from *Ustilago maydis*. *Mol Microbiol* **66**: 525-533.

Tulloch AP, Spencer JFT, Deinema MH (1968a) A new hydroxy fatty acid sophoroside from *Candida bogoriensis*. *Can J Chem* **46**: 345-348.

Van Bogaert INA, De Maeseneire SL, Develter D, Soetaert W, Vandamme EJ (2008a) Cloning and characterisation of the glyceraldehyde 3-phosphate dehydrogenase

gene of *Candida bombicola* and use of its promotor. *J Ind Microbiol Biotechnol* **35** : 1085-1092.

Van Bogaert INA, De Maeseneire SL, Develter D, Soetaert W, Vandamme EJ (2008b) Development of a transformation and selection system for the glycolipid producing yeast *Candida bombicola*. *Yeast* **25:** 272-278.

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.

Zerkowski JA, Solaiman DKY (2007) Polyhydroxy fatty acids derived from sophorolipids. *J Amer Oil Chem Soc* **84:** 463-471.

Zerkowski JA, Solaiman DKY, Ashby RD, Foglia TA (2006) Head group-modified sophorolipids : synthesis of new cationic, zwitterionic, and anionic surfactants. *J Surfactants Deterg* **9 :** 57-62.

Claims

1. A nucleic acid molecule consisting of the sequence as depicted by SEQ ID N° 1 encoding for an acetyltransferase, or a fragment thereof encoding for a protein retaining said acetyltransferase activity, or a variant thereof encoding for a protein having at least 50% sequence identity with SEQ ID N°2 and having said acetyltransferase activity.

2. A polypeptide consisting of the amino acid sequence as depicted by SEQ ID N° 2 and having acetyltransferase activity, or a fragment thereof retaining said acetyltransferase activity, or a variant thereof having at least 50% sequence identity with SEQ ID N°2 and having said acetyltransferase activity.

3. Use of a nucleic acid molecule according to claim 1 having lost its capability to encode for a functional acetyltransferase, or, use of a polypeptide according to claim 2 having lost its acetyltransferase activity to produce a mixture comprising entirely unacetylated sophorolipids.

4. Use of a fungal species which is capable of producing sophorolipids to produce a mixture comprising entirely unacetylated sophorolipids wherein said fungal species has at least one mutation in a nucleic acid molecule according to claim 1 and wherein said mixture comprises at least 50% of entirely unacetylated sophorolipids.

5. Use according to claim 4 wherein said entirely unacetylated sophorolipids comprises at least or equal to 70% of lactonic sophorolipids.

6. Use according to any of claims 4-5 wherein said yeast species is selected from the group consisting of *Candida bombicola, Candida apicola, Candida batistae, Candida floricola, Candida riodocensis, Candida stellata, Candida* sp. NRRL Y-27208, *Rhodotorula bogoriensis*, *Wickerhamiella domericqiae* and sophorolipid-producing species of the *Starmerella* clade.

7. Use according to claim 6, wherein said *Candida bombicola* is the strain *Candida (Starmerella) bombicola* ATCC 22214.

8. Use according to any of claims 4-7, wherein said mutation is a deletion and/or insertion and wherein said deletion and/or insertion results in a non-functional polypeptide.

9. A modified yeast strain belonging to a fungal species capable of producing sophorolipids , characterized in that said fungal strain, compared to an unmodified

wild type strain : a) has at least one mutation in a nucleic acid molecule according to claim 1 , and b) produces a mixture of entirely unacetylated sophorolipids comprising at least 50% of entirely unacetylated sophorolipids.

10. Use of a polypeptide having acetyltransferase activity according to claim 2 to acetylate carbohydrates or carbohydrate-containing compounds.

11. Use of a modified host strain expressing a polypeptide having acetyltransferase activity according to claim 2 to acetylate carbohydrates or carbohydrate-containing compounds.

12. Use according to claim 11 wherein said modified host strain is transformed with an exogenous nucleic acid molecule according to claim 1 or wherein said modified host strain over-expresses an endogenous nucleic acid molecule according to claim 1.

13. Use according to claims 11-12, wherein said modified host strain is a bacterium, a fungus, a yeast cell, an insect cell, a plant cell or an animal cell.

14. Use according to claim13 wherein said yeast is *Candida (Starmerella) bombicola* ATCC 22214.

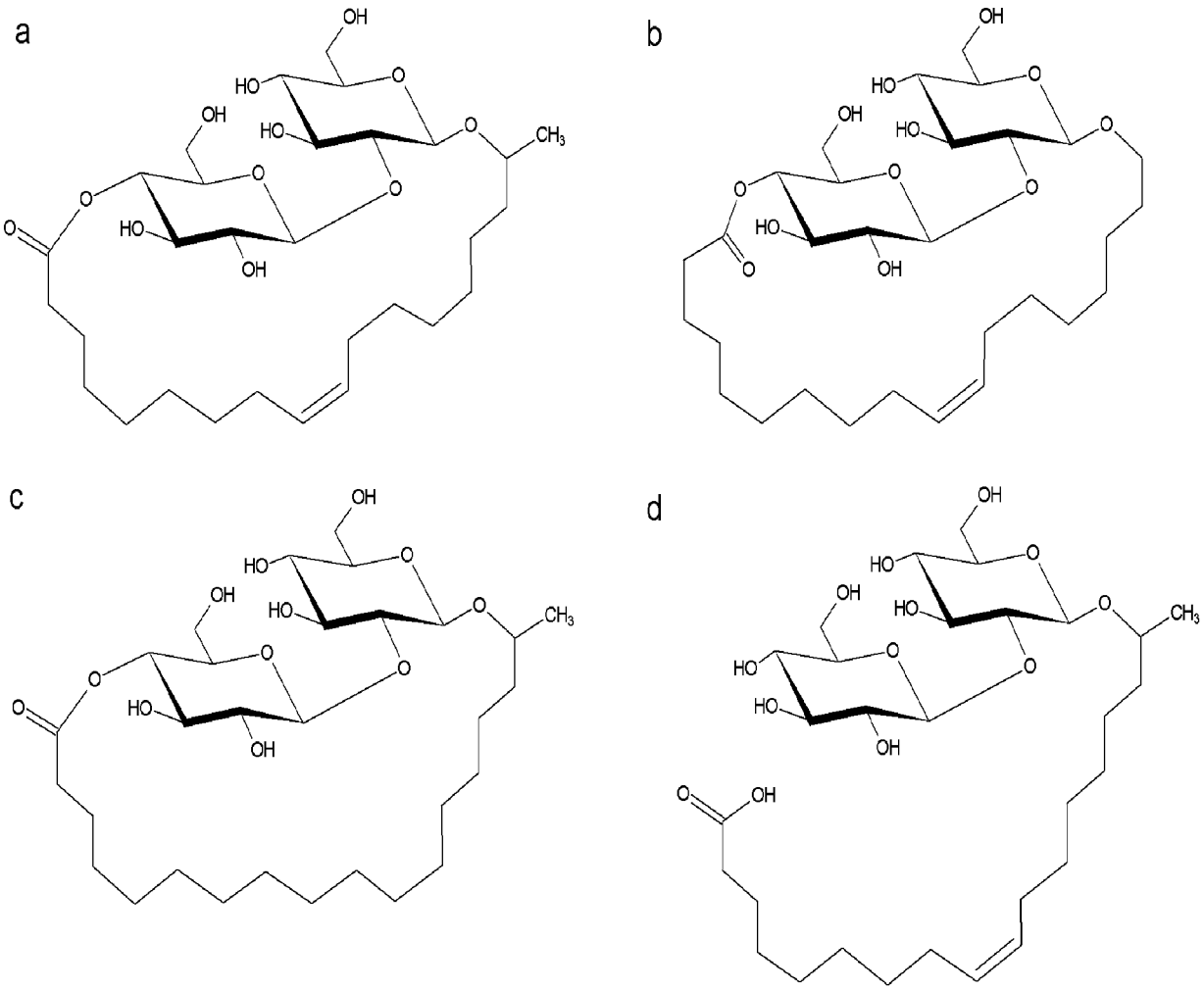
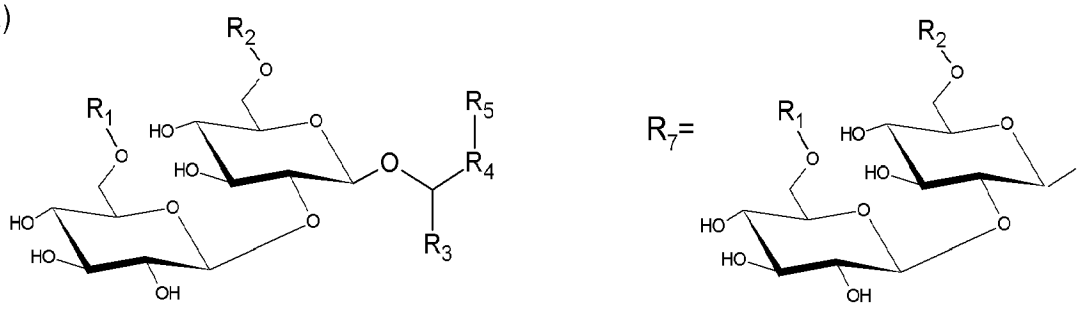
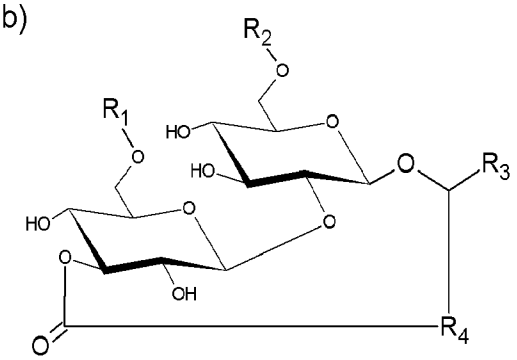


Fig. 1

(a)



(b)



R_1= -H or -C(O)CH_3

R_2= -H or -C(O)CH_3

R_3= saturated or unsaturated hydrocarbon chain with m carbon atoms

R_4= saturated or unsaturated hydrocarbon chain with n carbon atoms

R_5= -COOH, -CH_2(R_6) or -CH(R_6)CH_3

R_6= -H, -OH or OR_7

in which (m+n) >1 and <26 and m and n are independatnly of one another >=0 and <26

Fig. 2

```
   1  atggttgtaa actcctcgaa ggaccctcaa aacaaaggaa tgactcctag aaaagaaatt
        m  v  v    n  s  s    k  d  p  q    n  k  g    m  t  p    r  k  e  i
  61  gaccaggaaa tggtctcttg ggccaaaaaa aacctcaaaa acacccctgg caatgaaaac
        d  q  e    m  v  s    w  a  k  k    n  l  k    n  t  p    g  n  e  n
 121  tatgagaaga tggtctcagg agttccttac aatccatacg atccagatct tatgtttaga
        y  e  k    m  v  s    g  v  p  y    n  P  y    d  p  d    l  m  f  r
 181  gccctggcta ctagtgagaa agttagggag ttcaatacca ttgcaagtga aagtcgtact
        A  l  a    t  s  e    k  v  r  e    f  n  t    i  a  s    e  s  r  t
 241  tttgagtcaa atcacgctgc ttatatcaag aaggtcgaga ttctcaaaga cacttttggt
        f  e  s    n  h  a    a  y  i  k    k  v  e    i  l  k    d  t  f  g
 301  caaacaaagg atattgtctg gctgaccgct ccattctcag ttgattttgg attcaacatc
        q  t  k    d  i  v    w  l  t  a    p  f  s    v  D  f    g  f  n  i
 361  agcgtaggcg agcactttta cgccaacttc aacgtttgct tcttggactc ggctccaata
        s  v  g    e  h  f    Y  a  N  f    n  v  c    f  L  d    S  a  p  i
 421  atctttggtg atgaggtgat tgtagggccc aatacaacgt tcgtgactgc gactcatcct
        i  f  g    d  e  v    I  v  G  p    n  t  t    f  V  t    A  t  H  p
 481  attagccccg agaaacgtgc gaggagaatt gtgtatgctc ttcctatcaa ggtggggaat
        I  s  p    e  k  r    a  R  R  i    V  y  a    l  p  i    k  v  g  n
 541  aatgtatgga ttggtgcgaa tgtgactgtc ctgccgggtg ttacgattgg agatggctca
        n  v  W    i  G  A    n  v  t  v    L  p  g    v  t  i    g  d  g  s
 601  acaattgcgg ctggtgctgt cgttcgagaa gatgttcctc ctcgtactgt ggtgggagga
        t  i  A    A  g  a    V  v  R  E    d  v  p    p  r  t    v  v  G  G
 661  gtccctgcgc gaatcctcaa gcatattcca gaggaggatc ccgacgaggc tgaaggagag
        V  P  a    R  i  l    k  h  i  p    e  e  d    p  d  e    a  e  g  e
 721  gaactggaat tccttcttcc agttgaaatg aacgtcaata ccgctaacca gaaggtctag
        e  l  e    f  l  l    p  v  e  m    n  v  n    t  a  n    q  k  v  -
```

Fig. 3

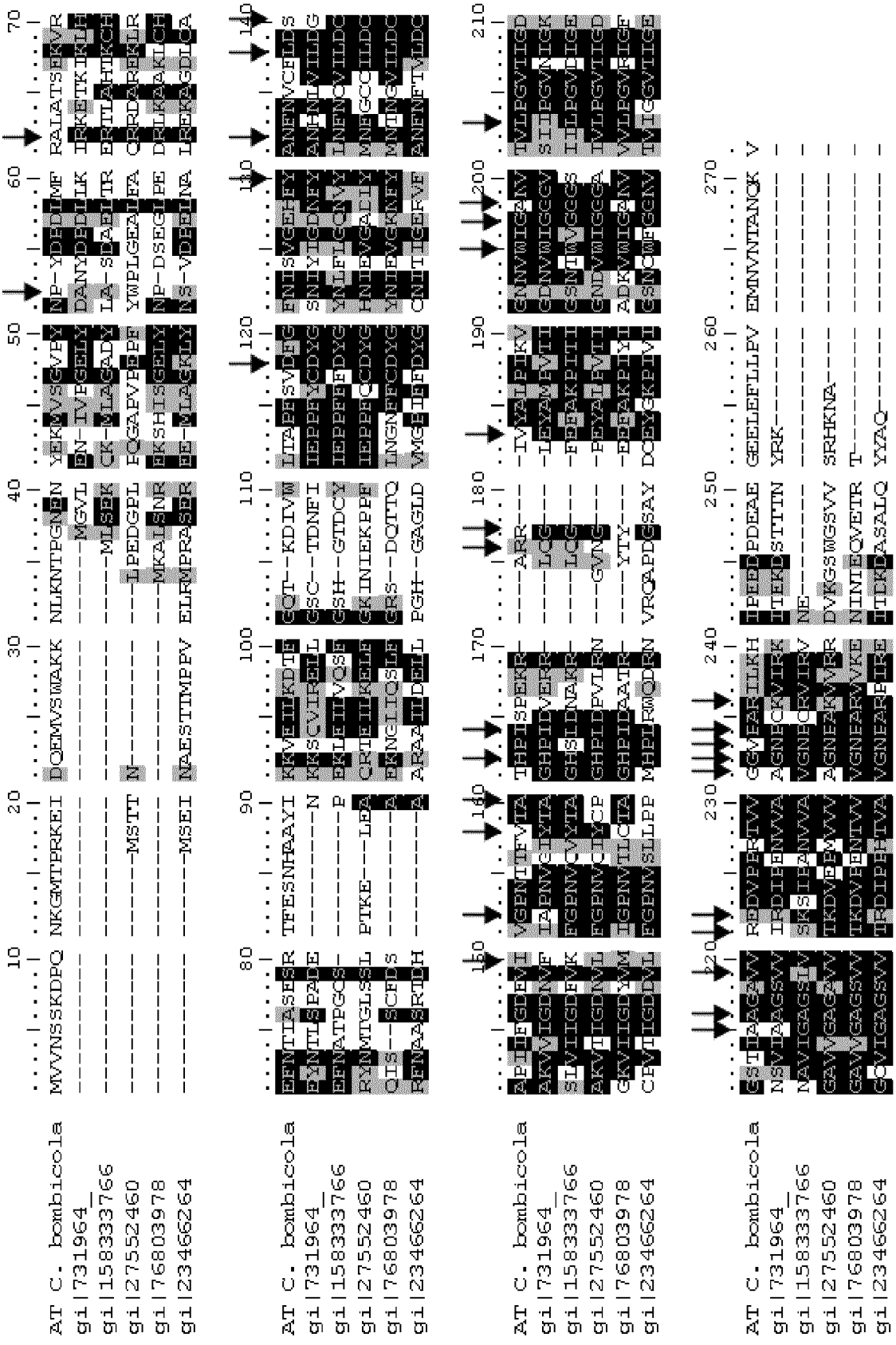
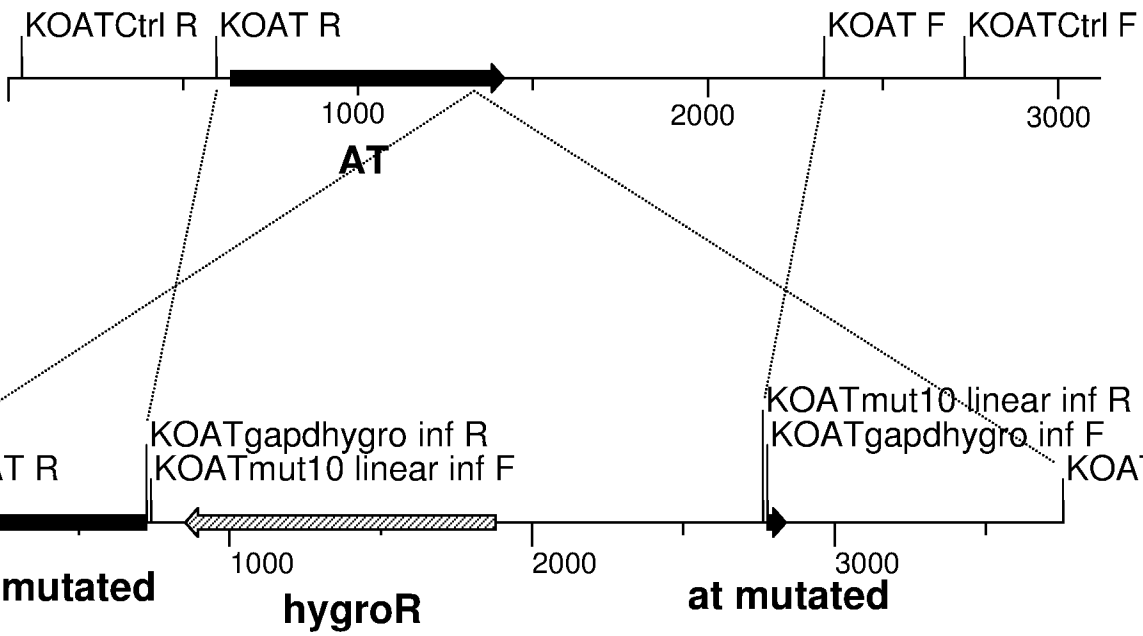

AT C. bombicola
gi|731964_
gi|158333766
gi|27552460
gi|76803978
gi|23466264

Fig. 4

A


KOATCtrl R
KOAT R
KOAT F
KOATCtrl F
1000
2000
3000
AT

B

KOAT R
KOATgapdhygro inf R
KOATmut10 linear inf F
KOATmut10 linear inf R
KOATgapdhygro inf F
KOAT F
1000
2000
3000
at mutated
hygroR
at mutated

C

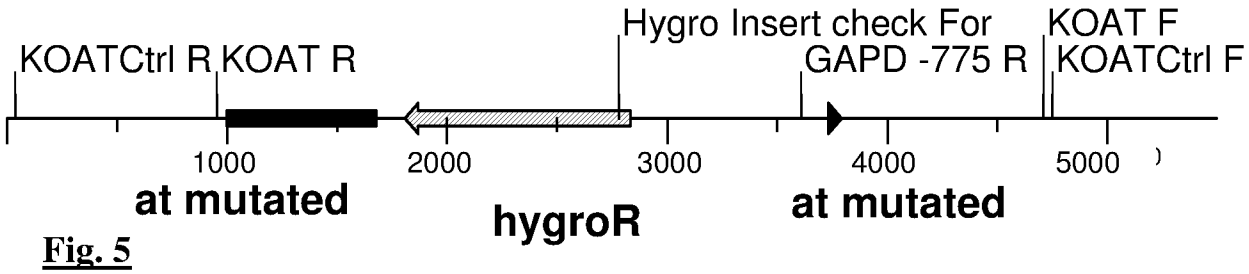

KOATCtrl R
KOAT R
Hygro Insert check For
GAPD -775 R
KOAT F
KOATCtrl F
1000
2000
3000
4000
5000
at mutated
hygroR
at mutated

Fig. 5

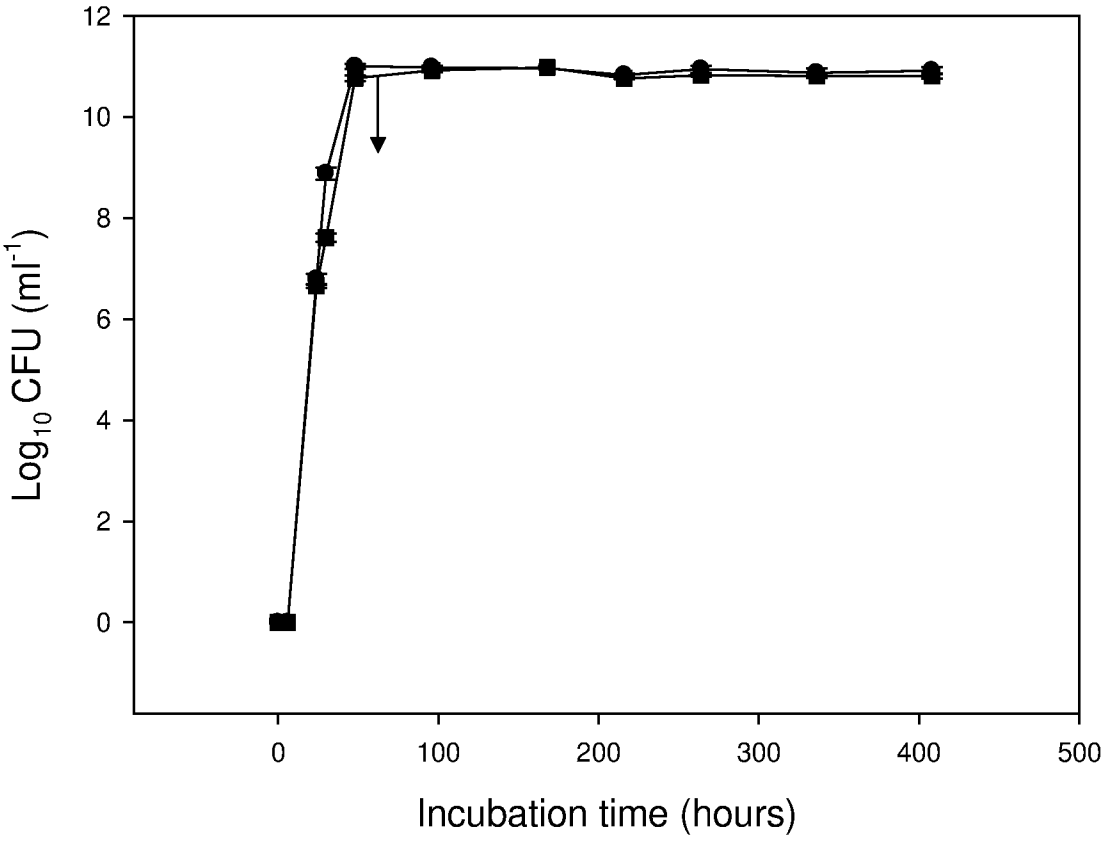

12
10
8
6
4
2
0
Log10 CFU (ml-1)
0
100
200
300
400
500
Incubation time (hours)

Fig. 6

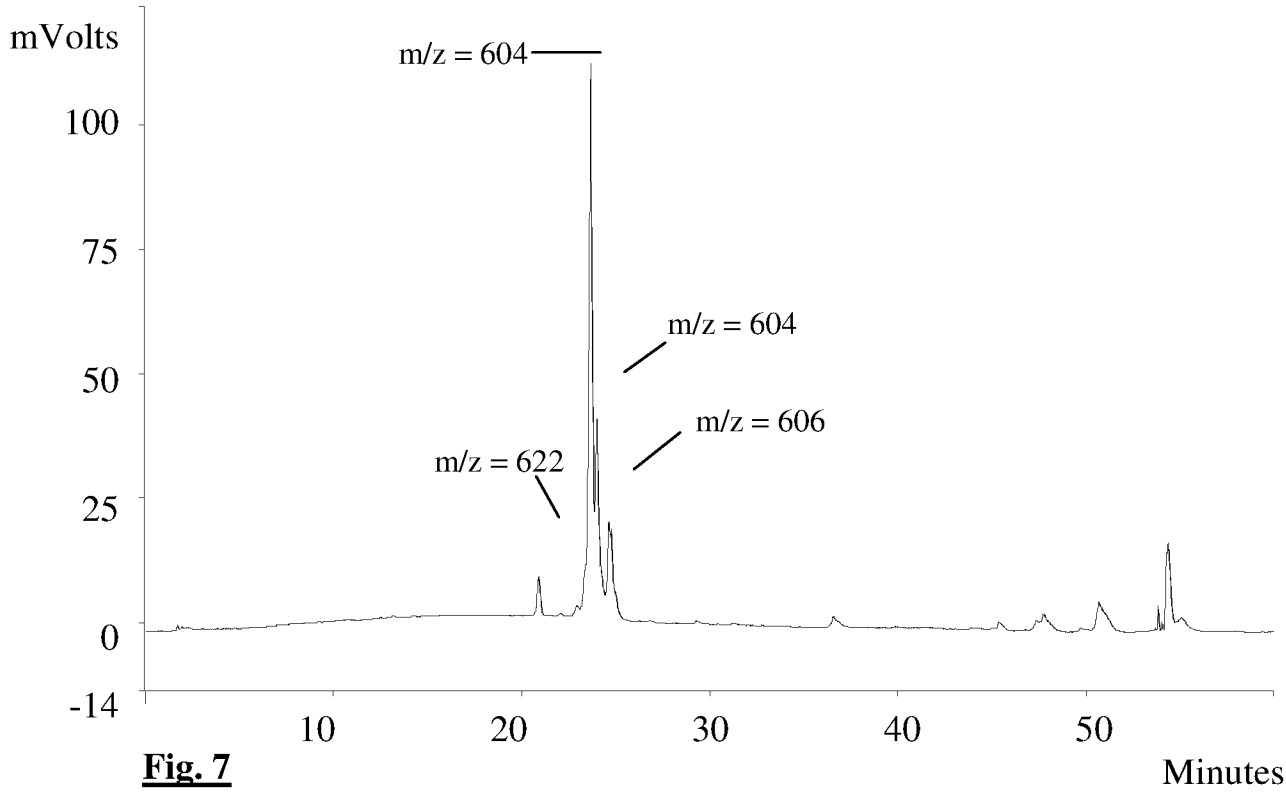

mVolts
100
75
50
25
0
-14
m/z = 604
m/z = 604
m/z = 606
m/z = 622
10
20
30
40
50
Minutes

Fig. 7

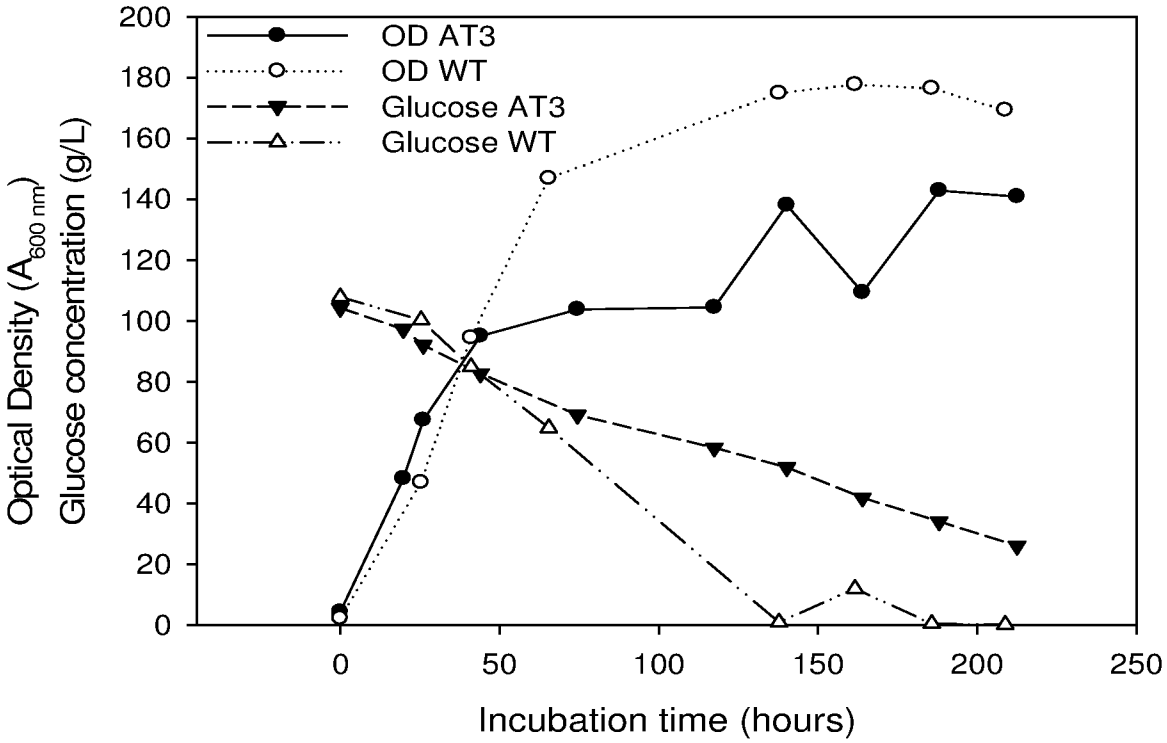

OD AT3
OD WT
Glucose AT3
Glucose WT
Optical Density ($A_{600\ nm}$)
Glucose concentration (g/L)
Incubation time (hours)
0
20
40
60
80
100
120
140
160
180
200
0
50
100
150
200
250

Fig 8

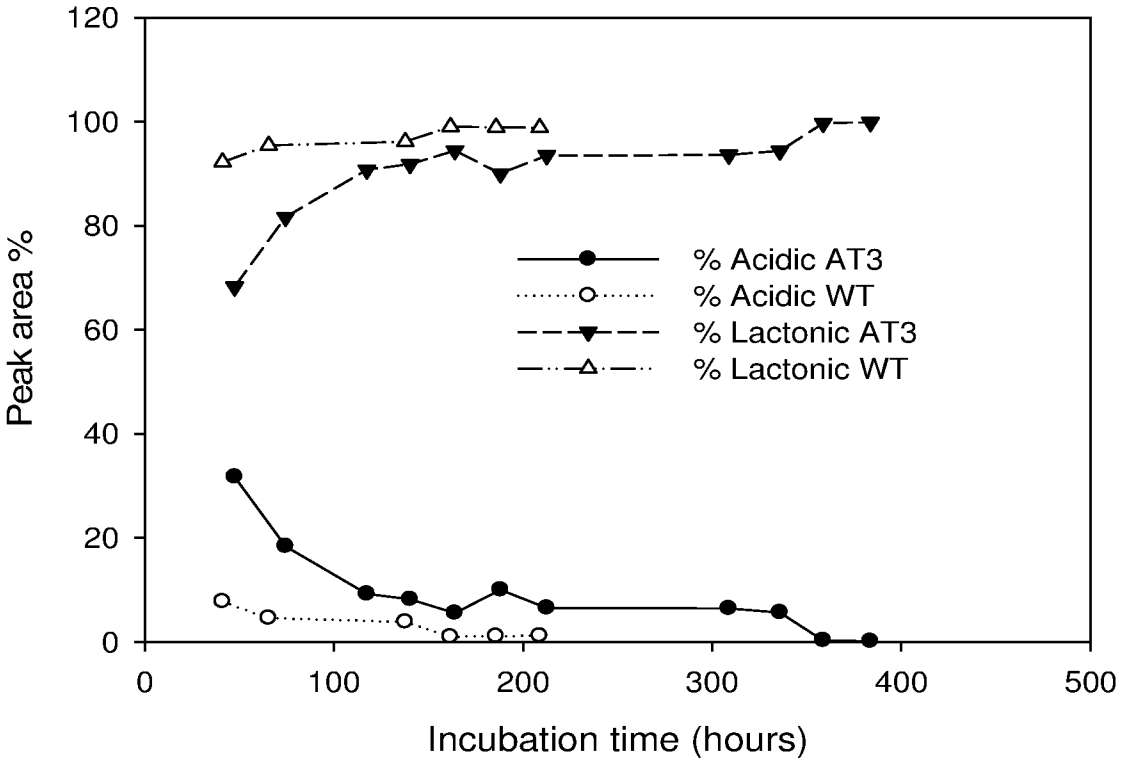

% Acidic AT3
% Acidic WT
% Lactonic AT3
% Lactonic WT
Peak area %
Incubation time (hours)
0
20
40
60
80
100
120
0
100
200
300
400
500

Fig 9

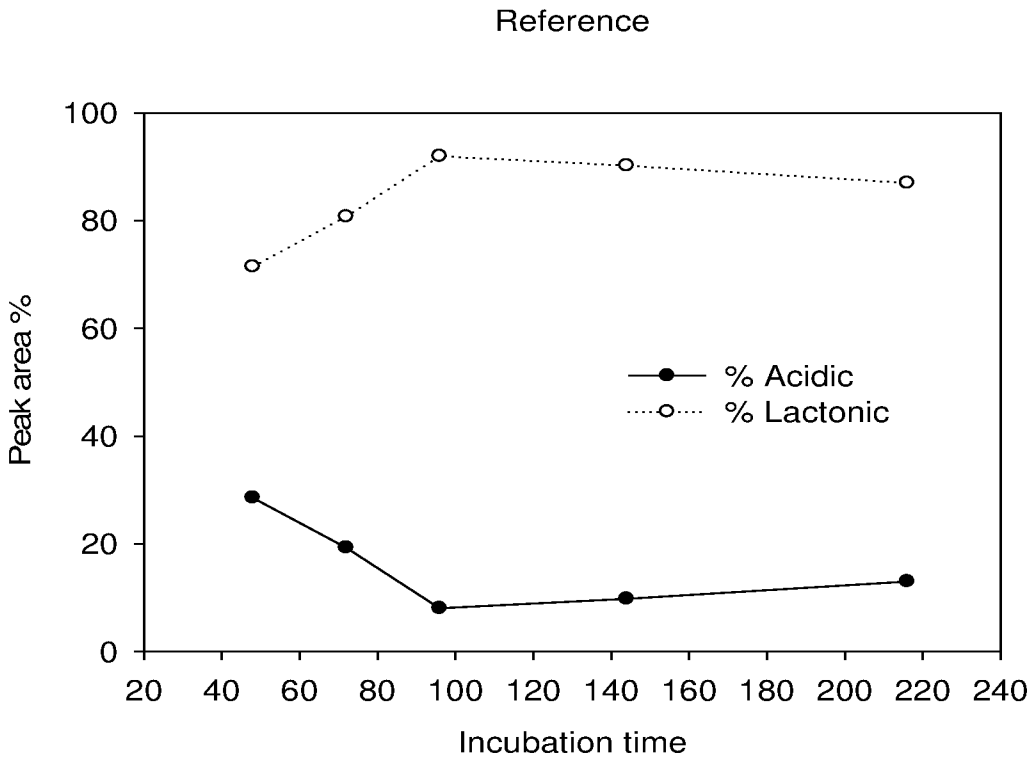


Fig 10A

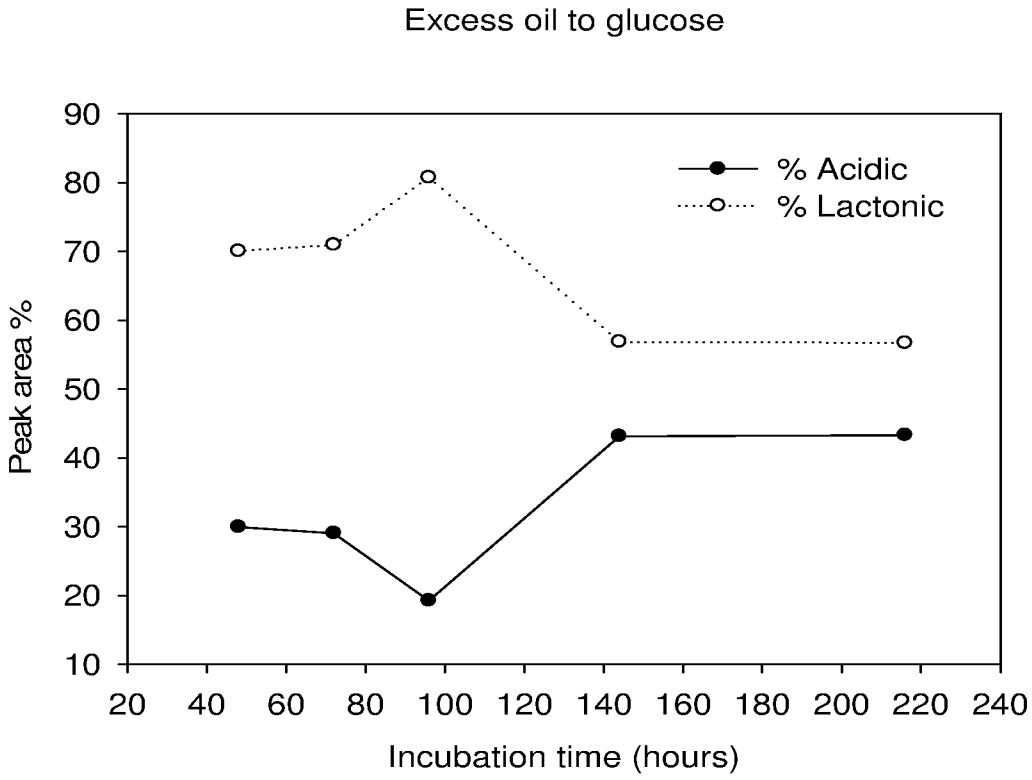


Fig 10B

YE 15g/L

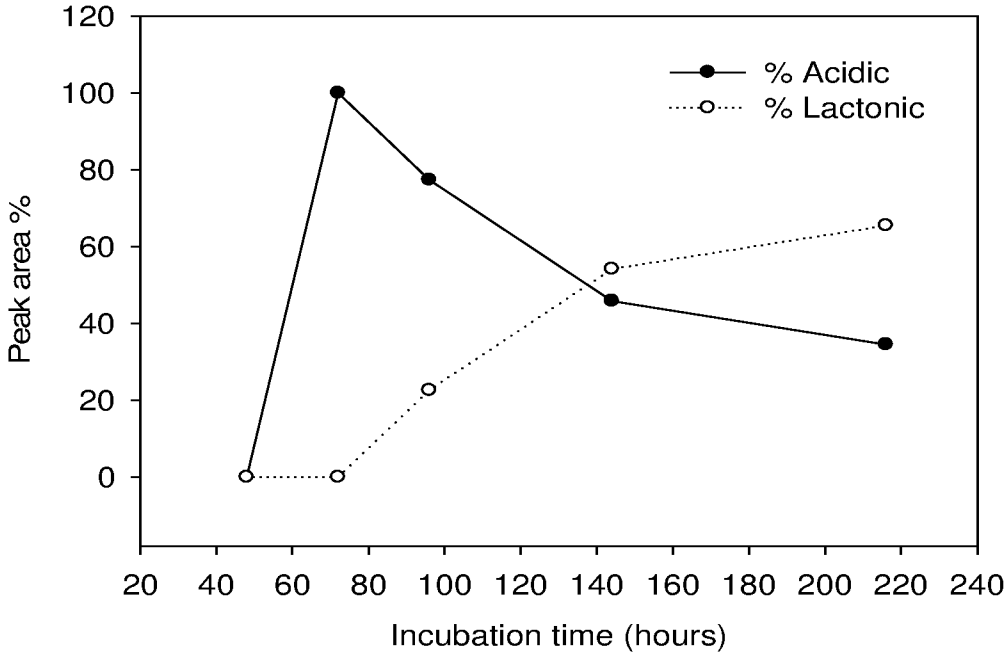


Fig 10C

Excess glucose to oil

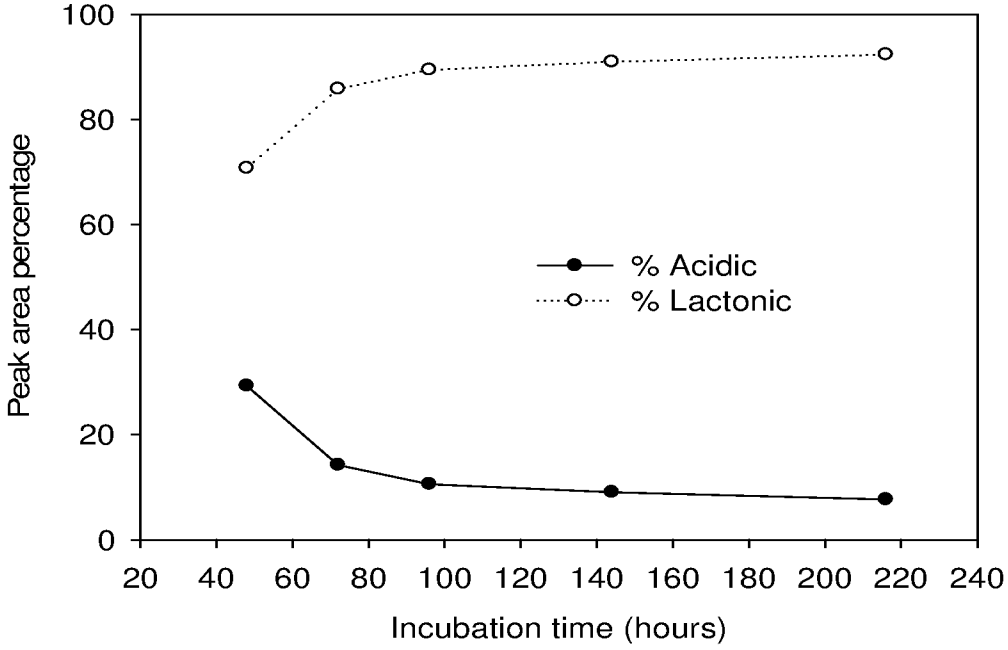


Fig 10D

YE 1g/L

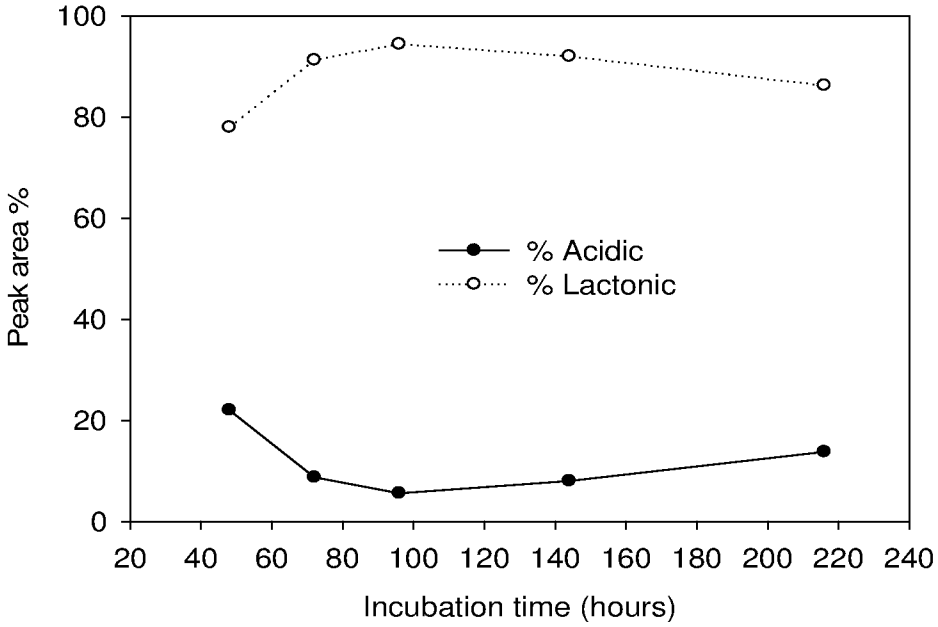


Fig 10E

No citrate

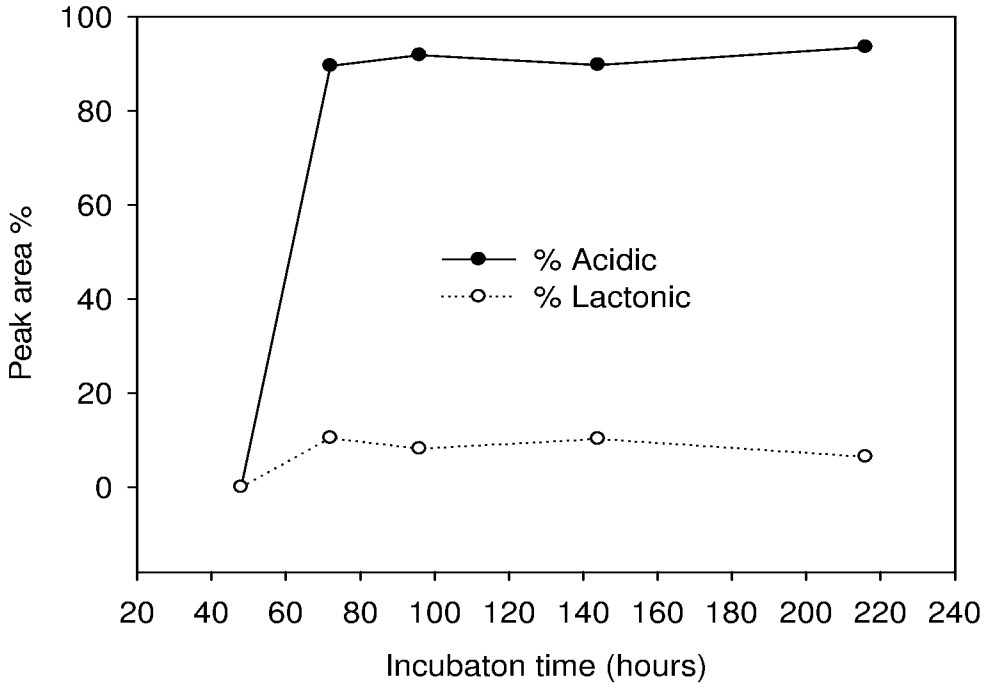


Fig 10F

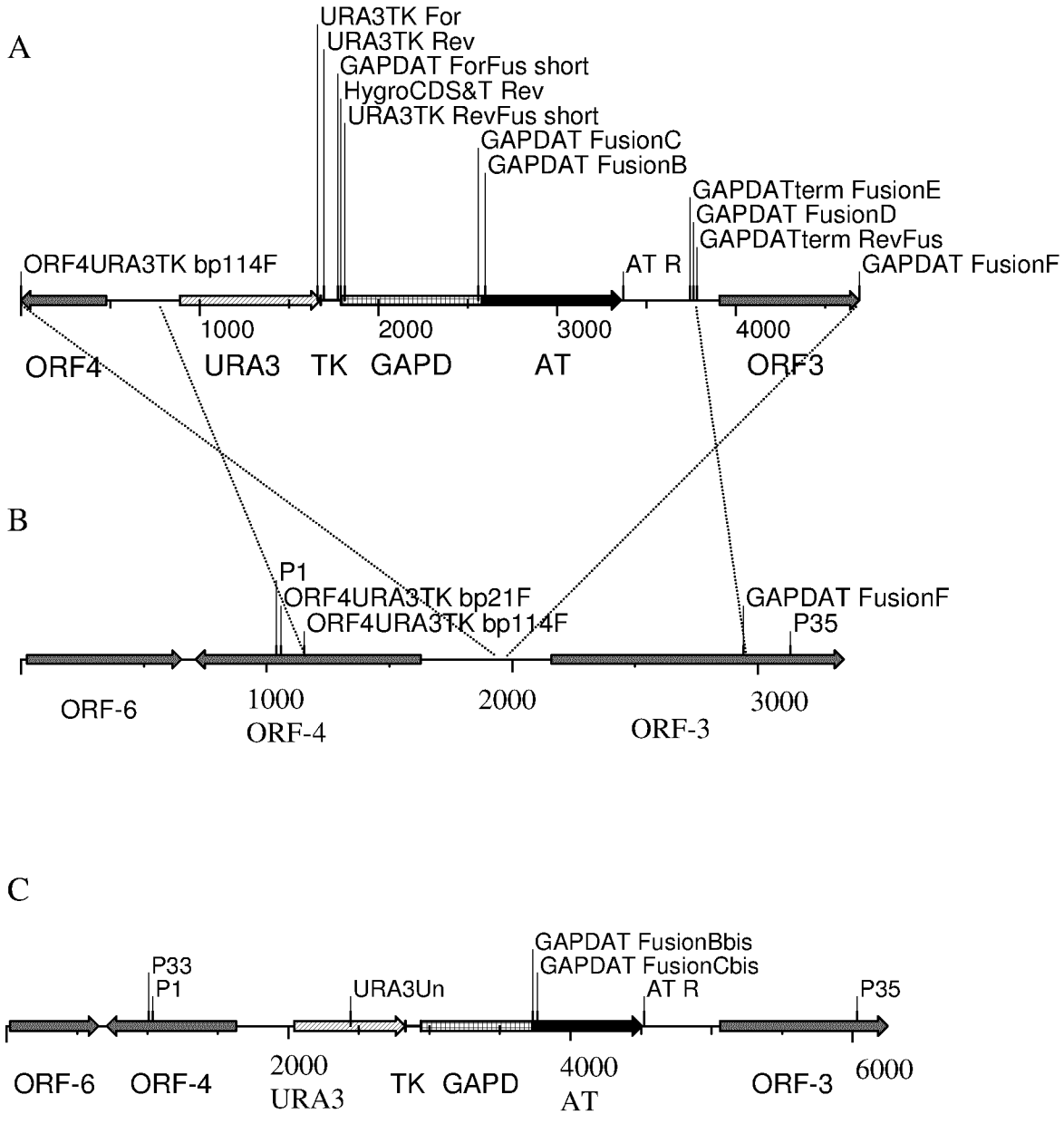

A
URA3TK For
URA3TK Rev
GAPDAT ForFus short
HygroCDS&T Rev
URA3TK RevFus short
GAPDAT FusionC
GAPDAT FusionB
GAPDATterm FusionE
GAPDAT FusionD
GAPDATterm RevFus
ORF4URA3TK bp114F
AT R
GAPDAT FusionF
1000
2000
3000
4000
ORF4
URA3
TK
GAPD
AT
ORF3
B
P1
ORF4URA3TK bp21F
ORF4URA3TK bp114F
GAPDAT FusionF
P35
ORF-6
1000
2000
3000
ORF-4
ORF-3
C
GAPDAT FusionBbis
GAPDAT FusionCbis
P33
P1
URA3Un
AT R
P35
ORF-6
ORF-4
2000
URA3
TK
GAPD
4000
AT
ORF-3
6000

Fig 11

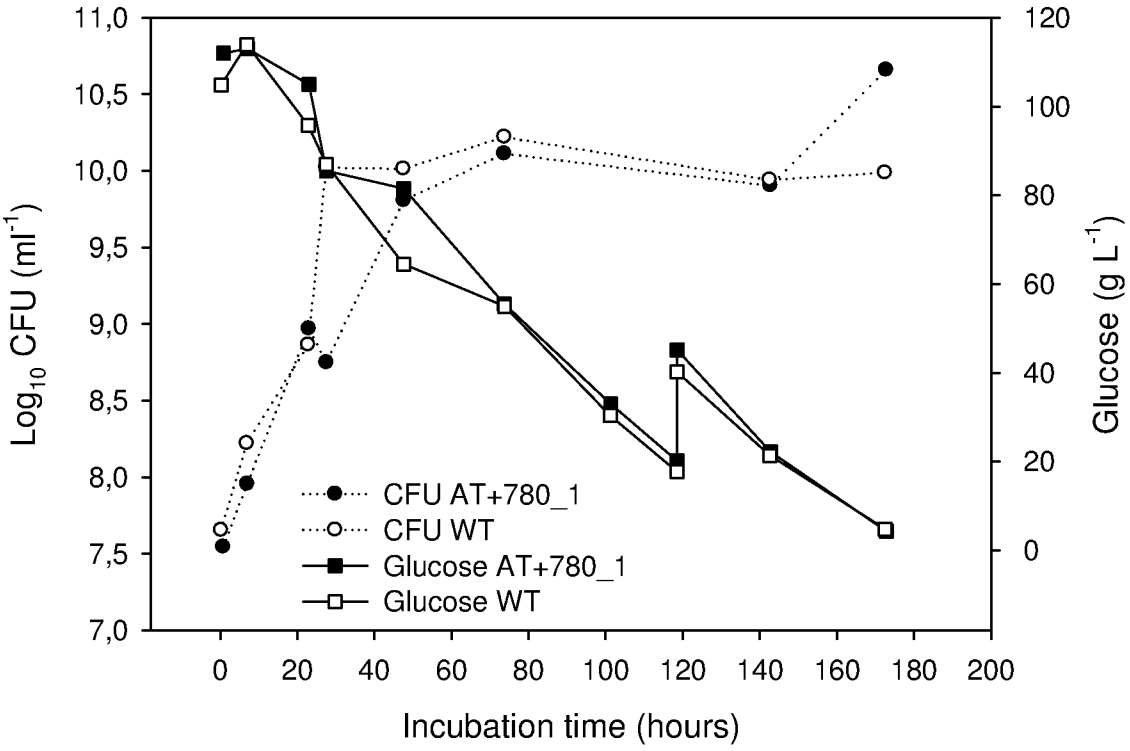

11,0
10,5
10,0
9,5
9,0
8,5
8,0
7,5
7,0
Log10 CFU (ml-1)
120
100
80
60
40
20
0
Glucose (g L-1)
CFU AT+780_1
CFU WT
Glucose AT+780_1
Glucose WT
0
20
40
60
80
100
120
140
160
180
200
Incubation time (hours)

Fig 12

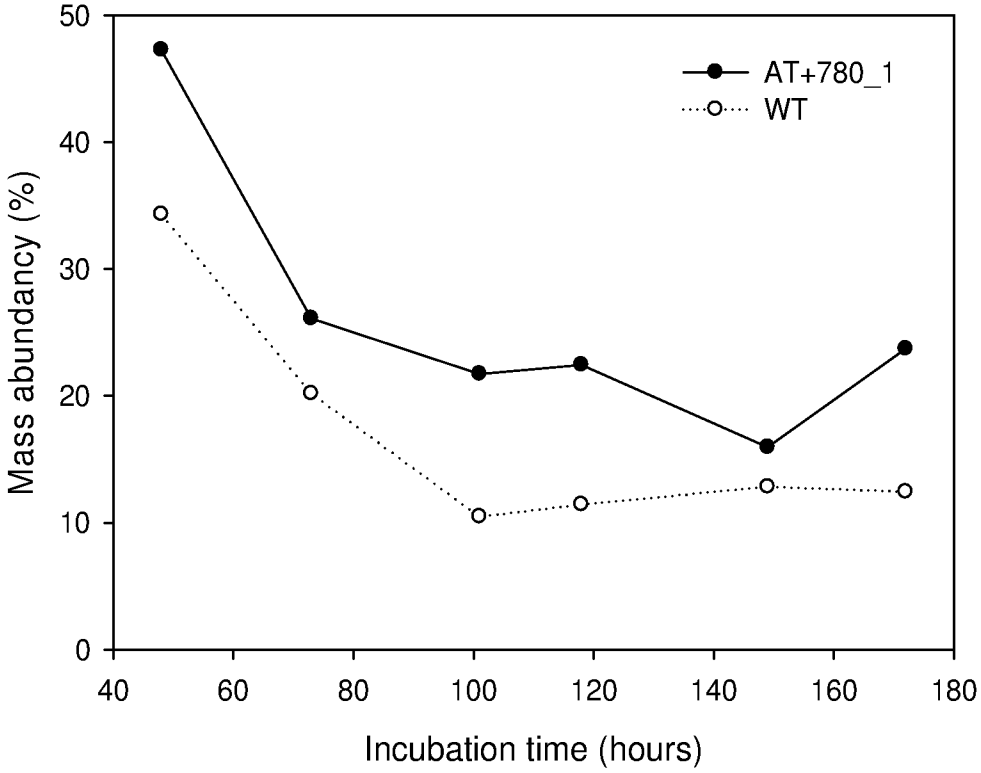

50
40
30
20
10
0
Mass abundancy (%)
AT+780_1
WT
40
60
80
100
120
140
160
180
Incubation time (hours)

Fig 13

INTERNATIONAL SEARCH REPORT

International application No
PCT/EP2011/072304

A. CLASSIFICATION OF SUBJECT MATTER
INV. C12N1/16 C12P19/44 C12N9/10
ADD.

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
C12N C12P

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)
EPO-Internal, BIOSIS, EMBASE, WPI Data

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	RAU ET AL.: "Sophorolipids: a source for novel compounds", INDUSTRIAL CROPS AND PRODUCTS, vol. 13, 8 May 2001 (2001-05-08), pages 85-92, XP002636232, cited in the application the whole document -----	1-14
A	KURTZMAN CPPRICE NPJRAY KJKUO TM: "Production of sophorolipid biosurfactants by multiple species of the Starmerella (Candida) bombicola yeast clade", FEMS MICROBIOL LETT, vol. 311, 20 October 2001 (2001-10-20), pages 140-146, XP002636233, the whole document -----	1

-/--

[X] Further documents are listed in the continuation of Box C.

[X] See patent family annex.

* Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier application or patent but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
14 May 2012	25/05/2012

Name and mailing address of the ISA/	Authorized officer
European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Fax: (+31-70) 340-3016	van Heusden, Miranda

2

Form PCT/ISA/210 (second sheet) (April 2005)

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	HU YONGMEI ET AL: "Lipase-mediated deacetylation and oligomerization of lactonic sophorolipids.", BIOTECHNOLOGY PROGRESS, vol. 19, no. 2, March 2003 (2003-03), pages 303-311, XP002636234, ISSN: 8756-7938 the whole document	1-14
A	BUCHOLTZ M L ET AL: "ACETYLATION OF 13 SOPHOROSYLOXY DOCOSANOIC-ACID BY AN ACETYL TRANSFERASE PURIFIED FROM CANDIDA-BOGORIENSIS", JOURNAL OF BIOLOGICAL CHEMISTRY, vol. 251, no. 2, 1976, pages 424-430, XP002636236, ISSN: 0021-9258 the whole document	1-14
A	ESDERS T W ET AL: "Glucosyl- and acetyltransferases involved in the biosynthesis of glycolipids from Candida bogoriensis.", THE JOURNAL OF BIOLOGICAL CHEMISTRY 10 MAR 1972 LNKD- PUBMED:5012313, vol. 247, no. 5, 10 March 1972 (1972-03-10), pages 1375-1386, XP002636235, ISSN: 0021-9258 the whole document	1-14
X,P	KAREN M.J. SAERENS ET AL: "One-step production of unacetylated sophorolipids by an acetyltransferase negative Candida bombicola", BIOTECHNOLOGY AND BIOENGINEERING, vol. 108, no. 12, 12 July 2011 (2011-07-12), pages 2923-2931, XP55027011, ISSN: 0006-3592, DOI: 10.1002/bit.23248 the whole document	1-14
X,P	WO 2011/061032 A2 (EVONIK DEGUSSA GMBH [DE]; SCHAFFER STEFFEN [DE]; WESSEL MIRJA [DE]; TH) 26 May 2011 (2011-05-26) page 3, line 13 - line 24 page 5, paragraph 3 - page 6, paragraph 1 page 13, paragraph 2 - page 14, paragraph 1 page 17, paragraph 2 - page 18, paragraph 1; examples 3,,5,8,11,13 SEQ ID NO: 9, sbg3	1-14

2

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No
PCT/EP2011/072304

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 2011061032 A2	26-05-2011	DE 102010014680 A1	18-08-2011
		WO 2011061032 A2	26-05-2011