Aerodynamically, the bumble bee shouldn't be able to fly, but the bumble bee doesn't know it so it goes on flying anyway.

**Mary Kay Ash**

Jury: Prof. Dr. ir. Norbert DE KIMPE
Prof. Dr. ir. Nico BOON
Lic. Dirk DEVELTER
Prof. Dr. ir. Monica HÖFTE
Prof. Dr. Andreas SCHMID
Prof. Dr. Els VAN DAMME
Prof. Dr. ir. Wim SOETAERT
Prof. Dr. ir. Erick VANDAMME

Promotors: Prof . Dr. ir. Erick VANDAMME
Prof. Dr. ir. Wim SOETAERT
Laboratory of Industrial Microbiology and Biocatalysis
Department of Biochemical and Microbial Technology
Ghent University

Dean: Prof. Dr. ir. Herman VAN LANGENHOVE

Rector: Prof. Dr. Paul VAN CAUWENBERGE

Ir. Inge Van Bogaert was supported by Ecover Belgium NV (Malle, Belgium) and a fellowship of the *Bijzonder Onderzoekfonds* of Ghent University (BOF).

The research was conducted at the Laboratory of Industrial Microbiology and Biocatalysis, Department of Biochemical and Microbial Technology, Ghent University.

ir. Inge Van Bogaert

# MICROBIAL SYNTHESIS OF SOPHOROLIPIDS BY THE YEAST *CANDIDA BOMBICOLA*

Thesis submitted in fulfillment of the requirements
for the degree of Doctor (PhD) in Applied Biological Sciences

Titel van het doctoraatsproefschrift in het Nederlands:
Microbiële synthese van sopohorolipiden door de gist *Candida bombicola*

Cover illustration: Cadzand on a stormy day by Inge Van Bogaert

Refer to this thesis:
Van Bogaert INA (2008) Microbial synthesis of sophorolipids by the yeast *Candida bombicola*. PhD-thesis, Faculty of Bioscience Engineering, Ghent University, Ghent, Belgium, 239 p.

ISBN-number: ISBN 978-90-5989-243-9

# Woord vooraf

Alhoewel het voorwoord meestal het laatste stukje is dat geschreven wordt, is het wel het eerste (en soms het enige) dat met volle aandacht gelezen wordt. En terecht trouwens, want hier worden even de mensen belicht die achter de wetenschappelijke tekst verscholen zitten en ze op eender welke manier hebben helpen vorm geven; een woordje van dank is hier dus wel op zijn plaats.

Ik denk hierbij in eerste instantie aan mijn promotoren, Prof. dr. ir. Erick Vandamme en Prof. dr. ir. Wim Soetaert die me alle vrijheid en vertrouwen gaven om in hun labo mijn eigen weg te zoeken en te gaan. Professor, het was me een waar genoegen om als laatste onder uw promotorschap een doctoraatstitel te behalen. En Wim, ik hoop dat LIMAB met jou nog een mooie weg mag gaan.

Uiteraard wil ik ook Prof. dr. ir. Nico Boon, Prof. dr. ir. Norbert De Kimpe, Lic. Dirk Develter, Prof. dr. ir. Monica Höfte, Prof. dr. Andreas Schmid en Prof. dr. Els Van Damme bedanken voor het kritisch beoordelen van dit proefschrift. Graag een extra woord van dank aan Dirk en Steve van Ecover Belgium NV voor de fijne samenwerking (en niet te vergeten de vele HPLC-analyses).

Het voltooien van dit doctoraat zou lang niet zo aangenaam geweest zijn zonder de (ex)-collega's. Bart liet me kennis maken met *Candida bombicola*: deze beestjes moest ik de komende jaren te vriend zien te houden en dat deed je volgens Katja het best door er af en toe eens een "klapke" tegen te doen wanneer ze aan het rondzwieren waren in een erlenmeyer of fermentor. Een half jaar later besloot ik mijn "wilde beestjes" te temmen met de moleculaire zweep; ik durf echt geen getal meer plakken op het aantal keren dat ik aan Sofies (en Marjans) bureautje gestaan heb. Gelukkig kreeg ik hulp van een aantal thesistudenten: bedankt Wouter, Marjolijn, Dries, Katrien, Ellen en Magda, zonder jullie zou dit doctoraat niet zijn wat het nu is.
En ja, soms moest er worden overgeschakeld op een plan B of zelfs een plan C, en vooral in de laatste rechte lijn liepen de frustraties soms hoog op door elektriciteitspannes,

boterzuurgeuren en defecte toestelen, deze werden gelukkig verzacht door de koffiepauzes, traktaties, een luisterend oor of een helpende hand: bedankt lieve collegaatjes! Dank ook aan mijn “speed-leescomité” Julia, Lieve, Marjan en Sofie om met zoveel enthousiasme mijn schrijfsels te verbeteren.

Uiteraard wil ik ook mijn ouders bedanken voor hun steun en geloof, en om er steeds te zijn voor mij. Onder het motto “All work and no play makes Jack a dull boy”, ben ik ook een woordje van dank verschuldigd aan mijn zus Julie, Stefan en Marleen, Marc, Eefje en de “oud-studiegenoten”. Meer dan één woordje dank gaat uit naar Peter: dankjewel om me te blijven steunen en opkoteren, om te verdragen dat dit doctoraat soms heel wat aandacht kreeg en je hebt gelijk: “Dat doctoraat kreeg me soms kwaad”.

Inge,

Mei 2008

# Table of contents

# Abbreviations

| | |
|---|---|
| 5-FOA | 5-fluoroorotic acid |
| AA | Amino Acids |
| Acetyl-CoA | Acetyl-Coenzyme A |
| Acyl-CoA | Acyl-Coenzyme A |
| APG | AlkylPolyGlucosides |
| APS1 | Adaptor Primer 1 |
| APS2 | Adaptor Primer 2 |
| ARL1 | ADP-Ribosylation factor-like 1 protein |
| ARS | Autonomous Replication Sequence |
| ATCC | American Type Culture Collection (Rockville, Maryland, USA) |
| CBS | Centraalbureau voor Schimmelcultures (Baarn-Delft, The Netherlands) |
| CCRC | Culture Collection and Research Centre (Taiwan) |
| cDNA | Copy DeoxyriboNucleic Acid |
| CDS | Coding Sequence |
| CDW | Cell Dry Weight |
| CFU | Colony Forming Units |
| CMC | Critical Micelle Concentration |
| CPR | NADPH Cytochrome P450 Oxidoreductase |
| DMSO | DiMethyl SulfOxide |
| DNA | DeoxyriboNucleic Acid |
| EDTA | Ethylene Diamine Tetra Acetate |
| ELSD | Evaporative Light Scattering Detection |
| ER | Endoplasmatic Reticulum |
| FAD | Flavin Adenine Dinucleotide |
| FAS | Fatty Acid Synthase |
| FMN | Flavin MonoNucleotide |
| GAPD | GlycerAldehyde-3-Phosphate Dehydrogenase |
| GC | Gas Chromatography |
| gDNA | Genomic DeoxyriboNucleic Acid |
| GPS | Gene Specific Primer |
| HAc | Acetic Acid |
| HLB | Hydrophilic-Lipophilic Balance |
| HPLC | High Performance Liquid Chromatography |
| HRT3 | High level expression Reduces Ty3 transposition protein 3 |
| IFO | Institute for Fermentation (Osaka, Japan) |
| JCM | Japan Collection of Microorganisms (Hirosawa, Wako, Japan) |
| LC | Liquid Chromatography |
| LCMS | Liquid Chromatography Mass Spectroscopy |
| MEL | MannosylErythritol Lipids |
| MFE-1 | MultiFunctional Enzyme type 1 |
| MFE-2 | MultiFunctional Enzyme type 2 |
| MIC | Minimum Inhibitory Concentration |
| min | mintutes |
| MS | Mass Spectroscopy |

| | |
|---|---|
| NADPH | Nicotinamide Adenine Dinucleotide Phosphate |
| n.d. | Not determined |
| ND | Not Detected |
| NRIC | Nodai Research Institute Culture Collection (Tokyo, Japan) |
| NRRL | Northern Regional Research Laboratory (Peoria, Illinois, USA) |
| o/n | OverNight |
| OD | Optical Density |
| ORF | Open Reading Frame |
| P450 | Cytochrome P450 Monooxygenase |
| PCR | Polymerase Chain Reaction |
| PEG | Polyethylene Glycol |
| POX | Acyl-CoA dehydrogenase |
| PTS | Peroxisomal Targeting Signals |
| RTF1P | RNA polymerase II Transcription elongation Factor 1 Protein |
| RNA | RiboNucleic Acid |
| rpm | Revolutions Per Minute |
| s | seconds |
| SD | Synthetic Dextrose |
| TDP | Thymidine DiPhosphate |
| TE | 10 mM Tris-Cl, 1 mM EDTA, pH 7.5 adjusted with HCl |
| TK | Tyrosine Kinase |
| URA3 | orotidine 5'-phosphate decarboxylase |
| vvm | Volumes of air per Volume of batch per Minute |
| YNO | Yeast Nitrogen base with Oleic acid |
| YPD | Yeast Peptone Dextrose |

# Introduction

Sophorolipids have been known for over 40 years as biological surface active compounds or surfactants. They can be produced by a selected number of yeast strains starting from renewable resources or even waste streams, dispose low ecotoxicity and are biodegradable. These features make them particularly attractive in our current society with growing environmental awareness. Millions of tons of surfactants are yearly used in a wide range of sectors (cleaning, chemical, textile, food and paper industry, cosmetics, personal and health care, agriculture, etc.), but the large majority of them is produced by chemical means. Especially in washing applications they inevitably end up in the environment where they are not always that readily degraded.

Sophorolipids synthesized by the yeast *Candida bombicola* are one of the most promising biosurfactants; they are synthesized by nonpathogenic yeasts, can be produced at high concentrations and find applications in various sectors due to either their emulsifying, antimicrobial or other beneficial properties.

Sophorolipids are composed of the disaccharide sophorose to which a hydroxy fatty acid is linked. The carbon chain length of these hydroxy fatty acids is limited to 16 to 18 carbon atoms. However, to increase the hydrophilic-lipophilic balance and the foaming capacities of the molecules, it would be interesting to obtain sophorolipids with a shorter fatty acid tail of 8 to 14 carbon atoms. These can on one hand be obtained by using unconventional hydrophilic carbon sources such as hydroxylated substrates or substrates with internal ester bonds which are post-fermentative modified.

On the other hand, medium-chain sophorolipids can be obtained via the genetic engineering of the producing yeast strain. Since no such research has been reported till now, a transformation and selection system for *C. bombicola* was developed. In order to achieve this, the *URA3* gene was isolated from the totally unknown *C. bombicola* genome. Several other genes were also isolated, among them the *GAPD* gene who's promoter sequence can be used for (heterologous) expression of genes in *C. bombicola*, and the *MFE-2* gene. The corresponding enzyme of this latter gene takes part in the β-oxidation pathway and knocking out of this gene resulted in improved medium-chain sophorolipid production on unconventional substrates.

Cytochrome P450 monooxygenases play a crucial role in sophorolipid synthesis: they control the chain length of the sophorolipid hydroxy fatty acid tail when using conventional substrates. Several of these genes were isolated from the *C. bombicola* genome and two of them were assigned to new families. It is expected that not all of them take part in sophorolipid synthesis, but one cytochrome P450 monooxygenase shows a strong induction upon sophorolipid synthesis; it can therefore be suggested that this enzyme contributes to the formation of sophorolipids.

This PhD-thesis presents the exciting first steps in the genetic engineering and in-depth study of the sophorolipid synthesis pathway of *C. bombicola*. The development of molecular tools and the use of unconventional substrates made it possible to produce medium-chain sophorolipids and gave better insights in the involved biochemical pathways.
Furthermore, the presented work opens up perspectives for future and long term research on sophorolipid synthesis and *C. bombicola*.

# Chapter I: Literature Review on microbial production and application of the biosurfactant sophorolipids

Part of this chapter has been published as:

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

# 1 Introduction

Sophorolipids are surface-active compounds synthesized by a selected number of yeast species. They have been known for over 40 years, but because of growing environmental awareness, they recently regained attention as biosurfactants due to their biodegradability, low ecotoxicity, and production via fermentation processes based on renewable resources. In this chapter, an introduction on surfactants and biosurfactants is presented, followed by a more specific part on sophorolipids.

In terms of production volume, surfactants belong to the most important classes of industrial chemicals with a current total world production of about 10 million ton per year. About half that volume is used in household and laundry detergents, the other half in a wide variety of industrial sectors, particularly the chemical, textile, food, and paper industry, cosmetics, personal, and health care, agriculture, etc.
The large majority of the currently used surfactants are petroleum-based and are produced by chemical means. These compounds are often toxic to the environment, and their use may lead to significant ecological problems, particularly in washing applications as these surfactants inevitably end up in the environment after use (Mann and Bidwell, 2001; Mann and Boddy, 2000). The ecotoxicity, bioaccumulation, and biodegradability of surfactants are therefore issues of increasing concern.

Glycolipid surfactants are composed of a carbohydrate head and a lipid tail. They are a class of nonionic surfactants that has significantly increased its market share during the last 10 years. They offer a vastly improved environmental compatibility as compared to traditional surfactants, combined with excellent functional properties. Whereas the first generation of glycolipids was produced from renewable resources through chemical means, *e.g.* alkylpolyglucosides (APG), the second generation of glycolipids is obtained from renewable resources through biotechnological means; indeed, glycolipids produced by fermentation are now entering the market, particularly rhamnolipids and sophorolipids. The latter molecules consist of the dimeric sugar sophorose linked to a long chain hydroxy fatty acid. They posses good surface active properties and show excellent skin compatibility, a property that is very

important for cosmetic and personal care applications. Furthermore, they can be used in various other sectors due to either their emulsifying, antimicrobial, or other beneficial properties. Sophorolipids are synthesized in high concentrations by nonpathogenic yeasts (in contrast to rhamnolipids, where the most productive strains are bacteria belonging to the species *Pseudomonas aeruginosa*). This fact makes sophorolipids particularly attractive for further commercial production and use.

# 2 Surfactants

## 2.1 General use and applications

Surfactants are surface-active compounds capable of reducing interfacial tension between liquids, solids and gases, thereby allowing them to mix and disperse readily in water or other liquids. Surfactants are amphiphilic molecules consisting of a hydrophilic and a hydrophobic moiety that interacts with the phase boundary in heterogeneous systems. The non-polar "tail" is typically a hydrocarbon chain whereas the polar "head" appears in many different varieties such as carbohydrates, alcohol alkoxylates, amino acids, carboxylates, sulphates, sulphonates and phosphates.

Surfactants are used for a wide variety of applications in households, industry and agriculture. They are extensively used in cleaning applications and as a formulation aid to promote solubilisation, emulsification and dispersion of other molecules in products ranging from chemicals, cosmetics, detergents, foods, textiles and pharmaceuticals. Surfactants are molecules that intervene in nearly every product and every aspect of human daily life.

In addition to their use as a formulation aid, certain surfactants can also be used as an active compound with antimicrobial, antitumor, antiviral or immunological properties or as inducers of cell differentiation. This has resulted in a number of potential applications and related developments in biomedical sciences (Rodrigues *et al.*, 2006; Singh and Cameotra, 2004). Also in plant protection, apart from their general use as a formulation and dispersion aid, certain surfactants are actually the active ingredient. Biosurfactants such as rhamnolipids are known to have very high and specific antimicrobial activity against the zoospores of *Phytophtora*, one of the most important phytopathogenic fungi (De Jonghe *et al.*, 2005).

Surfactants are one of the most important classes of industrial bulk chemicals with a total world production exceeding 10 million ton per year. About half of that is used in household and laundry detergents, the other half in a wide variety of industrial sectors, particularly the chemical industry, food industry, cosmetics and personal care, textile industry, health care, paper industry and agriculture.

The large majority of the currently used surfactants are petroleum-based and are produced by chemical means. These compounds are often toxic to the environment and their use may lead to significant environmental problems, particularly in washing applications as these surfactants inevitably end up in the environment after use. The ecotoxicity, bio-accumulation and biodegradability of surfactants are therefore issues of increasing concern. This is easily demonstrated by many reports describing a low rate of biodegradation and a high aquatic toxicity of some traditional surfactants (Mann and Bidwell, 2001; Mann and Boddy, 2000).

## 2.2 Characteristics related to surfactants

The activity of a surfactant is in general evaluated by its ability to lower the surface tension between water and air (72.80 mN/m). Surface tension is represented by the symbols $\sigma$ or $\gamma$ and is defined as the force F along the line with length L, where the force is parallel to the surface, but perpendicular to the line ($\gamma = F/L$). The SI unit of surface tension is N/m.
Another important parameter concerning the performance of a surfactant is its critical micelle concentration (CMC, Figure I-1).

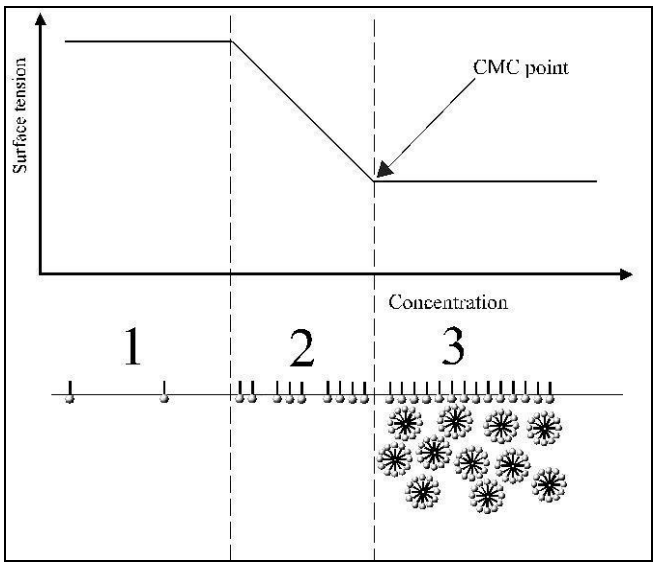


**Figure I-1**: Surfactant concentration in relation to surface tension. The CMC is reached when micelles start to form.

An amphiphilic molecule can arrange itself at the surface of the water such that the polar part interacts with the water and the non-polar part is pushed above the surface. The presence of these molecules on the surface disrupts the cohesive energy at the surface and thus lowers the surface tension. Upon gradual addition of the surfactant to water, the surface tension will decrease until the CMC is reached. Beyond this point, further increase of the surfactant concentration will cause no further decline in surface tension, since aggregates or micelles are formed. In those, the hydrophobic residues are oriented within the micelles and the hydrophilic parts are exposed to the solvent.

A third parameter for surfactant evaluation is the Hydrophilic-Lipophilic Balance (HLB). It indicates if the surfactant will stabilize an oil-in-water emulsion (high HLB) or rather stabilize a water-in-oil emulsion (low HLB).

The HLB can be determined theoretically according to the formula of Griffin (1954):

$$HLB = 20 * Mh / M$$

with

- Mh - Molecular mass of the hydrophilic portion of the molecule
- M - Molecular mass of the whole molecule.

An HLB value of 0 corresponds to a completely hydrophobic molecule, and a value of 20 would correspond to a molecule made up completely of hydrophilic components.

In 1957, Davies improved the formula by taking into account the effect of strong and less strong hydrophilic groups (Davies, 1957). All chemical groups are assigned a specific value. The formula is:

$$HLB = 7 + m * Hh + n * Hl$$

with:

- m - Number of hydrophilic groups in the molecule
- Hh - Value of the hydrophilic groups
- n - Number of lipophilic groups in the molecule
- Hl - Value of the lipophilic groups

The HLB value can be used to predict the surfactant properties of a molecule:
A value from 1 to 3 indicates an anti-foaming agent
A value from 3 to 6 indicates a W/O emulsifier
A value from 7 to 9 indicates a wetting agent
A value from 8 to 12 indicates an O/W emulsifier
A value from 12 to 15 is typical of detergents
A value from 15 to 20 indicates a solubiliser.

Oleic acid for example has an HLB of 1, the values for polyethylene lauryl ether (Brij 30) and polyoxyethylene sorbitan mono-oleate (Tween 80) are 9 and 15 respectively, while potasium oleate has an HLB value of 20.

# 3 Biosurfactants

## 3.1 Introduction

Surfactants intervene in nearly every product and every aspect of human daily life. The large majority of the currently used surfactants are produced by chemical means from fossil resources such as petroleum. However, during the last decades, the environmental awareness has become a more important issue in the study, development and application of surfactants. In this respect, microorganisms are a potential source of (novel) biodegradable surfactants or biosurfactants. With the advantages of biodegradability, low ecotoxicity and the fermentative production on renewable-resource substrates, biosurfactants may eventually replace their chemically synthesized counterparts.

The structure of biosurfactants is predominantly determined by the producing organism, but can to a certain extent be influenced by the culture conditions. Biosurfactants can be classified in four groups based on their chemical composition: glycolipids (1), oligopeptides and lipopeptides (2), phospholipids, fatty acids and neutral lipids (3) and polymeric biosurfactants (4). In addition to those four basic groups, there also exists biosurfactants build of carbohydrates, fatty acids and peptides, and sometimes external cell components or even whole cells show surface tension lowering properties. Table I-1 lists the most important microbial surfactants.

As the presented work focuses here on the glycolipids sophorolipids, this group will be discussed more in detail.

**Table I-1**: Microbial source and properties of important types of microbial surfactants (after Desai and Banat, 1997). n.d. = not determined.

| Biosurfactant | Organism | Surface tension (mN/m) | CMC (mg/L) |
|---|---|---|---|
| Glycolipids | | | |
| Rhamnolipids | *Pseudomonas aeruginosa* | 29 | 15-100 |
| | *Pseudomonas* sp. | 25–30 | 0.1-10 |
| Trehalolipids | *Rhodococus. erythropolis* | 32–36 | 4 |
| | *Nocardia erythropolis* | 30 | 20 |
| | *Mycobacterium* sp. | 38 | 0.3 |
| Sophorolipids | *Candida bombicola* | 33 | 40-100 |
| | *C. apicola* | 30 | 40-100 |
| | *Rhodotorula bogoriensis* | n.d. | n.d. |
| Mannosylerythritol lipids | *Pseudozyma antarctica* | 35 | 10 |
| Cellobiolipids | *Ustilago zeae, U. maydis* | n.d. | n.d. |
| Lipopeptids and lipoproteins | | | |
| Peptide-lipid | *Bacillus licheniformis* | 27 | 12-20 |
| Serrawettin | *Serratia marcescens* | 28–33 | n.d. |
| Viscosin | *Ps. fluorescens* | 26.5 | 150 |
| Surfactin | *B. subtilis* | 27-32 | 23-160 |
| Gramicidins | *B. brevis* | n.d. | n.d. |
| Polymyxins | *B. polymyxa* | n.d. | n.d. |
| Fatty acids, neutral lipids and phospholipids | | | |
| Fatty acids | *C. lepus* | 30 | 150 |
| Neutral lipids | *N. erythropolis* | 32 | n.d. |
| Phospholipids | *Thiobacillus thiooxidans* | n.d. | n.d. |
| Polymeric surfactants | | | |
| Emulsan | *Acinetobacter calcoaceticus* | n.d. | n.d. |
| Biodispersan | *A. calcoaceticus* | n.d. | n.d. |
| Mannan-lipid-protein | *C. tropicalis* | n.d. | n.d. |
| Liposan | *C. lipolytica* | n.d. | |
| Carbohydrate-protein-lipid | *P. fluorescens* *Debaryomyces polymorphis* | 27 | 10 |
| Protein PA | *Ps. aeruginosa* | n.d. | n.d. |
| Particulate biosurfactants | | | |
| Vesicles en fimbriae | *A. calcoaceticus* | n.d. | n.d. |
| Whole cells | *Variety of bacteria* | n.d. | n.d. |

## 3.2 Screening for biosurfactants

Microorganisms synthesizing biosurfactants are in general able to grow on hydrophobic substrates such as oils or hydrocarbons. Most screening programs therefore focus on those niches, although a lot of interesting biosurfactant producing species are isolated from totally different sources. A quick and easy way to screen for biosurfactants is the use of blood plates:

colonies producing biosurfactants will cause clearing zones on those plates because of lysis of the erythrocytes.

One can also screen by using the surface lowering properties of the biosurfactant: a droplet of water put on a hydrophobic surface will collapse and spread when a surfactant producing cell culture is added (Hildebrand, 1989). Van der Vegt *et al.* (1991) improved this method by developing the axisymmetric drop shape analysis: a drop culture broth is placed on a fluoro-ethylene-propylene surface and the profile of the droplet is determined with a contour monitor. The surface tension can be calculated from the recorded profile. There also exist several colorometric detection methods, but those are often specific for a certain type of surfactant. Shulga *et al.* (1993) developed a method for the estimation of anionic surfactants based on the reaction with a cationic substrate to form a coloured complex. There also exists a direct thin-layer chromatographic technique for rapid screening and characterisation of promising microorganisms (Matsuyama *et al.*, 1987).

## 3.3 Glycolipids

Glycolipids exist of one or more carbohydrates in combination with one or more fatty acids, hydroxy fatty acids or fatty alcohols. Because of their high production yield and the possibility to use renewable resources for this, they are the most promising for commercial production and utilisation (Kitamoto *et al.*, 2002). The best studied glycolipids are rhamnolipids synthesized by *Pseudomonas* sp. (Jarvis and Johnson, 1949), mannosylerythriol lipids synthesized by *Pseudozyma antarctica* (Kitamoto *et al.*, 1990), trehalose lipids synthesized by *Rhodococcus* sp., *Nocardia* sp. and *Mycobacterium* sp. (Lang and Wagner, 1987) and sophorolipids synthesized by *Candida* sp. (Cooper and Paddock, 1984). Table I-2 gives an overview of glycolipid producing microorganisms.

**Tabel 1-2**: Microbial source and yield for important glycolipids (Kitamoto *et al.*, 2002).

| Biosurfactant | Organism | Yield (g/L) | Carbon source |
|---|---|---|---|
| Mannosylerythritol lipids (MEL) | | | |
| MEL-A, -B and -C | *Pseudozyma antarctica T-34* | 140 | n-Octadecane |
| | | 47 | Soy oil |
| MEL-SY 16 | *Pseudozyma antarctica KCTC 7804* | 41 | Glycerol, oleic acid |
| MEL-A, -AB and -B | *Ustilago maydis DSM 4500* | 30 | Safflower oil |
| Rhamnolipids (RL) | | | |
| RL-1, -2, -3 and -4 | *Pseudomonas sp. DSM 2874* | 15 | n-Tetradecane |
| RL-1 and -2 | *Ps. aeruginosa DSM 7107* | 112 | Soy oil |
| RL-1 and -2 | *Ps. aeruginosa UI 29791* | 46 | Corn oil |
| RL-1 and -2 | *Ps. aeruginosa IF0 3924* | 32 | Ethanol |
| RL-A and -B | *Ps. aeruginosa BOP 101* | 14 | n-Paraffin |
| Trehalose lipids (TL) | | | |
| TL-1 and -2 | *Rhodococcus erythropolis DSM 43215* | 2 | n-Alkane |
| Trehalose-tetraëster (STL) | *R. elythropolis DSM 43215* | 32 | n-Decane |
| STL-1 and -2 | *R. erythropolis SD-74* | 40 | n-Hexadecane |
| Cellobiose lipids (CL) | | | |
| CL-A, -B and -C | *Ustilago maydis ATCC 14826* | 16 | Coconut oil |
| Sophorolipids (SL) | | | |
| SL mixture | *Candida bombicola ATCC 22214* | 422 | Whey, rapeseed oil |
| SL mixture | *C. bombicola ATCC 22214* | 317 | Glucose, rapeseed esters |
| SL mixture (SL-1: 73%) | *C. bombicola ATCC 22214* | 160 | Glucose, canola oil |
| SL (lacton) | *C. apicola IMET 43147* | 90 | Glucose, safflower oil |
| Oligosaccharide lipid | | | |
| GL-1, -2 and -3 | *Tsukumurella sp. DSM 44370* | 30 | Safflower oil |

### 3.3.1 Rhamnolipids

Rhamnolipids are produced by *Pseudomonas* sp.. *Ps. aeruginosa* is the best characterised producer for which yields higher than 100 g/L can be achieved. This relatively high production quantity makes rhamnolipids, together with sophorolipids, attractive compounds for further commercialization.

Like sophorolipids, microbially synthesized rhamnolipids are a mixture of a number of analogous molecules. The four most important ones comprise either one or two rhamnose molecules attached to one or two β-hydroxydecanoic acids (Syldatk *et al.*, 1985). Figure I-2 gives the structure of a rhamnolipid with two rhamnose and two β-hydroxydecanoic acid units (Rhamnolipid 3).

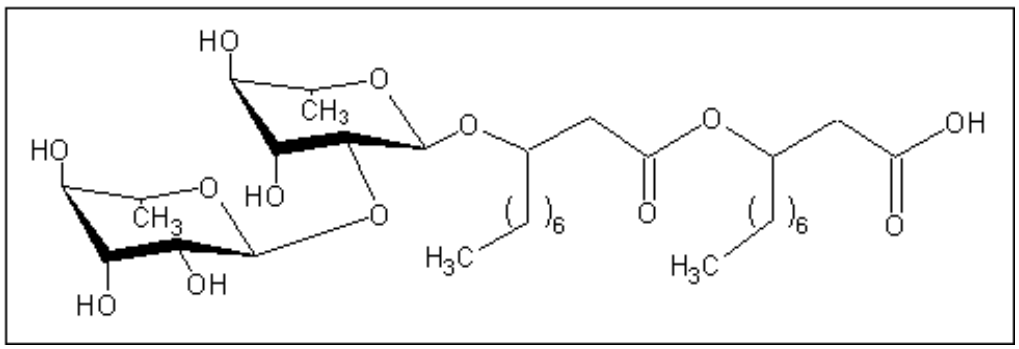


**Figure I-2:** Structure of rhamnolipid 3.

When looking at all biosurfactants, rhamnolipids are the ones for which most documentation is available on biosynthesis and genetic regulation. Biosynthesis proceeds by two sequential rhamnosyl-transfer reactions, each catalysed by a specific rhamnosyltransferase (Rt 1 and Rt 2) with TDP-rhamnose as donor. Rt 1 is composed of two polypeptides encoded by the *rhlA* and *rhlB* genes, located as one operon on a plasmid. This operon is among others regulated by RhlR en RhlI, proteins belonging to the LuxR and LuxI family which are known to be envolved in quorum-sensing responses. Also the gene encoding Rt2 is described (*rhlC*), as well as the pathways for synthesis of the precursors TDP-rhamnose and β-hydroxydecanoyl-S-CoA (Maier and Soberon-Chavez, 2000).

Rhamnolipids reduce the surface tension of water from 72.80 mN/m to 25 to 30 mN/m (Parra *et al.*, 1989). Because of their good biocompatibility and microbial degradation, they can be used in applications with extensive interference with the ecosphere such as tertiary petroleum recovery (Parra *et al.*, 1989), decontamination of marine oil pollution, soil remediation (Banat, 1995) and crop protection. In this latter application as crop protecting agent, rhamnolipids not only act as wetting agent to disperse the active molecule but they are themselves active against zoosporic phytopathogens such as *Pythium aphanidermatum*,

*Phytophtora capsici* and *Plasmopara lactucae-radicis* (Stanghellini and Miller, 1997). The rhamnolipid molecules are thought to interact with the phosphatidylethanolamine moiety of biological membrane systems. As such, they also show antimicrobial activity against Gram-positive and Gram-negative bacteria (Sanchez *et al.*, 2006). Furthermore, rhamnolipids can be applied in the food, cosmetic and pharmaceutical sectors.

One matter of concern in the production and commercialisation of rhamnolipids is the fact that *Pseudomonas aeruginosa* is a pathogenic organism. Other, non pathogenic *Pseudomonas* species are known, but the levels of produced glycolipids are much lower. Examples are *Ps. putida*, *Ps. chlororaphis* and *Ps. fluorescens* (Gunther *et al.*, 2007).

### 3.3.2 Mannosylerythritol lipids

Mannosylerythritol lipids (MEL) are synthesized by the yeast *Pseudozyma antarctica* as a mixture of four components; MEL-A and MEL-B as major ones, MEL-C and MEL-D as minor byproducts. The backbone of those molecules is a mannose-erythritol disaccharide on which short (2 to 8 carbon atoms) or long (10 to 18 carbon atoms) fatty acid chains are acetylated (Kitamota *et al.*, 1990).

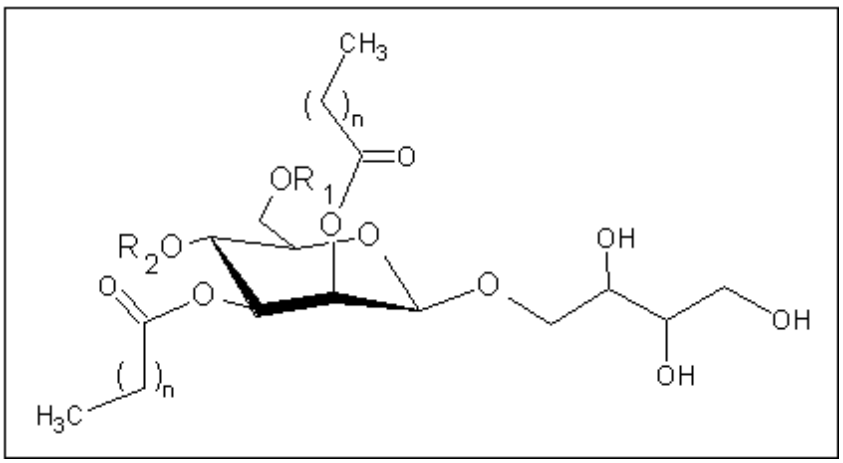


**Figure I-3:** The four possible component of mannosylerythritol lipids: Mel-A: $R_1$ = $R_2$ = acetyl, Mel-B: $R_1$ = acetyl and $R_2$ = H, Mel-C: $R_1$ = H, $R_2$ = acetyl, Mel-D: $R_1$ = $R_2$ = H, n = 6-12 (Kitamoto *et al.*, 1990).

MEL's can bring down the surface tension of water to 35 mN/m (Lang, 2002) and show versatile biochemical actions, including protein binding toward immunoglobulin G (Im *et al.*, 2003) and lectin (Konishi *et al.*, 2007), as well as induction of differentiation with respect to

different mammalian cells (Wakamatsu *et al.*, 2001). Because of this interesting biochemical behaviour, the pharmaceutical and medical sectors show al lot of interest in MEL's.

### 3.3.3 Trehalose lipids

Certain microorganisms synthesize glycolipids with trehalose sugar moieties (α-1,1 glucose disaccharide). Examples of known producers are bacteria such as *Mycobacterium* sp., *Corynebacterium* sp., *Nocardia* sp. and *Rhodococcus* sp., but production by *Rhodococcus erythropolis* has been best characterised (Desai and Banat, 1997). This species produces a complex mixture of surface active compounds, which remain cell-bound. The major constituents are trehalose mycolates (Figure I-4), but nitrogen-limited conditions lead also to the production of anionic trehalose tetraesters (Lang and Philp, 1998).

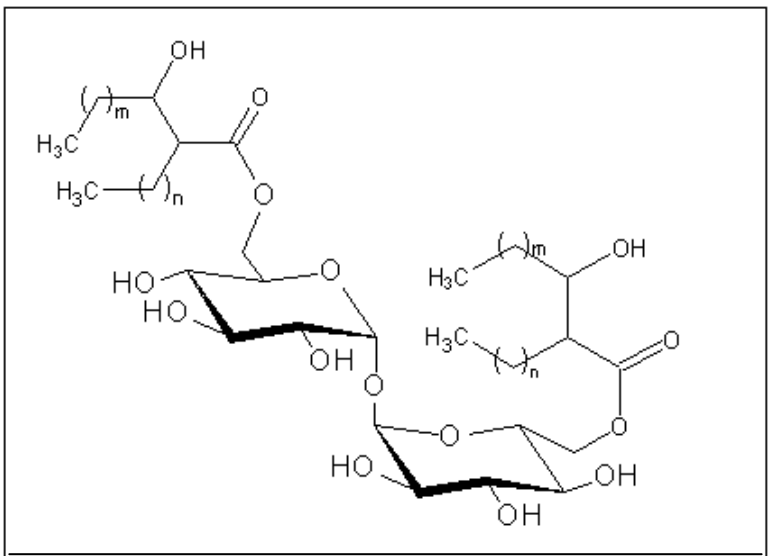


**Figure I-4:** Structure of nonionic trehalose-dicorynomycolates from *Rhodococcus erythropolis* DSM 43215. n + m = 27 to 30 (Rapp *et al.*, 1979).

## 3.4 Oligopeptides and lipopeptides

The molecules in this class of biosurfactants in general consist of cyclic peptides linked to a fatty acid. Several bacteria are known to produce those antibiotic-like molecules, among them *Bacillus subtilus*, which produces surfactin. This is one of the most powerful biosurfactants: it lowers the surface tension from 72.8 to 27.9 mN/m at a concentrations of 0.005 % (Arima *et*

*al.*, 1968). Furthermore, it possesses anti-bacterial, antiviral, anti-fungal, antimycoplasma and hemolytic activities. This lipopeptide consists of a hexapeptide lactonised to a hydroxy fatty acid (Figure I-5).

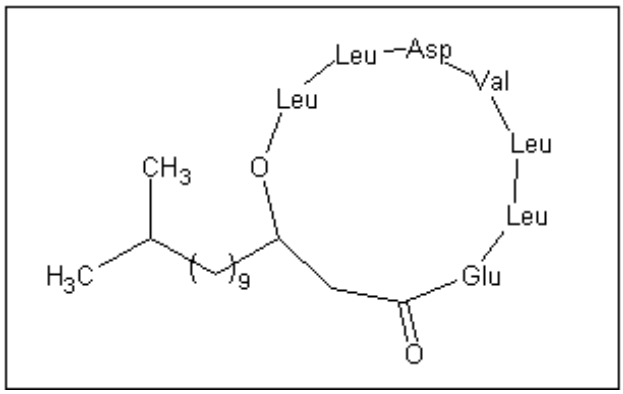


**Figure I-5:** Structure of surfactin.

## 3.5 Phospholipids, fatty acids and neutral lipids

Several fungi, yeasts and bacteria, that are able to grow on hydrophobic substrates such as alkanes, secrete large amounts of phospholipids, fatty acids or neutral lipids to facility the uptake of the carbon source. Examples are *Acinetobacter* sp. and *Aspergillus* sp. (Kappeli and Finnerty, 1979).

## 3.6 Polymeric biosurfactants

Polymeric biosurfactans are compiled from several components. Emulsan, synthesized by *Acinetobacter calcoaceticus*, is the best studied one. It consists of a heteropolysaccharide backbone to which fatty acids are covalently linked (Rosenberg *et al.*, 1988). Another example is liposan, a carbohydrate-protein complex synthesized by the yeast *Yarrowia lipolytica* (Cirigliano and Carman, 1984).

In addition to the four basic groups described above, there also exists biosurfactants build of carbohydrates, fatty acids and peptides, and sometimes external cell components or even whole cells show surface tension lowering properties (Table I-1).

# 4 Sophorolipids

## 4.1 Introduction

Sophorolipids are surface-active compounds synthesized by a selected number of yeast species. They have been known for over 40 years, but because of growing environmental awareness, they recently regained attention as biosurfactants due to their biodegradability, low ecotoxicity, and production based on renewable resources. In the next pages, an overview is given of the producing yeast strains and various aspects of fermentative sophorolipid production. Also, the biochemical pathways and regulatory mechanisms involved in sophorolipid biosynthesis are outlined. To conclude, a summary is given on possible applications of sophorolipids, either as native or modified molecules.

## 4.2 Producing microorganisms

In the early 1960s of the past century, Gorin *et al.* (1961) were the first to describe an extracellular glycolipid synthesized by the yeast *Torulopsis magnoliae*. However, the authors reported in 1968 that the producing strain was incorrectly identified and was actually *Torulopsis apicola* (Hajsig), currently known as *Candida apicola*. The structure of the hydroxy fatty acid sophoroside mixture was elucidated as a partially acetylated 2-O-β-D-glucopyranosyl-D-glucopyranose unit attached β-glycosidically to 17-L-hydroxyoctadecanoic or 17-L-hydroxy-Δ9-octadecenoic acid (Tulloch *et al.*, 1962; Tulloch and Spencer, 1968; Figure I-6).

In the same year, Tulloch *et al.* (1968a) also discovered a new sophorolipid from *Candida bogoriensis* (now known as *Rhodotorula bogoriensis*). The overall structure is similar to the sophorolipids of *Candida apicola*, but differs in its hydroxy fatty acid moiety: the sophorose unit is linked to 13-hydroxydocosanoic acid.

A third sophorolipid secreting yeast strain was identified by the same researchers as *Candida bombicola* (named initially *Torulopsis bombicola*); the glycolipids and production characteristics of this species are nearly identical to those of *Candida apicola* (Spencer *et al.*, 1970). In 1998, Rosa and Lachance described the novel yeast species *Starmerella bombicola*

and introduced it as the teleomorph of *Candida bombicola* based on the high 18S rDNA identity between both strains (more than 98%) and their ability to mate with each other to form ascospores.

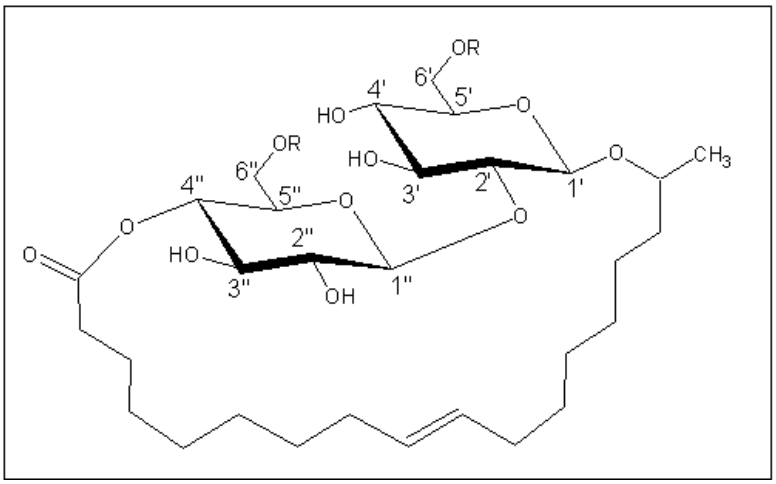


**Figure I-6:** Structure of a classic sophorolipid (lactonic form). R = H or $COCH_3$.

Recently, Chen *et al.* (2006a) proved sophorolipid synthesis in a new strain of *Wickerhamiella domericqiae*. They observed more than six glycolipids and identified one of the three main products as 17-L-(-oxy)-octadecanoic acid 1,4″-lactone 6′,6″-diacetate, which is identical to the major component of the sophorolipids of *C. apicola* and *C. bombicola*.

Regarding the fact that the production of sophorolipids is not restricted to one single yeast species, but to a number of related microorganisms, it is not unlikely to presume that other species belonging or related to the *Wickerhamiella*, *Starmerella* and *Rhodotorula* clades are also capable to synthesize some sort of sophorolipid.

This literature overview will mainly focus on the glycolipids synthesized by *C. bombicola* ATCC 22214 (Figure I-7). This is the strain preferred by most research groups active in the sophorolipid field; it can produce over 400 g/L sophorolipids and now is used for commercial production and applications.

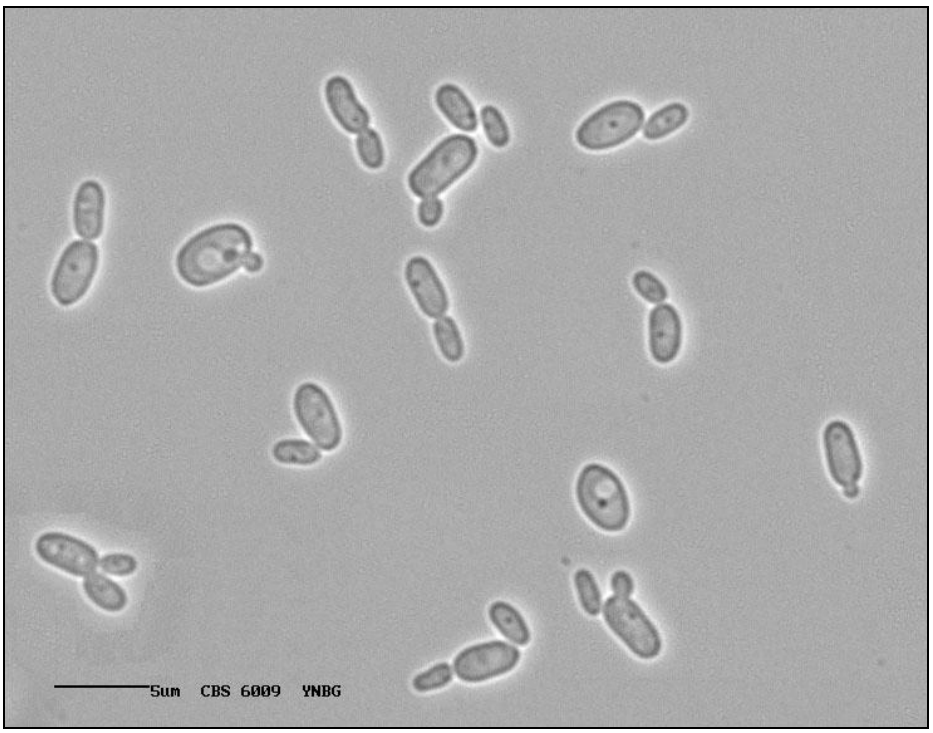


**Figure I-7:** Microscopic view of *Candida bombicola* cells, some of them showing budding. The length of the bar is 5 µm (http://www.cbs.knaw.nl/yeast/BioloMICS.aspx,).

### 4.2.1 A short description of *Candida bombicola*

#### 4.2.1.1 Origin and taxonomy

As its name already suggests, the first *Candida bombicola* isolate was obtained from the honey of a bumblebee or *Bombus* species (Spencer *et al.*, 1970). Just like *Candida bombicola*, many yeasts of the *Starmerella* clade are associated with bees or substrates visited by bees and it is suggested that a mutually beneficial interaction exists between them (Rosa *et al.*, 2003). As a consequence of its habitat, *Candida bombicola* is able to grow at high sugar concentrations and standard fermentations are usually started with a glucose level of 100 g/L or more.

Even though this yeast carries the genus name '*Candida*', this does not mean that it is closely related to the commonly known human pathogen *Candida albicans*. *C. bombicola* is a non pathogenic yeast and does not use the alternative translation of the CUG codon (serine instead

of leucine) as is described among others for *C. albicans*, *C. cylindracea*, *C. parapsilosis*, *C. melibiosica*, *C. zeylanoides*, *C. rugosa*, *C. maltosa*, *C. tropicalis*, *C. lusitaniae*, *C. guillermondii* and *C. viswanathii*, (Ohama *et al.*, 1993; Pesole *et al.*, 1995; Tuite and Santos, 1996). The full taxonomy of *Candida bombicola* is given below (The Dictionary of the Fungi, currently 9th edition, 2001; http://www.indexfungorum.org/Names/fundic.asp; consulted on 11 September 2007)

Dominio: *Eukaryota*
Regnum: *Fungi*
Phylum: *Ascomycota*
Subphylum: *Saccharomycotina*
Classis: *Saccharomycetes*
Subclassis: *Saccharomycetidae*
Ordo: *Saccharomycetales*
Familia: *Saccharomycetaceae*
Genus: *Candida* (anamorph) or *Starmerella* C.A. Rosa & Lachance 1998 (teleomorph)
Species: *Candida bombicola* J.F.T. Spencer, Gorin & A.P. Tulloch (anamorph) or *Starmerella bombicola* (teleomorph)

#### 4.2.1.2 Strain characteristics

The most important characteristics of *Candida bombicola* ATCC 22214 are listed in Table I-3. In the work described here, strain *Candida bombicola* ATCC 22214 is used, unless otherwise specified. References from other collections are CBS 6009, CCRC 21323, CCRC 22302, DBVPG 6870, IFO 1449, IFO 10243, JCM 9596, NRIC 1806 and NRRL Y-17069.

**Table I-3:** Characteristics of *Candida bombicola* ATCC 22214 (*Centraalbureau voor Schimmelcultures*, www.cbs.knaw.nl; National Collection of Yeast Cultures, www.ncyc.co.uk).

| CHARACTERISTIC | FEATURE | VALUE |
|---|---|---|
| **Temperature (°C)** | Optimum | 28.8 |
| | Minimum | 14 |
| | Maximum | 31.3 |
| **Cell wall constitution (%)** | Glucose | 43 |
| | Mannose | 57 |
| **Metabolites** | Sophorolipids | |
| **Cells and reproduction** | Cell shape | elongated |
| | Cell size | 2-4 µm (see Figure I-7) |
| | Cell division | multipolar budding |
| | Filamentous growth | absent |
| | Asexual spores | absent |
| | Sexual spores | absent |
| **Coenzyme Q system** | 9 | |
| **Carbon source** | Glucose | + |
| | Galactose | - |
| | Sucrose | + |
| | Maltose | - |
| | Cellobiose | - |
| | Trehalose | - |
| | Lactose | - |
| | Melibiose | - |
| | Raffinose | + |
| | Inulin | - |
| | Soluble Starch | - |
| | Me-α-D-Glucoside | - |
| | 50% Glucose | + |
| | 60% Glucose | + |
| **Nitrogen source** | $NH_4(SO_4)_2$ | + |
| | $KNO_3$ | - |
| | Ethylamine | + |

## 4.3 Structure and properties

As a surfactant molecule, sophorolipids are amphiphilic molecules interacting with the phase boundary in heterogeneous systems. They consist of a hydrophobic fatty acid tail of 16 or 18 carbon atoms and a hydrophilic carbohydrate head, sophorose. Sophorose is a glucose disaccharide with an unusual β-1,2 bond and can - in the case of sophorolipids - be acetylated on the 6′-and/or 6″positions (Figure I-6). One terminal or subterminal hydroxylated fatty acid is β-glycosidically linked to the sophorose molecule. The carboxylic end of this fatty acid is either free (acidic or open form) or internally esterified at the 4" or in some rare cases at the 6'-or 6"-position (lactonic form). The hydroxy fatty acid itself counts in general 16 or 18 carbon atoms and can have one ore more unsaturated bonds (Asmer *et al.*, 1988; Davila *et al.*, 1993). As such, the sophorolipids synthesized by *C. bombicola* are in fact a mixture of related molecules with differences in the fatty acid part (chain length, saturation, and position of hydroxylation) and the lactonization and acetylation pattern. Asmer *et al.* (1988) were the first to shed light on this structural variation. They separated the sophorolipid mixture by medium pressure liquid chromatography and thin layer chromatography, and mainly based on the lactonization and acetylation pattern, they put forward 14 components. However, differences in fatty acid length and hydroxylation pattern were not taken into account. Davila *et al.* (1993) separated the sophorolipid mixture by a gradient elution high-performance liquid chromatography (HPLC) method and used an evaporative light scattering for the detection of the individual sophorolipids. They spend special attention to the analysis of the fatty acid chain and identified over 20 components.

When sophorolipids are disolved in water, they lower the surface tension from 72.80 mN/m down to 40 to 30 mN/m, with a critical micelle concentration of 40 to 100 mg/L. The hydrophilic/lipophilic balance is 10 to 13, making sophorolipids useful as detergents or as stabilizers for oil-in-water emulsions. The different structural classes cause wide variation in physicochemical properties. Lactonized sophorolipids have different biological and physicochemical properties as compared to acidic forms. The biosurfactants' hydrophilic/lipophilic balance, foam formation capacity, and antimicrobial effects are all strongly influenced by the degree of lactone formation. In general, lactonic sophorolipids have better surface tension lowering and antimicrobial activity, whereas the acidic ones display better foam production and solubility. Also, the presence of acetyl groups can have a

profound effect on the properties of biosurfactants. Indeed, acetyl groups lower the hydrophilicity of sophorolipids and enhance their antiviral and cytokine stimulating effects (Shah *et al.*, 2005).

Sophorolipids are readily biodegradable surfactants as determined by standard manometric respirometry (Figure I-8) and stable metabolite studies. There is no evidence for the accumulation of stable metabolites in both the accumulation phase and the running out phase of simulated waste water systems; sophorolipids are totally degraded till 1 carbon metabolites (Renkin, 2003). Furthermore, they are far less toxic for aquatic organisms than the conventional detergents; the inhibitory effect on the crustacea *Daphnia magna*, the ciliate *Tetrahymena terhmophila* and the micro algae *Pseudokirchneriella subcapitata* is tenfold less as compared to conventional surfactants (Renkin, 2003).

Tests with crude and acidic sophorolipids pointed out that they are not irritating to the skin, do not trigger allergic reactions and have an oral safety level which is greater than or equal to 5 ml/kg weight (Hillon *et al.*, 1998).

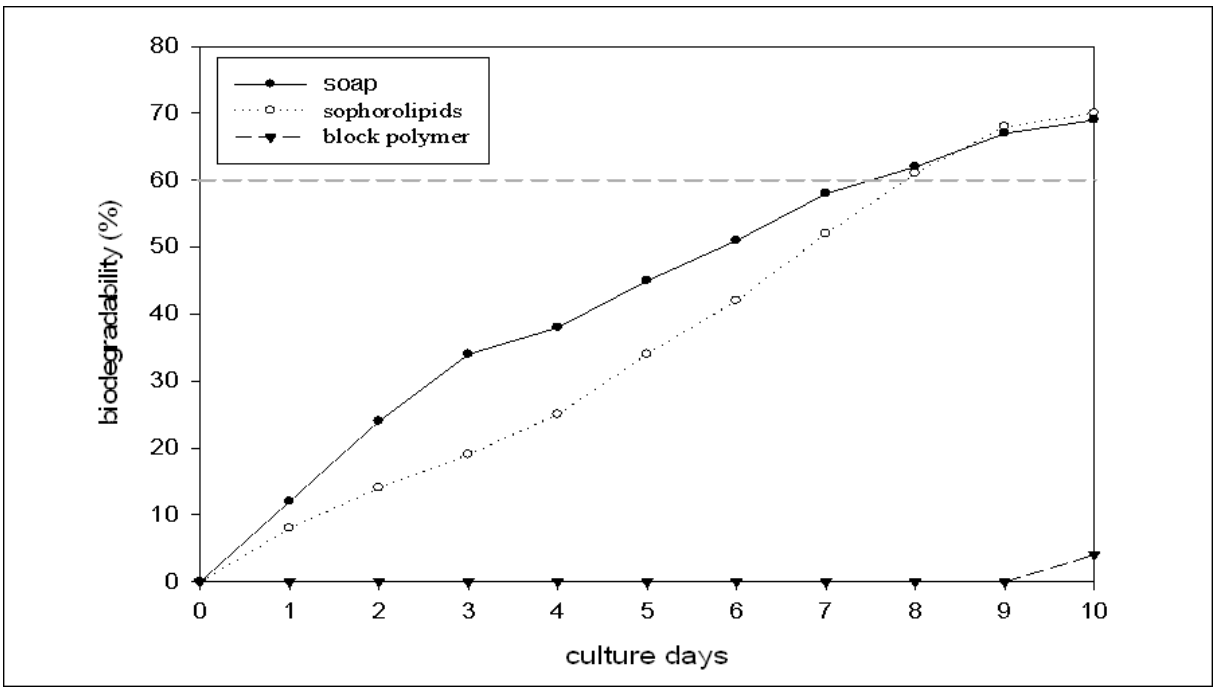


**Figure I-8:** Evaluation of the biodegradabilty of sophorolipids; they exceed the biodegradability (after product information on Sophoron from Saraya Co, LTD).

## 4.4 Biosynthesis

Figure I-9 gives a schematic overview of the biochemical pathways involved in sophorolipid synthesis. The building blocks for conventional sophorolipid synthesis are glucose and a fatty acid. Ideally, both can be provided in the production medium as such or, because free fatty acids can disturb the electron balance of the cells, sometimes fatty acid methyl or ethyl esters, or triglycerides are used. In this case, esterases will mediate gradual release of fatty acids. Since sophorolipid-producing yeast strains such as *C. bombicola* and *C. apicola* are capable of growing on alkanes, they posses the enzymes required for the terminal oxidation of alkanes, thereby generating fatty acids for further β-oxidation. Consequently, also alkanes or intermediates of the terminal oxidation pathway can act as feedstock.
If no hydrophobic substrate is present in the medium, fatty acids will be formed *de novo* starting from acetyl-CoA derived from the glycolysis pathway. On the other hand, and this especially when the glucose concentration is low, part of the fatty acids will be conducted toward the β-oxidation for cell maintenance instead of sophorolipid synthesis.

In a first step, the fatty acids are converted to a terminal (ω) or subterminal (ω-1) hydroxy fatty acid through the action of a membrane bound nicotinamide adenine dinucleotide phosphate (reduced form; NADPH) dependent monooxygenase enzyme, cytochrome P450 (Jones, 1968). Lottermoser *et al.* (1996) identified two cytochrome P450 monooxygenase genes from *C. apicola* (European Molecular Biology Laboratory/GenBank accession numbers X76225 and X87640). Based on the amino acid similarity, they were classified into the CYP52 family, which comprises cytochrome P450 enzymes of yeasts capable of hydroxylating alkanes and/or fatty acids (Nelson, 1998). However, the authors did not verify whether the corresponding gene products were involved in sophorolipid production or alkane assimilation, and if they were expressed at all. Yet, evolutionary history of cytochrome P450 genes is characterized by several events of gene duplication and conversion, resulting in a broad diversity among those genes also within the genome of a single organism, and it is not always clear what induces them or if they are expressed at all (Nebert and Gonzalez, 1987; Nelson, 1999). For *C. bombicola* ATCC 22214, we have identified eight different cytochrome P450 monooxygenase genes belonging to the CYP52 family. One of them exposes very high similarity (91% AA identity) to the CYP52E2 gene of *C. apicola*, whereas the others probably

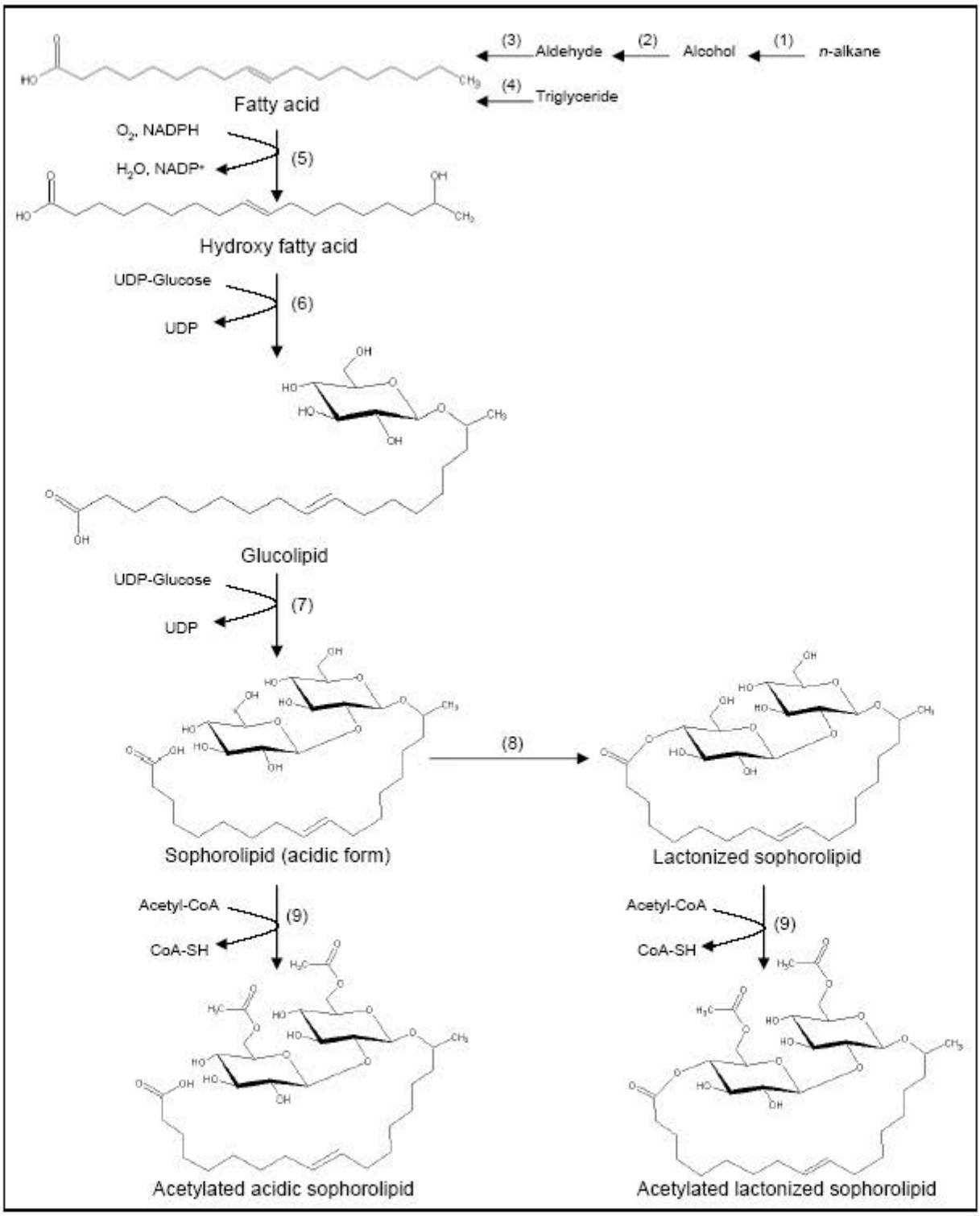


**Figure I-9:** Proposed sophorolipid biosynthetic pathway. 1, cytochrome P450 monooxygenase; 2, alcohol-dehydrogenase; 3, aldehyde-dehydrogenase; 4, lipase; 5, cytochrome P450 monooxygenase; 6, glucosyl-transferase I; 7, glucosyltransferase II; 8, lactonesterase; 9, acetyltransferase.

belong to one or more new CYP52 subfamilies. Isolation of these genes is described in Chapter VI.

In a second step, glucose is glycosidically coupled (position C1') to the hydroxyl group of the fatty acid through the action of a specific glycosyltransferase I. Experiments with $^{13}$C-labeled glucose pointed out that the bulk of the added glucose first passed through glycolysis, in this way supplementing trioses for the gluconeogenesis of glucose for sophorolipid synthesis (Hommel *et al.*, 1994). The transferase reaction requires nucleotide-activated glucose (uridine diphosphate (UDP)-glucose) as glucosyldonor (Breithaupt and Light, 1982).

In a subsequent step, a second glucose is glycosidically coupled to the C2' position of the first glucose moiety by glycosyltranferase II. Both glycosyltransferases involved in sophorolipid synthesis of *Rhodotorula bogoriensis* were partially purified. The two enzyme activities could however not be separated and highly purified samples exhibit a single major band of 52 kDa on sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE; Esders and Light, 1972; Breithaupt and Light, 1982). Therefore, it remains open for discussion whether the consecutive glucose transfers are carried out by two different (but copurified) enzymes or by one and the same (multi)enzyme. It is supposed that sophorolipid synthesis in *C. bombicola* involves analogous enzymes.

The sophorolipids obtained after the action of glucosyltransferase II are as such detected in the sophorolipid mixture as the acidic, nonacetylated molecules. The majority of the sophorolipids are however further modified by both internal esterification (lactonization) and by acetylation of the carbohydrate head. Lactonic sophorolipids are formed by an esterifaction reaction of the carboxyl group of the hydroxy fatty acid with a hydroxyl group of sophorose (Figure I-6 and I-9). The vast majority of lactones are esterified at the 4"-position, whereas a small percentage is esterified at either the 6'-or 6"-position (Asmer *et al.*, 1988). Since several commercial lipases are able to introduce such a 6'-or 6"-ester linkage, it is suggested that those bonds are formed by cellular lipases, whereas esterification at the 4"-position is believed to be catalyzed by a specific lactone esterase. Neither of both esterases has been identified in *C. bombicola* or in other sophorolipid producing species.

The acetylation at the 6'- and/or 6"-position is carried out by an acetyl-coenzyme A (CoA) dependent acetyl transferase. The transferase from *R. bogoriensis* has been partially purified

(Esders and Light, 1972; Bucholtz and Light, 1976), but the corresponding enzyme has not yet been identified in *C. bombicola*.

## 4.5 The fermentation process

### 4.5.1 Culture conditions

Together with mannosylerythritol lipids, sophorolipids are the only surfactants produced in large quantities by yeasts.

*C. bombicola* ATCC 22214 is mainly used as producing strain, whereas some research groups in Germany prefer(red) working with *C. apicola* IMET 42747 (Hommel *et al.*, 1987). This latter strain is however closely related to *C. bombicola*, and mechanisms and characteristics for sophorolipid production can be considered more or less the same for both yeast species.

The optimal growth temperature of *C. bombicola* ATCC 2214 is 28.8 °C. Gobbert *et al.* (1984) pointed out that the optimal temperature for sophorolipid formation is 21 °C. The authors themselves however advised to use higher temperatures for better handling (e.g., sample taking and oil addition). Most fermentations are run at 25 or 30 °C. The amount of obtained sophorolipid is nearly identical for both temperatures, whereas for fermentations at 25 °C, biomass growth is lower, and the glucose consumption rate is higher as compared to the fermentation at 30 °C (Casas and Garcia-Ochoa, 1999).

During the exponentional growth phase, pH drops tremendously and must further be maintained at the value of 3.5 by the addition of NaOH for optimal sophorolipid production (Gobbert *et al.*, 1984). This low pH and the antimicrobial effect of sophorolipids protect the fermentation broth against contamination, even when fed-batch processes of more than 200 h are run. Sophorolipid synthesis starts when the yeast cells enter stationary phase and takes place under conditions of nitrogen limitation, making the synthesis not growth-associated (Davila *et al.*, 1992). Albrecht and colleagues linked the synthesis also to phosphate depletion and suggested the following mechanism. Phosphate- or nitrogen-limiting conditions cause decline in the specific activities of $NAD^+$- and $NADP^+$-dependent isocitrate dehydrogenase, which leads to accumulation of isocitrate and subsequently citrate in the mitochondria. Both

are transported into the cytosol and citrate is cleaved by adenosine triphosphate (ATP): citrate synthase to give rise to acetyl-CoA, the precursor for fatty acid synthesis (Albrecht *et al.*, 1996).

Oxygen supply is very important throughout the whole fermentation process; the yeast cells are very sensitive to oxygen limitation during their exponentional growth, and good aeration conditions are important for sophorolipid production. Guilmanov *et al.* (2002) investigated the effect of aeration by means of shake-flask experiments. The optimal aeration for high sophorolipid yield, expressed in terms of oxygen transfer rate, lays between 50 and 80 mM $O_2$/L h.

### 4.5.2 Hydrophilic and lipophilic carbon sources and their influence on sophorolipid formation

When the yeast cells are supplied with only one type of carbon source, such as glucose (Hommel *et al.*, 1994) or n-alkanes (Jones and Howe, 1968), sophorolipid formation is observed. The production is however considerably higher when two types of carbon sources, hydrophilic (glycidic) and a hydrophobic (lipidic) one, are provided. In most cases, glucose is used as the hydrophilic carbon source. Sucrose can also act as substrate, but the obtained sophorolipid level is lower (Klekner *et al.*, 1991). In an attempt to reduce substrate costs, cheese whey was proposed as hydrophilic carbon source. Zhou and Kosaric (1993, 1995) first investigated the fermentation with galactose and lactose, the main sugar components of whey. *C. bombicola* was not able to grow when only lactose was present, but when also olive, canola, or safflower oil was supplemented, growth and sophorolipid formation were observed. Daniel *et al.*, (1998a) then investigated the production of sophorolipids in a medium with deproteinized whey concentrate and rapeseed oil. They obtained high levels of sophorolipid production (280 g/L), but strangely enough, the lactose was not consumed, nor could β-galactosidase activity in the supernatant or crude cell extract be detected. Still, the same researchers managed to fully use the deproteinized whey as a substrate for sophorolipid production by a two-stage cultivation process. They first cultivated the oleagineous yeast *Cryptococus curvatus* on the whey. The cells, which accumulated a high single cell oil level, were harvested and disrupted and then served as lipidic substrate for the *C. bombicola* cells

(Daniel *et al.*, 1998b). Furthermore, also low-cost soy molasses can act as glucose substitute, but again, lower yields were a observed (Solaiman *et al.*, 2004). As *C. bombicola* was originally isolated from honey, some researchers have tested it as a substrate. It was however only supplemented at the end of the fermentation process when the initially added glucose was consumed (Pekin *et al.*, 2005).

A lot of substrates can act as hydrophobic carbon source: oils, fatty acids, and their corresponding esters, alkanes, etc. The level of sophorolipid formation during fermentations based on alkanes as hydrophobic feed-stock largely depends on the chain length of the used substrate. Hexadecane and octadecane give the best production yields. They appear to be directly converted into hydroxy fatty acids and incorporated into the sophorolipid molecules, in this way strongly influencing the fatty acid composition of the sophorolipid mixture. Davila *et al.* (1994), for example, detected over 70% of hydroxylated hexadecanoic and octadecanoic acid for fermentations on hexadecane and octadecane, respectively. Shorter alkanes are only to a minor extent incorporated, whereas the vast majority is either elongated to C16 or C18 fatty acids, or metabolized via β-oxidation. More or less the same is true for eicosane (n-C20) or longer alkanes. A few percentages of eicosane could be detected in the sophorolipid fatty acid moiety, whereas none were observed for the longer alkanes. The alkanes are metabolized by β-oxidation, either completely or partially to give rise to shorter fatty acids which can be incorporated into the sophorolipid molecules. The yields are however higher when compared to the shorter alkanes and even comparable to those obtained for n-C16 and n-C18 (Tulloch *et al.*, 1962; Jones and Howe, 1968). The same trend is observed for fatty acids or fatty acid methyl esters. The best results are obtained for oleic acid (C18:1). It turns out that the fatty acid chain length determines the rate and position of hydroxylation and consequently governs the incorporation into the sophorolipid molecule. Pentadecanoic acid is too short for hydroxylation, but palmitic acid (22.55 Å) is not. It is predominantly hydroxylated at the terminal position. The longer the chain, the more the terminal/subterminal oxidation ratio declines; for stearic acid (25.05 Å) for example, no terminal oxidation is observed while this does occur for oleic (24.22 Å) and heptadecanoic acid (23.80 Å). Linolenic acid (C18:3), however, does not follow this rule; no sophorolipids with this fatty acid tail were ever observed. It is suggested that the 15–16 double bound is too close to the enzymatic reaction site (Tulloch *et al.*, 1962; Davila *et al.*, 1994). In both sophorolipid mixtures on alkanes and

fatty acids with direct incorporation, desaturation is observed (Brett *et al.*, 1971; Davila *et al.*, 1994).

Furthermore, oils (especially those of vegetable origin) are widely used as lipidic carbon source. The most common vegetable oils are comprised of saturated or unsaturated fatty acids with chain lengths of 16 or 18 carbon atoms, making them an ideal substrate for direct incorporation and the consequent high sophorolipid production and yield. Examples of tested oils are canola, corn, safflower, sunflower, olive, rapeseed, grape seed, palm, coconut, fish, and soybean oil (Cooper and Paddock, 1984; Lee and Kim, 1993; Davila *et al.*, 1994; Zhou and Kosaric, 1995; Rau *et al.*, 1996; Kim *et al.*, 1997; Casas and Garcia-Ochoa, 1999; Cavalero and Cooper, 2003; Pekin *et al.*, 2005). As vegetable oils are renewable resources, sophorolipid synthesis with those carbon sources contributes to the environmental friendly character of this surfactant. Furthermore, waste streams such as biodiesel by-product streams, soybean dark oil and waste frying oil can be used (Ashby *et al.*, 2005; Kim *et al.*, 2005; Fleurackers, 2006). It is difficult to compare the effects of various oil sources because different media and culture conditions were used. Casas and Garcia-Ochoa (1999), however, compared olive, grape seed, sunflower, corn, and coconut oil and concluded that sunflower oil gave the best results. Coconut oil is less suitable, as it in fact contains high amounts of lauric and myristic acid. Davila *et al.* (1994) tested palm, sunflower, fish, and rapeseed oil. The latter one turned out to be most suitable for sophorolipid production due to its high oleic acid content. These researchers could further increase the production and yield from 255 g/L and 0.53 g/g up to 340 g/L and 0.65 g/g when using the methyl or ethyl esters of rapeseed oil. Nonincorporated substrates, either alkanes, fatty acids, or esters, are mainly oxidized to $CO_2$.

Brakemeier *et al.* (1995, 1998a) developed a method to circumvent the length-dependent and restricted incorporation. They used secondary alcohols (C12 to C16) as the lipophilic carbon source. The majority of the substrate was incorporated into the glycolipids without any further modifications, although small percentages of (ω-1)-alkandiols and hydroxy fatty acids were observed. The resulting compounds display better surface active properties as compared to native sophorolipids. Another remarkable event is the possible coupling of sophorose units at both sites of the alkandiols, leading to a glycolipid with two sophorose units separated by a hydrocarbonic spacer; these molecules are however only slightly soluble in water (Figure I-10). The greatest disadvantage of this method is the high cost of secondary alcohols. Primary

alcohols are much cheaper substrates, but are most likely converted to fatty acids and metabolized in the β-oxidation route. Their metabolization can nevertheless be avoided in favor of the incorporation in sophorolipids by increasing the level of glucose (150 g/L) and yeast extract (4 g/L). Tests with 2-, 3-, or 4-dodecanones showed that those ketones were reduced into their corresponding alcohols by *C. bombicola* and subsequently incorporated into sophorolipid molecules (Brakemeier *et al.*, 1998b).

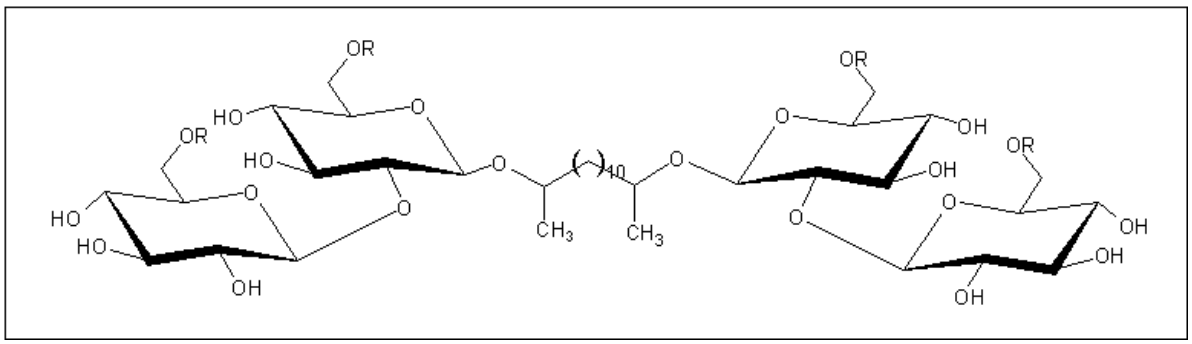


**Figure I-10**: Glycolipid with two sophorose units, described by Braekemeier *et al.* (1998a). R = H or $COCH_3$.

### 4.5.3 Effect of other medium components on sophorolipid production

The level of sophorolipid production largely depends on the medium composition and the addition of both the glycidic and lipidic carbon source. Tulloch *et al.* (1968b) obtained 40 g/L using a simple medium containing glucose, yeast extract, urea, and n-octadecane. Certain researchers kept using this medium and could increase the sophorolipid formation by better parameter control of the fermentation process and/or stepwise addition of the hydrophobic carbon source (Asmer *et al.*, 1988; Rau *et al.*, 1996; Casas and Garcia-Ochoa, 1999; Guilmanov *et al.*, 2002; Cavalero and Cooper, 2003; Solaiman *et al.*, 2004). Pekin *et al.* (2005), for instance, managed to produce up to 400 g/L sophorolipids. Others use media comprising potassium phosphate, citrate, and ammonium instead of urea and minerals such as iron, magnesium, and calcium and achieve similar results (Daniel *et al.*, 1998b; Davila *et al.*, 1994). Typically, a carbon conversion yield between 60 and 70 % is obtained.

Yeast extract is essential for both cell growth and sophorolipid formation. Substitution with ureum or peptone negatively influenced the biomass and glycolipids yield. The most favorable concentration remains however unclear. Cooper and Paddock (1984) set the optimum at 5 g/L, whereas Zhou *et al.* (1992) obtained the highest sophorolipid concentration

using only 2 to 3 g/L yeast extract. Casas and Garcia-Ochoa (1999) further decreased the yeast extract content to 1 g/L. They pointed out that higher concentrations were favorable for the development of biomass but decreased the glycolipids production due to depletion of the carbon sources at the expense of cell growth. It stays however tricky to draw conclusions because all three research groups set this optimum using different media.

## 4.6 Factors influencing the sophorolipid composition

As discussed in the above section, the fatty acid moiety of the sophorolipid molecules is to a large extent influenced by the type of supplemented hydrophobic carbon source. The glycidic part as such always is sophorose, whatever the employed culture conditions or hydrophilic carbon source might be.

However, the sophorose unit can be mono- or diacetylated and esterified to the fatty acid carboxylic site. What exactly determines the degree of lactonization remains unclear, but the lactonic/acidic balance is to a certain point influenced by fermentation conditions. Especially when the alkanes hexadecane, heptadecane, or octodecane are used, 85 % or more of the sophorolipids are diacetylated lactones. This high structural homology involves another interesting phenomenon: these sophorolipids tend to form white crystals instead of the traditional brown and viscous oily mixture. This feature makes them relatively easy to isolate (Davila *et al.*, 1994; Cavalero and Cooper, 2003). Such high degrees of lactonization are more difficult to achieve with oils or fatty acids, but can be obtained in fed-batch fermentation with limited feeding of the lipidic carbon source (Davila *et al.*, 1994; Rau *et al.*, 1996). According to Davila *et al.* (1994), sophorolipids produced from oils always exhibited a higher level of diacetylated lactones than the ones produced from the corresponding esters. In addition, the fatty acid composition of the lipidic carbon source also influences the sophorolipid composition. When substrates rich in polyunsaturated fatty acids are used (e.g., sunflower or linseed oil), increased levels of acidic sophorolipids were observed. Loss of flexibility of the sophorolipidic fatty acid tail due to the two double bounds can possibly hamper internal esterification.

The partition in structural classes also tends to depend on the level of yeast extract in the medium. A concentration of 1 g/L mainly triggers the lactonic form, whereas 20 g/L lead to

the synthesis of acidic sophorolipids. Furthermore, the same investigators noticed a higher concentration of acidic sophorolipids in shake-flask compared to fermentor experiments, suggesting that also the oxygen supply is a determining factor (Casas and Garcia-Ochoa, 1999). Finally, the lactonic/acidic balance seems to change during the course of fermentation. In the beginning, acidic forms are predominant, whereas later on, conversion to the lactonic and acetylated forms is observed (Davila *et al.*, 1992). It is therefore suggested that in energetically less favorable conditions, more acidic structures are found (Davila *et al.*, 1997).

## 4.7 Downstream processing

On laboratory scale, sophorolipids can be extracted from the culture broth with organic solvents such as ethyl acetate. Residual lipidic carbon source can however be coextracted and cause difficulties during further applications. For this reason, additional extraction with hexane is most frequently used, but other solvents such as pentane (Cavalero and Cooper, 2003) or t-butyl methyl ether (Rau *et al.*, 2001) can also be applied.

As sophorolipids are heavier than water, it is often possible to centrifuge them down or to just decant the sophorolipids from the fermentation medium, after heating if necessary. This method is very convenient when working with large volumes and high yields. Further elimination of water and impurities can be achieved by addition of polyhydric alcohols and subsequent distillation (Inoue *et al.*, 1980). Ultrafiltration is also often used in the recovery of biosurfactants, but this technique has not yet been optimized for sophorolipids (Mulligan and Gibbs, 1990).

As discussed in the previous section, certain fermentation conditions can give rise to high concentrations of lactonic sophorolipids resulting in a crystalline product, which is obviously much easier to separate and purify. Hu and Ju (2001) developed a method for the separation of lactonic sophorolipids from the crude, viscous mixture based on crystallization in phthalate and phosphate buffers. If one is interested in a specific sophorolipid structure, e.g., for pharmaceutical applications, chromatographic purification with silica gel or preparative reversed phase columns is necessary (Lin, 1996).

## 4.8 Physiological role of sophorolipids in *C. bombicola* ATCC 22214

Synthesis of biosurfactants is often associated with the assimilation of hydrophobic substrates (Ito *et al.*, 1980); this theory is however not commonly accepted for sophorolipid formation, especially because the molecules are also formed when no hydrophobic substrate is present and in amounts largely exceeding the concentration required for emulsification. As sophorolipid synthesis is associated with nitrogen starvation, it is suggested that formation of glycolipids is some sort of overflow metabolism, by means of extracellular storage material. This hypothesis is supported by the findings of Hommel *et al.* (1994), regarding the biosynthesis of the sophorose moiety, which resembles the trehalose synthesis *of Saccharomyces cerevisiae* under anaerobic conditions. It was also demonstrated that sophorolipids can be used as sole carbon source (Garcia-Ochoa and Casas, 1997). As *C. bombicola* and *C. apicola* by nature occur in environments with high osmotic strength, sophorolipid production may be a way of dealing with the high sugar concentrations by converting, storing, and making them less available for other organisms. Furthermore, sophorolipids display antimicrobial activity against certain yeasts such as *Candida* and *Pichia* species (Ito *et al.*, 1980) and Gram-positive bacteria. The mono-and diacetylated lactones have the strongest inhibitory effect (Lang *et al.*, 1989). We believe that the physiological role of sophorolipid synthesis is extracellular carbon source storage, combined with dealing with a high-sugar niche and defending this against other competing microorganisms.

## 4.9 Applications of native sophorolipids

The most recognized feature of sophorolipids is their ability to act as a surfactant. Surfactants are widely used in the food, pharmaceutical, cosmetic, and cleaning industries. Throughout the years, surfactants have been produced from petrochemical raw materials. During the last decades, the environmental awareness has become a more important issue in the study, development, and application of surfactants. In this respect, biosurfactants such as sophorolipids offer the advantages of biodegradability, low ecotoxicity and the production on renewable-resource substrates.

Those characteristics draw the attention of the Belgian company Ecover NV (http://www.ecover.com), a manufacturer of ecological detergents and cleansing agents and

also active in the natural cosmetics and professional cleansing sector. They saw potential in the use of sophorolipids in hard surface cleaners such as multi surface cleaner, floor soap and window cleaner (Develter *et al.*, 2007). Four novel products were created (Techno Ceramica, Boat Wax, Auto wax and Dashboard Cleaner) and the composition of the currently available hard surface cleaners will be redesigned based on a sophorolipid input.

The Japanese company Saraya Co, LTD (http://www.saraya.com) has commercialized sophoron, a dish washer containing sophorolipids as cleaning agent (Futura *et al.*, 2002). Sophorolipids can also be applied in laundry detergents (Hall *et al.*, 1996). As sophorolipids are nonionic molecules, they preserve their surface lowering properties despite high salt concentrations. In addition, they are active across a wide temperature range. One feature one must keep however in mind is their instability at pH values higher than 7.0 to 7.5: beyond this point, irreversible hydrolysis of the acetyl groups and ester bonds is observed. The emulsifying properties of sophorolipids can also be exploited in the petroleum industry; they are useful in secondary oil recovery, in removing hydrocarbons from drill material, and in the regeneration of hydrocarbons from dregs and muds (Baviere *et al.*, 1994; Marchal *et al.*, 1999; Pesce, 2002). Sophorolipids can also be applied for decontaminating porous media such as soils and groundwater tables polluted by hydrocarbons (Ducreux *et al.*, 1997) and in the removal of heavy metals from sediments (Mulligan *et al.*, 2001). Furthermore, the emulsifying property of sophorolipids can be used in the food industry to improve the quality of wheat flour products (Akari and Akari, 1987) and in the cold storage transportation in air conditioning systems for the prevention of ice particle formation (Masaru *et al.*, 2001).

Moreover, sophorolipids find application in various cosmetic formulations. The French company Soliance (http://www.groupesoliance.com) produces sophorolipid-based cosmetics for body and skin. In addition to its role as emulsifier, the glycolipid acts as a bacteriocidal agent in the treatment of acne, dandruff, and body odours (Mager *et al.*, 1987). Furthermore, they are claimed to trigger several beneficial events regarding the protection of hair and skin, making them attractive components in cosmetic, hygienic, and pharmaco-dermatological products. They stimulate the dermal fibroblast metabolism and collagen neosynthesis, inhibit free radical and elastase activity, possess macrophage-activating and fibrinolytic properties, and act as desquamating (i.e., eliminating the surface portion of the protective layer of the

epidermis as part of the wound healing process) and depigmenting agents (Hillion *et al.*, 1998; Borzeix, 1999; Maingault, 1999). Sophorolipids are also believed to stimulate the leptin synthesis through adipocytes, in this way reducing the subcutaneous fat overload (Pellecier and André, 2004).

As mentioned above, sophorolipids possess antimicrobial properties. For this reason, they can be applied in germicidal mixtures suitable for cleaning fruits and vegetables (Pierce and Heilman, 1998). The antimicrobial action is not merely restricted toward bacteria; sophorolipids also act as antifungal agents against plant pathogenic fungi such as *Phytophthora* sp. and *Pythium* sp. (Yoo *et al.*, 2005), inhibit algal bloom (Gi, 2004), and are even claimed to have anti-human immunodeficiency virus and sperm-immobilizing activities (Shah *et al.*, 2005). Other medical beneficial effects of sophorolipids are their ability to trigger cell differentiation instead of cell proliferation and the inhibition of protein kinase C activity of the human promyelocytic leukemia cell line HL60. The anticancer action is not caused by a simple detergent-like effect, but is attributed to a specific interaction with the plasma membrane (Isoda *et al.*, 1997). The above-described experiment was conducted with a crude sophorolipid mixture. Chen *et al.* (2006a), however, conducted anticancer tests with the purified diacetylated lacton. This component exhibits cytotoxic effects on several human cancer cell lines. The cytotoxic effect on the human liver cancer cells H7402 was further investigated and turned out to be attributed to the molecule's ability to induce apoptosis (Chen *et al.*, 2006b). Furthermore, sophorolipids tend to decrease mortality caused by septic shock in a rat model. It stays however a matter of debate whether the effect is caused by direct modulation of immune and inflammatory responses or by the antibacterial properties of the sophorolipid molecules (Bluth *et al.*, 2006; Napolitano, 2006).

Finally, sophorolipids are a source of difficult to synthesize ω and ω-1 hydroxy fatty acids. Those fatty acids can be used in polymerization reactions or can be lactonized into macrocyclic esters, which find application in the perfume and fragrance industry (Inoue and Miyamoto, 1980). Zerkowski and Solaiman (2006) produced fatty amines starting from sophorolipid derived 17hydroxy oleic acid. The compounds could be of interest in the preparation of highly functionalized polymers and surfactants.

## 4.10 Chemical or enzymatic modified sophorolipids and their applications

The simplest modification of sophorolipids, which at the same time reduces their structural variability, is conversion into the deacetylated acidic form by alkaline hydrolysis in an aqueous environment. If one wants to obtain sophorolipids merely lacking acetyl groups, enzymes such as acetylesterase (E.C. 3.1.1.6) or cutinase from *Fusarium solani* must be used. Acetylesterase removes both acetyl groups, whereas cutinase specifically hydrolyses the 6′ position (Asmer *et al.*, 1988; De Koster *et al.*, 1995).

The majority of the other modifications are carried out at the carboxylic end of the fatty acid. The one first reported in 1971 is the synthesis of sophorolipid alkyl esters to enhance the characteristics of prepared food products such as bakery and oily emulsions (Figure I-11). The authors found that the beneficial effects of the molecules increased with the chain length of the ester (Allingham, 1971). Zhang *et al.* (2004) observed the same trend. They synthesized and compared the properties of sophorolipid methyl, ethyl, propyl, and butyl esters and found that the critical micelle concentration decreases to about one half per additional carbon group to the ester moiety. Ashby *et al.* (2006) obtained sophorolipid methyl esters by simply applying fatty acid methyl esters of soy oil as the hydrophilic carbon source. However, the sophorolipid methyl esters comprise only 48 % of the total mixture, and no esterification is observed when ethyl or propyl soyate is used. The authors suggest that a lipase with only a partial activity towards methyl groups may be responsible for this phenomenon. There are a number of patents on the use of sophorolipid esters in cosmetics (Abe *et al.*, 1981).

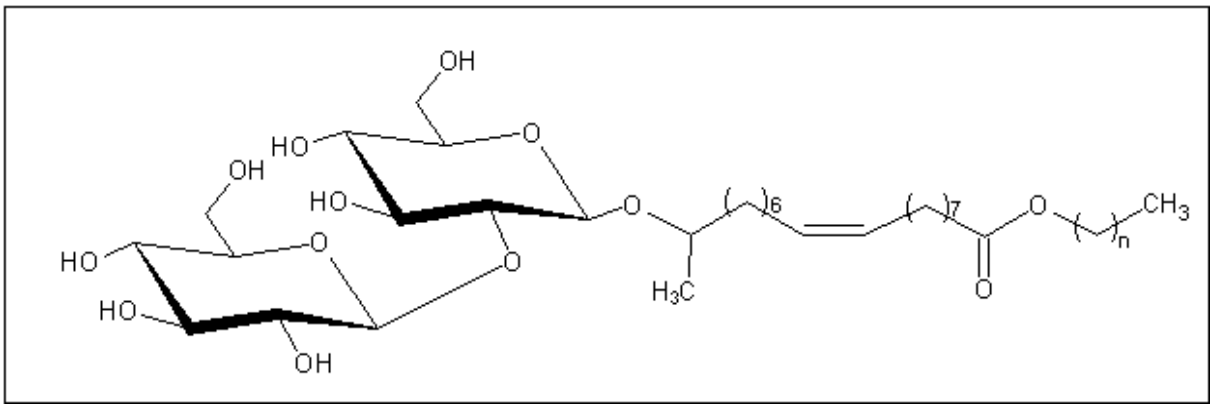


**Figure I-11:** Sophorolipid alkyl esters, n = 0, 1, 2, 3 or 5 (Zhang *et al.*, 2004).

The sugar moiety hydroxyl groups of the sophorolipid ester can be substituted with hydroxyalkyl groups, giving rise to hydroxypropyl-etherified glycolipids ester. These esters have been used in pencil-shaped lip rouge, lip cream, and eye shadow, in powdered compressed cosmetic material as well as in aqueous solutions (Kawano *et al.*, 1981a, b).
Sophorolipid esters can be reacetylated at the 6′-and 6″positions by Lipase Novozym 435 from *Pseudozyma* (*Candida*) *antarctica*. When sophorolipid methyl esters are reacted with the acylating agent at a concentration less than equimolar, lactonic sophorolipids are formed. The fatty acid carboxyl group is however linked to the sophorose unit by the 6″-instead of the 4″-hydroxyl found in the native form (Figure I-12). The lactonic sophorolipid can further be acetylated by Novozym 435 to lead to the formation of a 6′-monoacyl derivate (Bisht *et al.*, 1999). Singh *et al.* (2003) on the other hand developed a method for the direct regioselective acetylation of sophorolipid ethyl esters. With Novozym 435 or Lipase PS-C, they could mediate mono-acetylation at the 6′- or 6″-position, respectively.

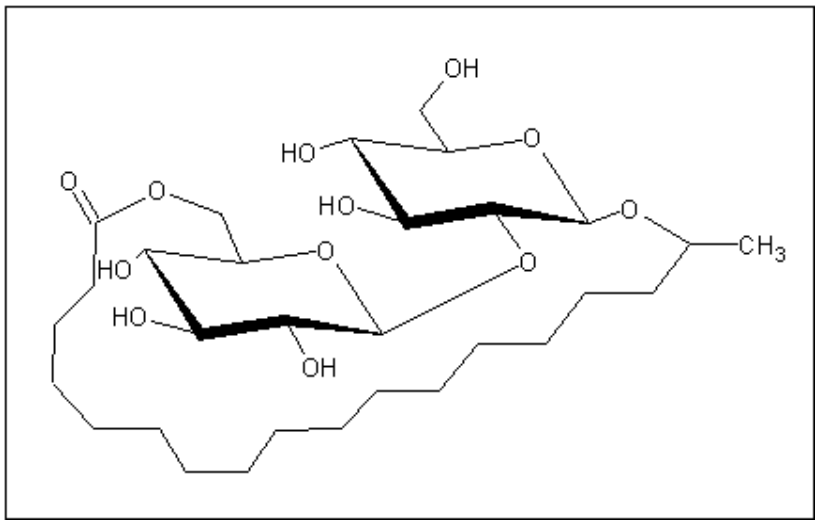


**Figure I-12:** Lactonic sophorolipid esterified at the 6”-position.

Furthermore, the ester itself can be subjected to further alterations. Carr and Bisht (2003) started from sophorolipid methyl-esters with all free hydroxyl groups of the sugar moiety blocked by peracylation. Novozym 435 turned out to be the only enzyme that could transesterify the peracetylated sophorolipid methyl esters with 1-butanol or 2-methylpropanol. Nuñez *et al.* (2003) applied more or less the same strategy to obtain a galactopyranose-sophorolipid. After reacetylation of a sophorolipid methyl-ester, they transesterified it with

1,2,3,4-di-O-isopropylidene-D-galactopyranose and again Novozym 435 as a catalyst. After acidic removal of the sugar hydroxyl group protection, a galactopyranose C6-linked to the carboxylic end of a nonacetylated sophorolipid molecule was obtained (Figure I-13). Other researchers synthesized amide derivates with Novozym 435 and stated that those derivates may have potential as tuneable immunoregulators. The introduction of methacryl or tyrosine groups on the other hand allows the molecules to be functional in polymerization processes (Singh *et al.*, 2003). However, sophorolipids can also be subjected to direct enzymatic polymerization. Hu and Ju (2003) optimized the reaction conditions for lipase-mediated conversion of diacetylated lactonic sophorolipids to monoacetylated lactonic sophorolipids, which were in the same reaction polymerized to oligomers and polymers, probably through ring-opening polymerization. Four different lipases were tested and polymerization took place in all cases, demonstrating the high promiscuity of lipase enzymes.

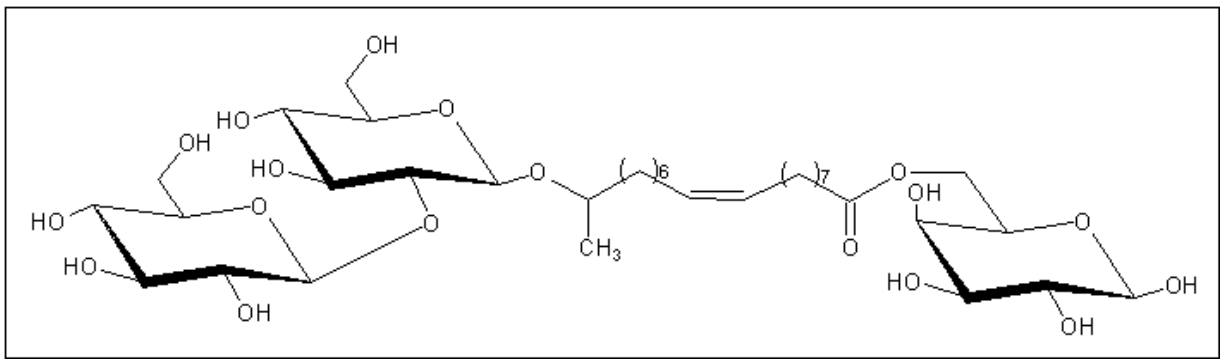


**Figure I-13:** Galactopyranose sophorolipid ester (Nuñez *et al.*, 2003).

A recently developed family of derivatives are the amino acid sophorolipid conjugates. The amino acids are coupled to the carboxylic end of acidic sophorolipids by using (di) carbodiimide. In this way, the nonionic sophorolipid can be converted to a cationic, zwitterionic, or anionic surfactant with increased water solubility and polar head groups that allow further chemical derivatization (Zerkowski *et al.*, 2006). Azim *et al.* (2006) evaluated the antibacterial, anti-HIV, and spermicidal activity of their conjugates. All molecules exhibited the desired action, but leucinesophorolipid was the most effective one.

A lot of examples are given on modification of the fatty acid carboxylic end and the sophorose acetylation pattern, but as far as we know, only one report on alteration of the sugar moiety itself is available. Rau and co-workers (1999) subjected deacetylated acidic sophoroli-

pids to various glycosidases. Hesperidinase from *Aspergillus niger* (E.C. 3.2.1.40) turned out to be the most active in the specific release of one glucose molecule. The surface active properties of this glucolipid are comparable to those of the acid sophorolipid, but its solubility in water is smaller.

# 5 Conclusion

Sophorolipids belong to the most promising biosurfactants. In contrast to rhamnolipids, which are synthesized by *Pseudomonas aeruginosa*, the producing strains are nonpathogenic. Very high production yields can be achieved (over 400 g/L), and this is based on renewable resources or even waste streams. The current production price amounts to 2 to 5 €/kg, depending on substrate cost and production scale. As sophorolipids find applications in the cleaning, environmental, and food industry as well as in the personal care, cosmetic, and pharmaceutical sectors, it is clear that their economical competitiveness depends on their final utilization. If sophorolipids are for example used in the cleaning industry, they have to vie with other environmentally friendly surfactants, such as the alkyl-polyglucosides (APGs), which have a market price of 2 €/kg. However, in the cosmetic or pharmaceutical sectors, higher price dimensions are standard, and therefore, sophorolipids can here easily compete.

Most research has been performed on the optimization of the fermentation process, but as far as we know, hardly any work has been published on the genetics of the producing yeast strain. However, genetic engineering of these yeast species could open up perspectives for higher yields and modification of the glycolipid mixture produced. In this context, a transformation and selection system was developed for *C. bombicola* (Chapter III) and several genes possibly involved in sophorolipid synthesis were cloned (Chapter VI).

# Chapter II: Synthesis of sophorolipids based on unconventional hydrophobic carbon sources.

Part of this chapter will be published as:

Van Bogaert INA, Develter D, Soetaert W, Vandamme EJ (2008) Production of new-to-nature sophorolipids by cultivating the yeast *C. bombicola* on unconventional substrates. Biotechnol Lett, in preparation

# 1 Introduction

As described in Section 4.3 of Chapter I, sophorolipids are a diverse group of molecules differing in their degree of acetylation, presence or absence of lactonization, position of the fatty acid hydroxyl group, fatty acid chain length and saturation. Despite these variations, the fatty acid tail of *de novo* sophorolipids is in general limited to 16 or 18 carbon atoms, with stearic acid (C18:0) being the dominant component (Table II-1, sophorolipids on glucose only). Consequently, vegetable oils - which are rich in C16-18 fatty acids - are very suitable as hydrophobic carbon source for sophorolipid synthesis and are readily incorporated into the sophorolipid molecule. Vegetable oils have a low content of saturated fatty acids, and only contain stearic acid as a minor component. However, excellent results are obtained with oils rich in oleic acid such as rapeseed oil (Davila *et al.*, 1994).

In most cases, the fatty acid composition of the vegetable oil used is reflected in the fatty acid pattern of the sophorolipid mixture (Table II-1), illustrating the direct incorporation of the substrates.

**Table II-1:** (Hydroxy) fatty acid composition of vegetable oil feed stocks and of the corresponding sophorolipids. After Davila *et al.* (1994).

| | **(Hydroxy) fatty acid composition (%)** | | | | |
|---|---|---|---|---|---|
| **PRODUCT** | **C16:0** | **C18:0** | **C18:1** | **C18:2** | **C18:3** |
| Rapeseed oil | 4.9 | 1.5 | 60.1 | 23.1 | 10.4 |
| Sophorolipids on rapeseed oil | 3.1 | 2.0 | 81.0 | 13.9 | - |
| Safflower oil | 6.8 | 4.9 | 22.8 | 65.5 | <0.1 |
| Sophorolipids on safflower oil | 4.4 | 8.7 | 37.9 | 49 | - |
| Palm oil | 43.2 | 5.6 | 39.7 | 11.5 | <0.1 |
| Sophorolipids on palm oil | 31.8 | 12.9 | 47.8 | 7.5 | - |
| Sophorolipids on glucose only | 27.6 | 38.4 | 34.0 | <0.2 | - |

Nevertheless, when a hydrophobic substrate with medium-chain fatty acids such as coconut oil is used, the fatty acid pattern is not reflected in the sophorolipid mixture. Those fatty acids are too short for efficient hydroxylation by the cytochrome P450 monooxygenase (P450) to

the corresponding hydroxy fatty acid and consequently can not be incorporated into the sophorolipid molecule. It turns out that the affinity of the P450 enzyme for a certain substrate depends on its absolute length: the enzyme converts fatty acids of 22.55 Å to 25.05 Å very easily, but once the length diverts from this optimal range, no or only very low activity is observed. For further information about this phenomenon, please see Sections 4.4 and 4.5.2 of Chapter I. The effect of this length specificity is illustrated by the efficiency of incorporation of fatty acids from vegetable oils and their derived fatty acid ethyl esters (Table II-2). Stearic and oleic acid are readily incorporated, while the affinity for palmitic acid is already much lower.

**Table II-2:** Incorporation of fatty acids from oil or ester feed stocks (Davila *et al.*, 1994)

| | **Efficiency for incorporation of fatty acids (mol/mol)** | | | | |
|---|---|---|---|---|---|
| **Feed stock** | **C16:0** | **C18:0** | **C18:1** | **C18:2** | **C18:3** |
| Rapeseed esters | 0.44 | 1.21 | 1.14 | 0.60 | <0.2 |
| Rapeseed oil | 0.44 | 0.98 | 0.96 | 0.48 | <0.2 |
| Safflower esters | 0.48 | 1.05 | 0.92 | 0.59 | - |
| Safflower oil | 0.37 | 1.01 | 0.95 | 0.43 | - |
| Palm esters | 0.81 | 0.94 | 1.11 | 0.94 | - |
| Palm oil | 0.51 | 1.10 | 0.84 | 0.45 | - |

In the first part of this chapter, the strict upper and lower boundaries of hydrophobic carbon source incorporation are verified. However, in order to force *C. bombicola* to synthesize sophorolipids with a non conventional hydroxy fatty acid moiety, the substrate specificity of the cytochrome P450 monooxygenase must be circumvented. This can be done on two ways. The first method comprises the addition of already hydroxylated substrates, in this way skipping the controlling effect of the P450 enzyme. The other method is the addition of a hydrophobic carbon source with a stearic acid-like structure or chain length, which can then be hydroxylated by P450 and build into the sophorolipid molecules. In some cases, one can even further chemically modify the fatty acid moiety of the sophorolipid molecule after the

fermentation (Develter and Fleurackers, 2007). The use of both methods for the creation of new-to-nature sophorolipids is described in this chapter as well.

# 2 Materials and Methods

## 2.1 Strains and culture conditions

*Candida bombicola* ATCC 22214 was used in all experiments. The medium described by Lang *et al.* (2000) was applied for sophorolipid production. Shake flask cultures (200 mL culture medium) were incubated at 30 °C and 200 rpm. Feeding of the hydrophobic carbon source was started 48 hours after inoculation, with a total amount of 37.5 g/L unless stated otherwise. The hydrophobic carbon sources used in the shake flask and fermentor scale experiments are listed in Table II-3. Incubation is stopped 8 days after the first addition of hydrophobic carbon source unless stated otherwise.

**Table II-3:** Hydrophobic carbon sources used in this chapter.

| Hydrophobic carbon source | Supplier |
|---|---|
| Meadowfoam oil | Natural Plant Products Inc. |
| Coconut oil | Sigma |
| Caprylic acid | Sigma |
| Capric acid | Sigma |
| Lauric acid | Sigma |
| 12-Hydroxydodecanoic acid | Sigma |
| 1,12-Dodecanediol | Sigma |
| Dodecyl glutarate* | Ecover Belgium NV |
| Dodecyl and myristyl malonate* | Ecover Belgium NV |
| Pentenyl dodecanoate* | Ecover Belgium NV |
| Dodecyl pentanoate* | Ecover Belgium NV |

*Substrates were specially synthesized for the described experiments by esterification of a fatty alcohol and a (di)carboxylic acid.

Fed-batch fermentations were run in a Biostat® B culture vessel (Sartorius-BBI Systems) with a maximum working volume of 1.5 to 3 L. Temperature (30 °C), pH (3.5), stirring rate (800

rpm) and airflow rate (1 vvm) were controlled by the Biostat® B control unit. 100 mL of an overnight grown shake flask culture was used to inoculate the fermentor and together with the inoculum, 4 g/L rapeseed oil was added to adapt the yeast to the hydrophobic substrates and to stimulate lipid body formation. For maintaining pH, 5 N NaOH was used. There was no correction for a too alkaline pH and fermentations started at pH 5.8 and were allowed to drop spontaneously till 3.5. Later unalterable incensement was seen as the end of the fermentation process. Feeding of the hydrophobic carbon source was started on average 48 hours after inoculation, the exact amount and feeding rate depended on the specific substrate used. Additional glucose was added on average 150 hours after inoculation and depending on the fermentation, if needed more supplementary glucose was added later. For some fermentations, foam could be controlled by addition of the hydrophobic carbon source, but in some cases a solution of 10 % silicone antifoaming agent (BDH 331512K from VWR Int Ltd.) had to be added to the culture vessel.

## 2.2 Sampling

### 2.2.1 Biomass formation

Cell growth of the cultures was monitored in different ways. The cell density was determined by measuring the optical density at 600 nm using an Uvikon 922 spectrophotometer (BRS). Samples with an optical density higher then 1 were diluted with distilled water, allowing measurement in the linear range of the spectrophotometer. This method is reliable in the exponentional growth phase of the cells, but becomes less valid in the stationary phase when lipophilic substrate is added and sophorolipids are formed. Both substances can interfere with optical density measurements.

Cell dry weight was measured by centrifugation of 2 mL culture broth for 5 min at 9 000 g. Pellets were washed two times with ethanol to remove sophorolipids and lipophilic substrate and finally dissolved in distilled water. The suspension was transferred to a cellulose nitrate filter with a pore diameter of 0.45 µm (Sartorius) and the dry weight was determined in the XM60 automatic oven from Precisa Instruments Ltd.

Because some substrates are toxic to *C. bombicola*, biomass formation was also monitored by determination of the colony forming units (CFU). Decimal dilutions till $10^{-7}$ were plated on agar plates with 10 % glucose, 1 % yeast extract and 0.1 % urea and incubated at 30 °C for three days. In case of presumed high toxicity, cell viability was also checked by light microscopy; affected cells are much smaller and thinner than normal cells and do not posses the big lipid bodies as seen for healthy ones.

### 2.2.2 Glucose concentration

1 mL culture broth was centrifuged for 5 min at 9 000 g and the supernatant was analysed with the 2700 Select Biochemistry Analyzer (YSI Inc.). Samples were diluted in such way that they did not exceed the upper limit of efficient detection of 7 g/L.

### 2.2.3 Sophorolipid formation

Analytical sophorolipid samples were prepared as follows: 440 µL ethyl acetate and 11 µL acetic acid were added to 1 mL culture broth and shaken vigorously for 5 min. After centrifugation at 9 000 g for 5 min, the upper solvent layer was removed and put into a fresh eppendrof tube with 600 µL ethanol. Samples were analysed as described in Section 2.4.

## 2.3 Final sophorolipid extraction from culture broth

This protocol has been developed by Steve Fleurackers from Ecover Belgium NV (Belgium). The yeast cells were removed by adding 3 volumes of ethanol to the residual fermentation medium. After centrifugation at 1500 g, the supernatants was transferred to a boiling flask and the water-ethanol mixture is evaporated in a rotavapor. 2 volumes of ethanol were added to dissolve the sophorolipids and the residual hydrophobic carbon source. Water soluble components will not dissolve. The ethanol mixture was transferred to a Whatman filter by vacuum filtration; the filtrate (free of water soluble components) was put into a fresh boiling flask and the ethanol was evaporated in a rotavapor. 1.5 volume of diethyl ether was added to dissolve the residual hydrophobic carbon source and the diethyl ether mixture was transferred to a Whatman filter with vacuum filtration. The residue which was left in the boiling flask

consists of the sophorolipids and is dissolved again in one volume of ethanol. The filtrate contained the residual hydrophobic carbon source and was put into a fresh boiling flask to evaporate the diethyl ether. The solid residue left after evaporation is the residual hydrophobic carbon source.

Weights of the different fractions can be determined by determining the empty weights of the boiling flask prior to the extraction. HPLC sample can be taken during and at the end of the extraction procedure.

## 2.4 High performance liquid chromatography analysis

Sophorolipid samples were analysed by High Performance Liquid Chromatography (HPLC) on a Varian Prostar HPLC system using a Chromolith® Performance RP-18e 100-4.6 mm column from Merck KGaA at 30 °C and Evaporative Light Scattering Detection (ELSD, Alltech). A gradient of two eluents, a 0.5 % acetic acid aqueous solution and acetonitrile, had to be used to separate the components. The gradient started at 5 % acetonitrile and linearly increased till 95 % in 40 min. The mixture was kept this way for 10 min and was then brought back to 5 % acetonitrile in 5 min. A flow rate of 1 mL/min was applied.

## 2.5 Liquid chromatography mass spectroscopy analysis

Liquid chromatography mass spectroscopy (LCMS) analysis was performed by Intertek ASG (Manchester, UK) with a Micromass Quattro Ultima LIMS 1107 (Waters). The detection range was set at m/z 100 to 1000 and the negative ion mode was applied. The same column and LC conditions as described for the HPLC analysis were used (Section 2.4).

## 2.6 Gas chromatography and mass spectroscopy analysis

The fatty acid part of the sophorolipids was analysed by gas chromatography (GC) and mass spectroscopy (MS) analysis. Prior the analysis, the hydroxy fatty acids of the sophorolipids were split of and modified to fatty acid methyl esters (FAME's) by the following procedure.

50 mg of dried sophorolipids were resolved in 1 mL methanol and transferred to a glass tube with a well closing screw cap. 1 mL $H_2SO_4$ dissolved in methanol (2 %) was added and the

tubes were placed in a warm water bath at 100 °C for 45 min. The mixture was transferred to a 15 mL falcon tube. 2 mL MQ-water and 3 mL hexane was added and the mixture was shaken vigorously prior to centrifuge at 1000 g for 2 min. The upper hexane phase was transferred to a glass tube with screw cap which was put into a warm sand bath were the hexane was evaporated under a nitrogen flow. The residue was resolved in 30-50 µL hexane and a small amount of $Na_2SO_4$ was added. This sample can be injected on the GC-MS apparatus.

GC-MS analysis was performed with the 6890N Network GC System and the 5973 Network Mass Secetive Detector from Agilent Technologies. Samples were brought on a CP-Wax 52 CB column (Varian) by direct injection at a temperature of 240 °C and a split ratio of 50/1. Helium was applied as carrier gas at a flow rate of 1 mL/min. The column oven begin temperature was set at 45 °C and increased till 240 °C with a rate of 10 °C/min.

## 2.7 Alkaline hydrolysis of sophorolipids

This protocol was developed and performed by Steve Fleurackers from Ecover Belgium NV (Belgium). 18.7 g sophorolipids were dissolved in 100 mL water and the pH was set at a value higher then 12 with a 50 % KOH solution. This was incubated overnight at room temperature. A clearly alkaline pH (> 10) indicated that hydrolysis was complete. If this was not the case, additional KOH solution was added and the mixture was again incubated overnight. If necessary, this was repeated.

The solution was neutralized by adding a few drops of 100 % acetic acid to set pH at 7.3. An additional centrifugation step at 16000 g for 5 min could be necessary if the solution is troubled by the release of *e.g.* glutaric acid or acetic acid.

# 3 Results and Discussion

## 3.1 Fermentation run on vegetable oils with a special fatty acid composition

In order to obtain sophorolipids with an unconventional fatty acid tail, oils with a special fatty acid composition were used as hydrophobic carbon source in shake flask experiments. A culture medium with high glucose content was used (120 g/L), as it is thought that a high glucose concentration could more or less prevent that the fatty acids are metabolized in the β-oxidation cycle (Lang *et al.*, 2000).

### 3.1.1 Meadowfoam oil

Meadowfoam oil is pressed from the seeds of meadowfoam (*Limnanthes alba*) and contains over 98 % fatty acids having 20 carbon atoms or more. The exact fatty acid composition is given in Table II-4.

**Table II-4:** Fatty acid composition of meadowfoam oil.

| Fatty acid | Percentage (%) |
|---|---|
| <C:20 | 2 |
| C20:0 | 1 |
| C20:1 Δ5 | 63 |
| C22:1 Δ5 | 4 |
| C22:1 Δ13 | 12 |
| C22:2 Δ5,13 | 17 |
| others | 1 |

The formation of CFU's was similar as compared to fermentations on conventional hydrophobic carbon sources; meadowfoam oil was not toxic to the *C. bombicola* cells.
During the incubation time, the oil stayed present on the culture surface and was still visible at the end of the experiment. This indicated that meadowfoam oil was not or not completely

used as a substrate for sophorolipid production. After separation of the sophorolipids and the residual oil, it turned out that 77 % of the initially added oil was still present and consequently not incorporated into sophorolipids. Furthermore, the overall sophorolipid yield was not so high (only 16.7 g/l) and is comparable to yields achieved on media without addition of any hydrophobic carbon source (*de novo* synthesis of sophorolipids). The same trend is observed when looking at the HPLC analysis of the sophorolipid mixture: the conventional *de novo* formed sophorolipids pattern was retrieved. Nevertheless, a very small hydrophobic peak (only a few percentages of the total sophorolipid mixture) was detected, which could probably by a lactonic sophorolipid molecule with a C20 fatty acid. GC-MS analysis did however not confirm this finding.

The above results led to the conclusion that meadowfoam oil is hardly incorporated in sophorolipid molecules and consequently is not suited as hydrophobic carbon source. The same phenomenon was observed for fish oil, an oil which even so contains fatty acids of 20 or more carbon atoms (Davila *et al.*, 1994).

On the other hand, Phadtare and colleagues (2003, 2004) succeeded in producing 20-hydroxyeicosatetraenoic acid with immobilised *C. bombicola* cells starting from arachidonic acid (C20:4 Δ5,8,11,14). The absolute length of arachidonic acid is however much shorter than the length of eicosenoic acid (C20:1) due to the four unsaturated bonds and it is therefore believed that arachidonic acid is hydroxylated much easier than eicosenoic acid.

### 3.1.2 Coconut oil

Coconut oil is one of the rare vegetable oils which are rich in medium-chain saturated fatty acids. The fatty acid composition of commercially available coconut oil is given in Table II-5. As also reported by other researchers, the use of coconut oil for sophorolipid production did not result in the synthesis of medium-chain sophorolipids (Davila *et al.*, 1994). For some fermentations on coconut oil, a reduction in biomass formation was observed. This confirmed the results of Ogawa and Ota (2000). Especially when the glucose concentration was low, the number of colony forming units dropped. This could be explained by the fact that the yeast starts using coconut oil as carbon source and that the presence of medium-chain fatty acids - released from the coconut triacylglycerides by esterases - has a toxic effect on the yeasts cells.

It is demonstrated that capric acid (C10:0), lauric acid (C12:0) and their 1-monoglycerides have not only microbicidal activity against enveloped viruses and various bacteria, but also against the yeast *C. albicans*. Even at concentrations lower than 10 mM and short incubation times, a clear inhibitory effect was observed due to the disruption or disintegration of the plasma membrane which causes disorganisation and shrinkage of the cytoplasm (Bergsson *et al.*, 2001). Because of the potential toxic effect of free fatty acids, it was decided to test their effect on *C. bombicola* cells in the stationary phase. Fatty acids were added after 48 h cultivation time at a concentration of 10 mM, in this way imitating the addition of the hydrophobic carbon source. The results are shown in Figure II-1. While caprylic acid (C8:0) seems to be extremely toxic with a 100 % killing of the yeast cells already after one hour of incubation, the maximal effect of capric acid (C10:0) was reached after a few hours of incubation and was not complete. For fatty acids with a chain length over 10 carbon atoms, no lethal effects were observed at the tested concentration of 10 mM. After this first investigation of the toxicity of fatty acids, further experiments with different concentrations were conducted.

**Table II-5:** Fatty acid composition of coconut oil.

| Fatty acid | Percentage (%) |
|---|---|
| C6:0 | ND - 0.7 |
| C8:0 | 4.6 - 10.0 |
| C10:0 | 5.0 - 8.0 |
| C12:0 | 45.1- 53.2 |
| C14:0 | 16.8 - 21.0 |
| C16:0 | 7.5 - 10.2 |
| C18:0 | 2.0 - 4.0 |
| C18:1 | 5.0 - 10.0 |
| C18:2 | 1.0 - 2.5 |
| C18:3 | ND - 0.2 |

ND: not detected

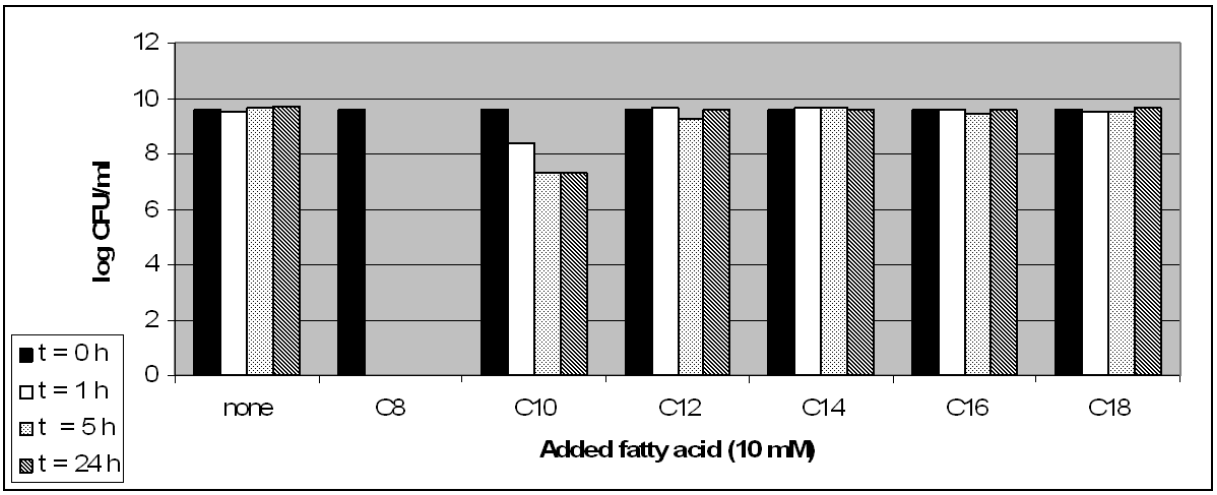


**Figure II-1:** Effect of 10 mM fatty acids on CFU formation after 1, 5 and 24 hours of incubation.

Because of the extreme effect of caprilic acid, this fatty acid was further tested in lower concentrations to find the level for which no toxicity is observed. As demonstrated in Figure II-2, it turned out that 2 mM could kill all *C. bombicola* cells after 5 hours of exposure, while 5 mM already has the same effect after 30 minutes. Only the very low concentration of 1 mM did not seem to have any effect, also after 24 hours of incubation. These results hereby confirm the severe toxicity of caprylic acid toward *C. bombicola* cells.

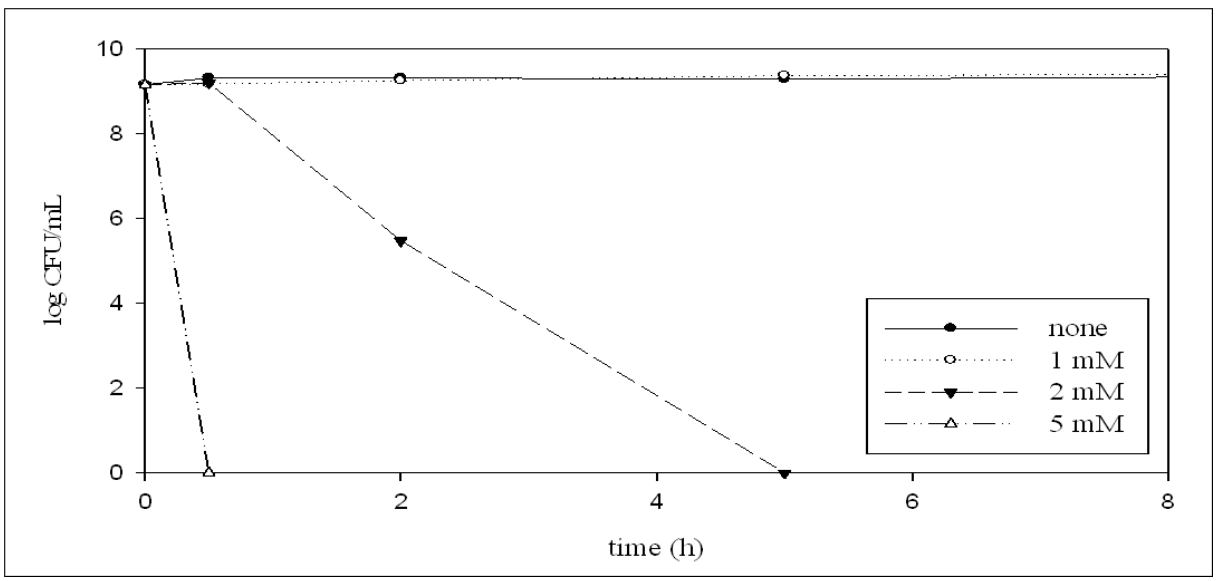


**Figure II-2:** Effect of different concentrations of caprilic acid on cell viability. The results after 24 h were the same as those after 5 h and are therefore not shown in the graph.

Also the toxicity of capric acid was investigated more in detail (Figure II-3). Both concentrations higher and lower than 10 mM were tested. Concentrations of 5 mM or more turned out to have an inhibitory effect. However, this effect was not that drastic and was not observed so fast compared to caprylic acid. Even at a concentration at 100 mM and 24 hours of incubation, there were still viable cells. When looking at the killing curves, one can see that there is a certain resistance towards the fatty acid and that the effect of 40, 70 and 100 mM is more or less the same.

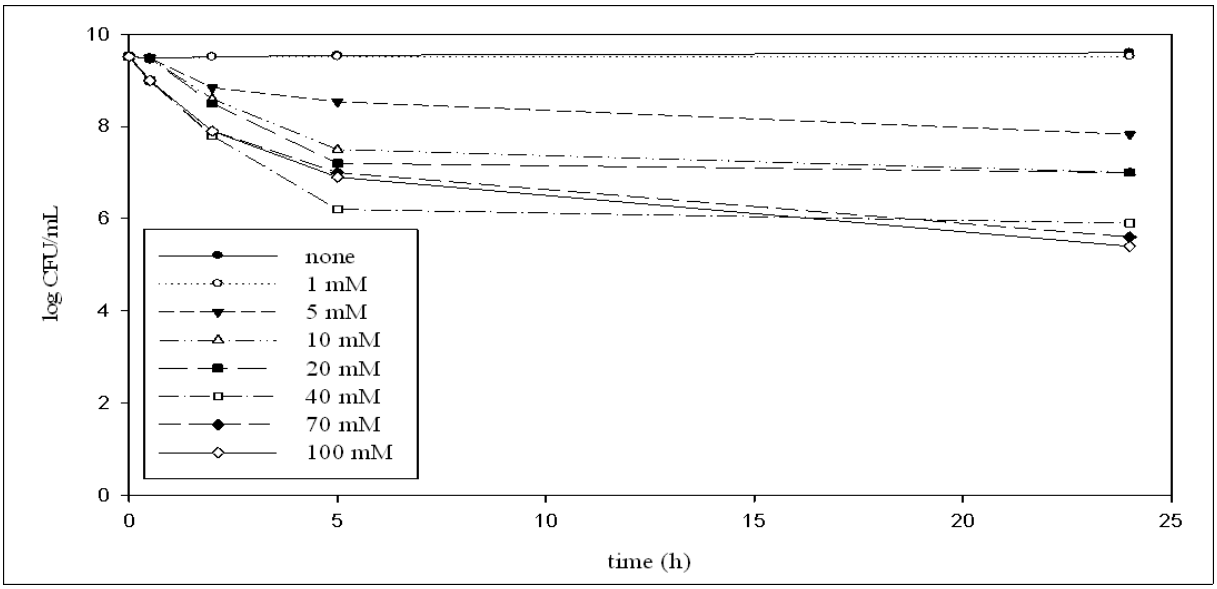


**Figure II-3:** Effect of different concentrations of capric acid on cell viability.

Finally, also the toxicity of lauric acid was further tested (Figure II-4). This fatty acid has a low toxicity towards *C. bombicola* cells. Only for the very high concentrations of 70 and 100 mM there is a minor effect.

The above experiments illustrate the toxicity of medium-chain fatty acids towards microorganisms by their effect on *C. bombicola*. The killing effect on the yeast cells increases when the fatty acid chain length decreases, as also demonstrated by Bergersson *et al.* for *C. albicans* (2001). Although *C. bombicola* is rather resistant to those substances when compared to *C. albicans*, this yeast turns out to be quite sensitive to caprylic acid: 2 mM can kill all cells in a fermentation medium after 5 hours of incubation. When calculating this back to coconut oil (triacylglycerides, with 7 % caprylic acid), this means that a theoretical value of

4.4 g/l coconut oil has a total lethal effect while 2.2 g/l, corresponding with 1 mM caprylic acid, has no effect. Nevertheless, when 37.5 g/l coconut oil is added to the fermentation medium, no direct and such extreme toxicity is observed. This could be explained by the fact that the caprylic acid in the oil is present as part of a triacylglyceride molecule and is only released gradually as free fatty acid in the medium as a result of the activity of esterases.

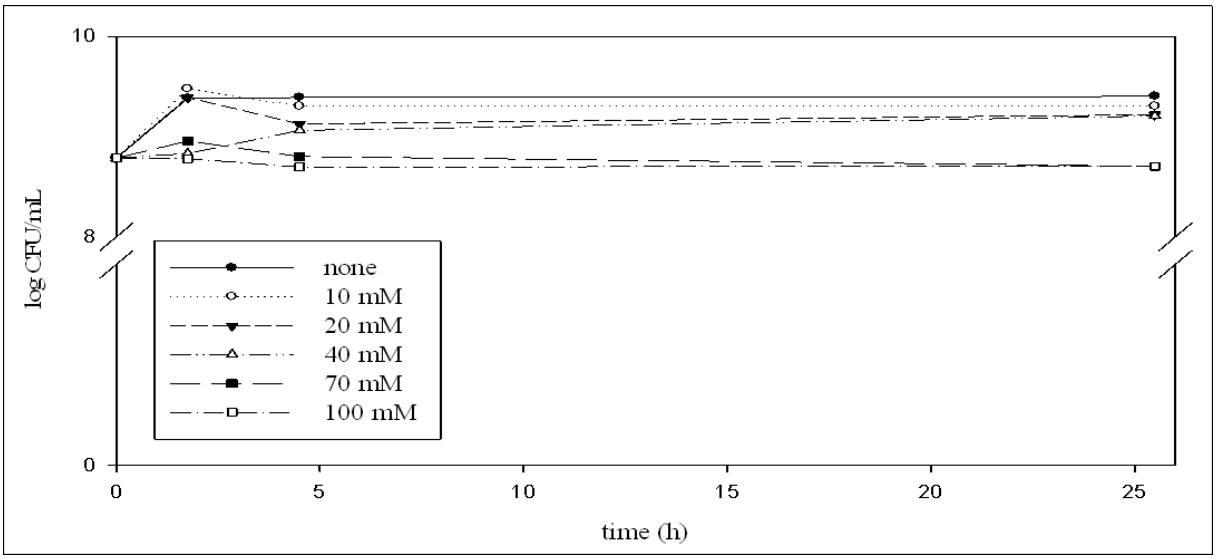


**Figure II-4:** Effect of different concentrations of lauric acid on cell viability. Please notice that the intercept of the Y-axis with the X-axis differs from the previous graphs.

## 3.2 Fermentations run on already hydroxylated substrates

As mentioned in the general introduction (Section 1 of this chapter), one method to circumvent the selectivity towards C16 and C18 fatty acids incorporation in sophorolipids is the addition of already hydroxylated substrates. Here the use of two of such substrates will be described.

### 3.2.1 12-Hydroxydodecanoic acid

Hydroxylated fatty acids are not always that easy to purchase; one can find some terminal hydroxylated fatty acids such as 12-hydroxydodecanoic acid and 16-hydroxyhexadecanoic acid at high prices, but subterminal hydroxylated fatty acids are not available on the commercial market. Since 16-hydroxyhexadecanoic acid is an intermediate product in the conventional sophorolipid synthesis giving rise to a sophorolipid molecule with a C16 fatty acid tail, it was decided to test 12-hydroxydodecanoic acid (Figure II-5). This hydrophobic substrate could possibly lead to the synthesis of unconventional sophorolipids. The core fatty acid itself, lauric acid, is to short for efficient incorporation; integration into the sophorolipid molecule should consequently be totally contributed by the terminal hydroxyl group.

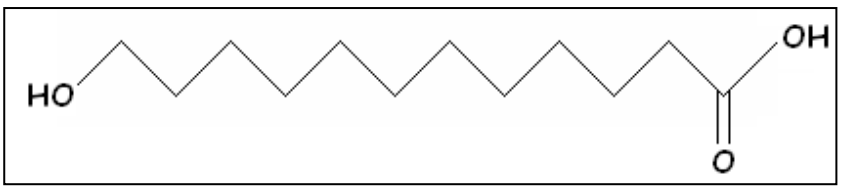


**Figure II-5**: Chemical structure of 12-hydroxydodecanoic acid.

Because unconventional substrates can be toxic for *C. bombicola* cells, 12-hydroxydodecanoic acid was added little by little: 2, 3, 6 and 8 days after inoculation 10, 5, 7.5 and 7.5 g/L was added respectively, making up a total of 30 g/L. A gradual decline in cell viability was indeed observed: the CFU's dropped from $3.10^9$ CFU/mL in the onset to $7.10^7$ CFU/mL at the end of the fermentation period. Nevertheless, glucose was almost completely consumed after this period, indicating the metabolic activity of the cells. The total extraction of sophorolipids from the culture broth yielded 22 g/L end-product, which is higher as

compared to simultaneously performed cultivations without addition of any hydrophobic carbon source (15 g/L) or with the addition of dodecanoic acid ethyl ester as the hydrophobic carbon source (11 g/L).

Upon HPLC-analysis, a clear peak in the hydrophilic region was detected, whereas in a parallel control experiment with dodecanoic acid ethyl ester, no new peaks were observed. This new peak is very likely caused by a new sophorolipid molecule containing a 12-hydroxydodecanoic acid tail. To support this hypothesis, a GC-MS analysis on the fatty acid moiety of the sophorolipid mixture was conducted as described in the Materials and Methods section. Because in this procedure fatty acid methyl esters are made not only from the sophorolipid fatty acid tail, but also from free fatty acids (FAME's), triacylglycerides or other esterified fatty acids, an additional purification was performed before the methylation procedure in order to remove any residual 12-hydroxydodecanoic acid which could intervene with a correct determination of the amount of incorporated substrate. Several solvents were tested and evaluated by HPLC-analysis; it turned out that washing a small amount of sophorolipids (50 mg) 3 times with diethyleter could totally remove the residual 12-hydroxydodecanoic acid. GC-MS analysis indeed confirmed that 12-hydroxydodecanoic acid was incorporated into the sophorolipid molecules; at least 51 % of the sophorolipid derived fatty acids were 12-hydroxydodecanoic acids.

The described results illustrate that medium-chain fatty acids can be incorporated into sophorolipid molecules if they are already hydroxylated upon addition to the fermentation medium, in this way circumventing the selective activity of the cytochrome P450 monooxygenase(s) responsible for hydroxylating of the C16 and C18 fatty acids in the *C. bombicola* cells.

### 3.2.2 1,12-Dodecanediol

Brakemeier and colleagues (1995, 1998a, 1998b) demonstrated that *C. bombicola* can synthesise sophorolipids when primary and secondary alcohols with medium-chain length are used as hydrophobic carbon source. The unusual hydrophobic carbon sources are to a large extent incorporated into the sophorolipid molecules. This method to obtain sophorolipids with a medium-chain fatty acid tail is unfortunately covered by several patents (Lang *et al.*, 2002;

Wullbrandt *et al.*, 1996) and the optimal substrate – secondary alcohols – are quite expensive to use for commercial fermentations. Diols are however not covered by the patents and are – compared to secondary alcohols – less expensive. 1,12-Dodecanediol (Figure II-6) was selected here since the hydroxyl groups positioned on both ends of the carbon chain make the molecule very attractive for incorporation.

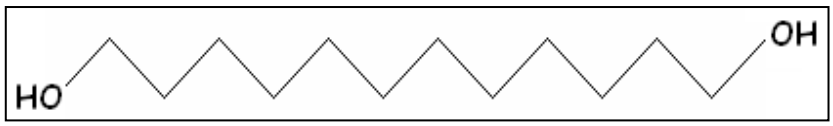


**Figure II-6**: Chemical structure of 1,12-dodecanediol.

### 3.2.2.1 Shake flask experiment

In a first shake flask experiment, 1,12-dodecanediol was added in two portions to circumvent any potential toxic effects: a first amount of 18.75 g/L was added 2 days after inoculation, while a second equal amount was added 4 days after the first supplementation. The *C. bombicola* cells turned out to be quite tolerant towards this diol; during the first half of the incubation period the CFU numbers were similar to those from a parallel set-up without addition of any hydrophobic carbon source. The CFU's declined only slightly at the end of the incubation period (no more than 0.6 log units compared to the highest obtained value).

A total yield of 45 g/L sophorolipids was obtained, which is quite high for shake flask fermentations. Furthermore, there was hardly any residual substrate at the end of the incubation period. HPLC-analysis indeed revealed several new peaks in the hydrophilic area (Figure II-7). The substrate itself, 1,12-dodecanediol, also elutes in this region, but can clearly be distinguished from the sophorolipids. The appearance of so many hydrophilic peaks is somewhat unusual; it is suggested that sophorolipids with sophorose molecules at both ends of the 1,12-dodecanediol are formed. Such molecules were also described by Braekemeier *et al.* (1998a) when using secondary alcohols (Figure I-10). These molecules can further be mono- or di-acetylated, resulting in a pattern of peaks. Another remarkable phenomenon is the discrepancy between the samples of the daily extractions with ethyl acetate and 2.5 % acetic acid (Figure II-8) and the end extraction with water and ethanol (Figure II-7). It turns out that when applying the first method, not all hydrophilic sophorolipids are extracted.

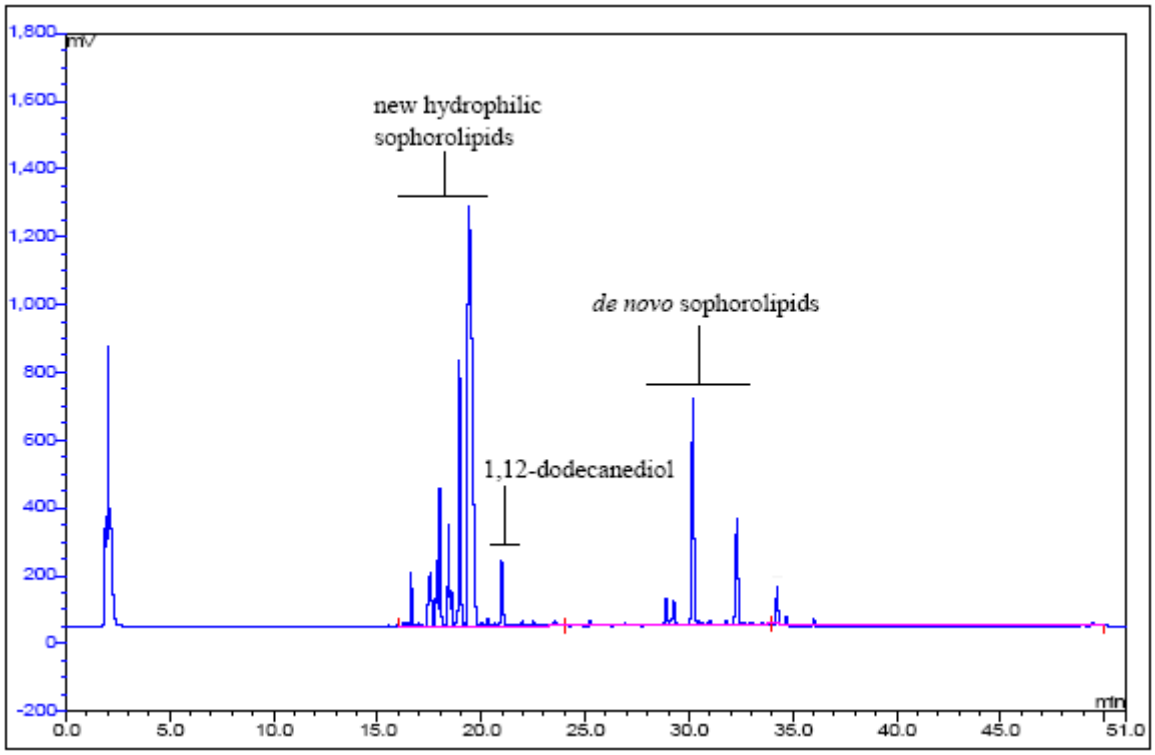


**Figure II-7:** HPLC-chromatogram of sophorolipids obtained from a fermentation with 1,12-dodedanediol as hydrophobic carbon source. Sophorolipids were extracted with water and ethanol (end extraction) as described in Section 2.3 of this chapter.

Analysis by GC-MS confirmed the incorporation of 1,12-dodecandiol into sophorolipids; 77 % of the sophorolipid fatty acids were derived from 1,12-dodecanediol. One fourth of it was present as 12-hydroxydodecanoic acid, indicating the *C. bombicola* started oxidizing the substrate. At this stage, it can still be build into the sophorolipid molecules (see Section 3.2.1) or it can be metabolized in the β-oxidation pathway.

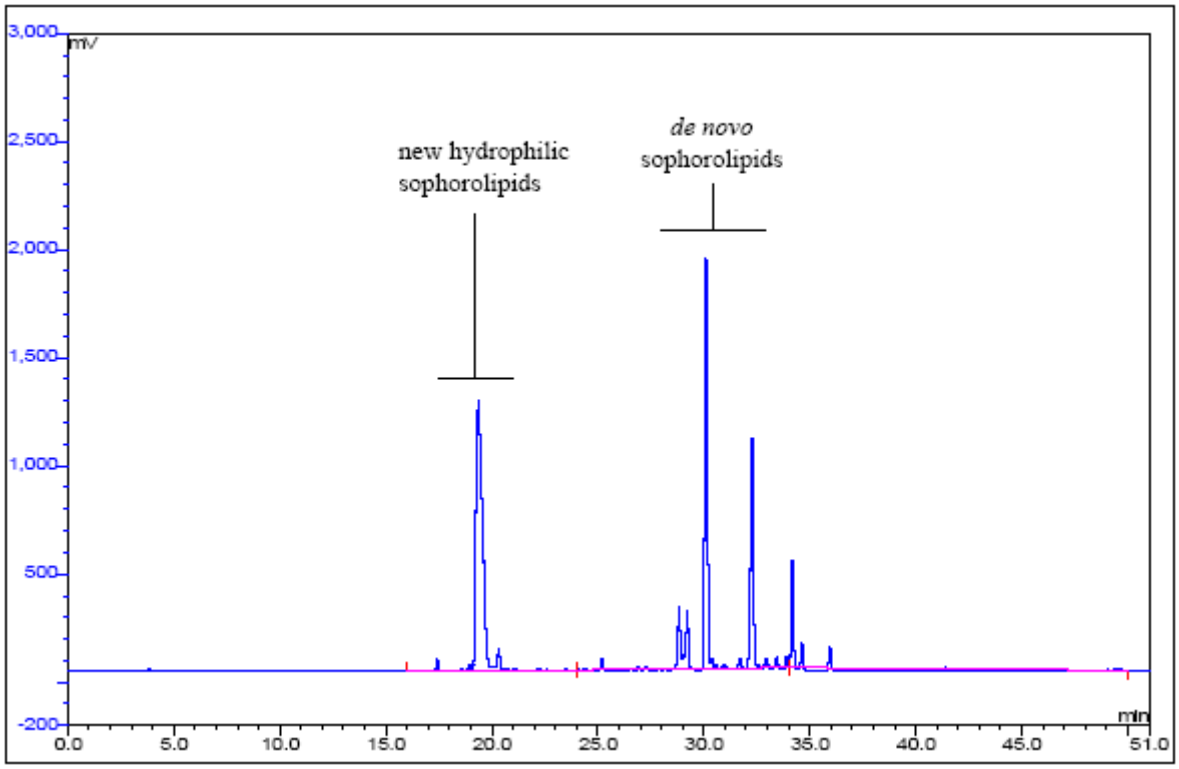


**Figure II-8:** HPLC-chromatogram of sophorolipids obtained from a fermentation with 1,12-dodedanediol as hydrophobic carbon source. Sophorolipids were extracted with ethyl acetate and 2.5 % acetic acid (daily sampling) as described in Section 2.1.2.3 of this chapter.

### 3.2.2.2 Fermentor scale experiment

Because of the good results obtained in the shake fask experiment, a larger fermentation was run in a batch fermentor (1.5 L working volume). Since the poor solubility of the substrate, 1,12-dodecanediol was added daily in small amounts (7.5 to 12.5 g) once the stationary phase was reached (Figure II-9). After 145 hours of fermentation, only a few grams per litre of glucose was left in the medium and 50 g/L glucose was added. When stopping the fermentation after 242 hours of cultivation, glucose was totally consumed and 101.3 g sophorolipids were extracted from the remaining 1.2 L culture broth (84.4 g/L). A total of 194 g glucose and 82.5 g 1,12-dodecanediol was added and only 0.5 g/L residual hydrophobic carbon source was removed from the crude sophorolipid mixture.

Theoretically, 0.408 mol (or 82.5 g) 1,12-dodecanediol can give rise to 0.408 mol (or 214.8 g) sophorolipids (MW 526 g/mol, with the simplification that they possess only one

sophorosemolecule). The obtained mass of sophorolipids was yet much lower; when assuming that all formed sophorolipids have incorporated 1,12-dodecanediol, only 0.1924 mol is formed. This brings the yield to 0.47 mol sophorolipids per mol 1,12-dodecanediol; this is still an overestimation since the *de novo* formed sophorolipids are not taken into account. For example, when assuming an incorporation of 77 % (GC-MS analysis), a yield of 0.36 mol/mol is obtained. Furthermore, since hardly any residual substrate was left, it is suggested that 1,12-dodecanediol was to some extent metabolized by the yeast.

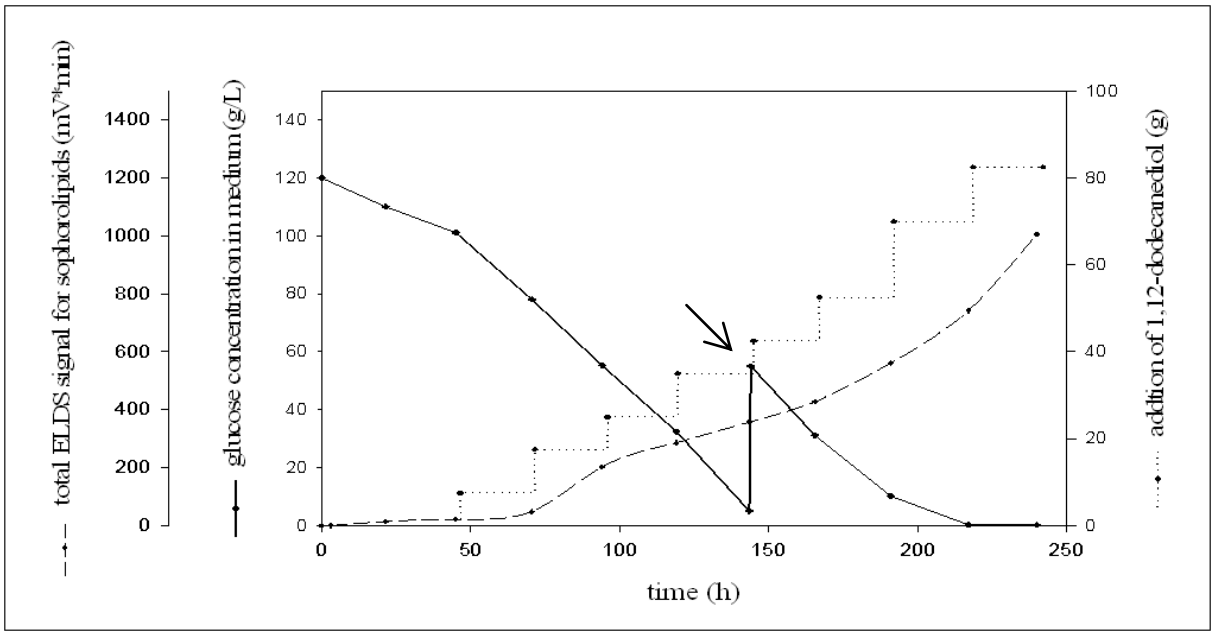


**Figure II-9:** Values for the glucose concentration in the medium (supplementation of additional glucose is marked with an arrow) and the addition of the hydrophobic carbon source for the fermentor scale production of sophorolipids on 1,12-dodecanediol. The amount of 1,12-dodedanol is not given in g/L since the volume of the culture broth decreased during fermentation because of evaporation and sampling. The ELSD signal form the HPLC-analysis is also plotted; all peaks originating from sophorolipids were added. Please notice that this method does not allow a correct quantification, but rather gives an indication of the sophorolipid production profile.

The pH-profile was as expected; once the value of 3.5 was reached no increase was observed until the end of the incubation period (Figure II-10). The $pO_2$ values were quite low despite a good aeration (1 vvm) and high stirring velocity (800 rpm); this is probably due to the deposition of sophorolipids on the probe, a fact which has been observed for some other

fermentations as well. However, one can still observe a small oscillating pattern, probably caused by the periodic addition of 1,12-dodedacanediol.

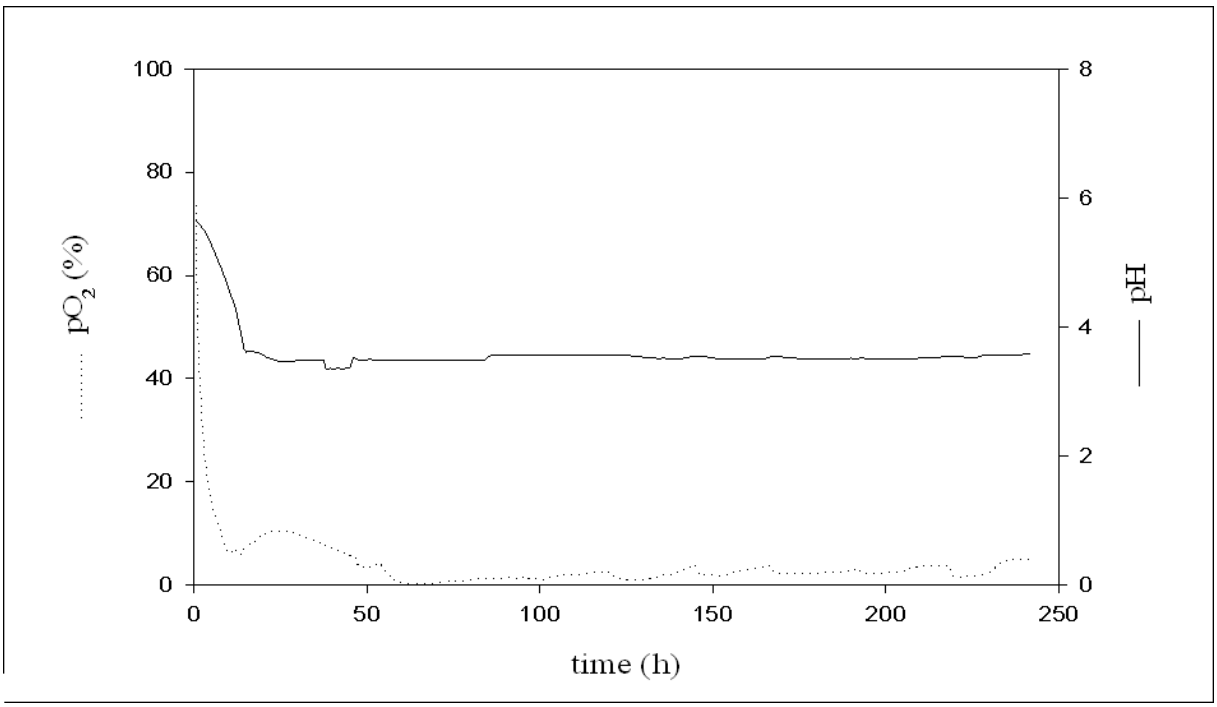


**Figure II-10**: pH and $pO_2$ profile during fermentor scale production of sophorolipids on 1,12-dodedacanediol.

Figure II-11 shows the measurements regarding cell growth. After 50 hours of fermentation, the values become more or less stabilized. The optical density and the cell dry weight still steadily increase, probably due to cell enlargement by intracellular formation of lipid bodies, as it can be seen that the values for the colony forming units do not follow the same trend. At the end of the fermentation, a CDW of 25 g/L is obtained; this CDW level and the other cell growth related values are comparable to those from similar fermentations on rapeseed oil. Therefore, it can be concluded that 1,12-dodecanediol is not toxic to *C. bombicola* cells and that it can be used in the production of medium-chain sophorolipids.

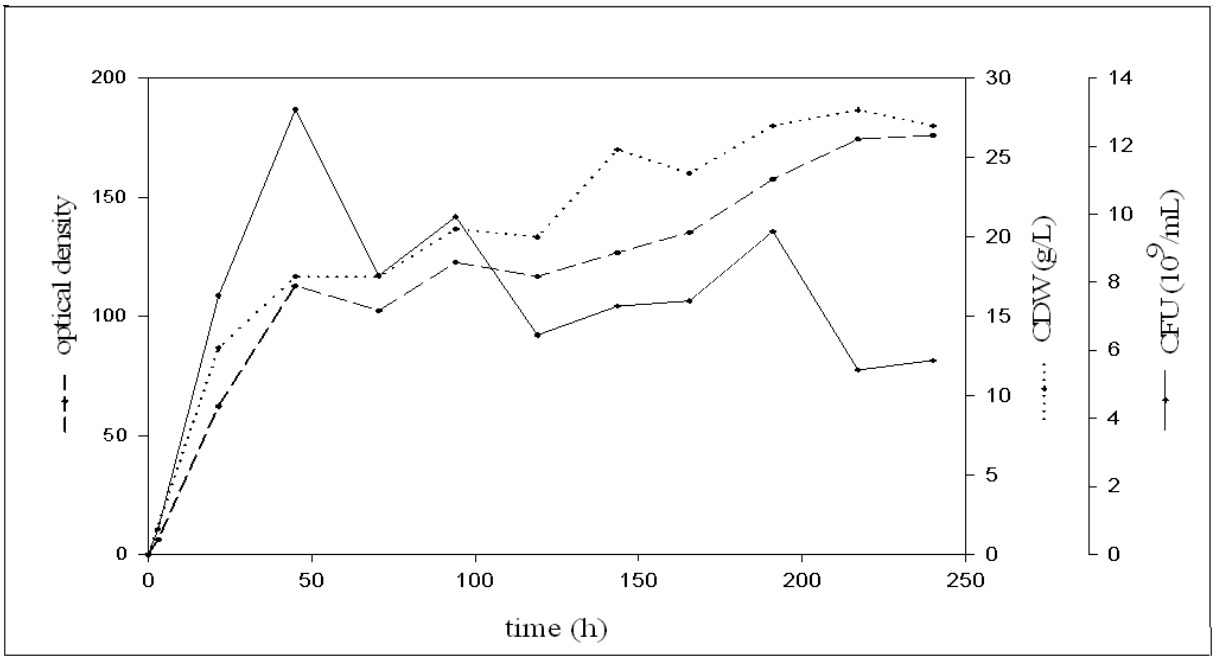


**Figure II-11:** Values for the CDW, CFU and OD of the *C. bombicola* cells during a fermentation on 1,12-dodecanediol.

## 3.3 Fermentations run on stearic acid-like substrates with an internal ester bond

### 3.3.1 Dodecyl glutarate

#### 3.3.1.1 Introduction

As stated in the general introduction of this chapter, *C. bombicola* has a certain preference for stearic acid above all other natural occurring fatty acids for the incorporation into sophorolipids. This fatty acid is readily hydroxylated by a cytochrome P450 monooxygenase and consequently covalently linked to a glucose molecule by a glucosyltransferase. The same is true for other saturated or unsaturated fatty acids with a chain length from 16 to 18 carbon atoms. The cytochrome P450 monooxygenase does not seem to have a strong specificity towards the shape of the substrate; its activity is very likely determined by the carbon chain length of the presented molecule.

The chemical structure of dodecyl glutarate resembles very much the one of stearic acid (Figure II-12), and for this reason it is believed that *C. bombicola* can incorporate this unusual substrate into its sophorolipid molecules. Dodecyl glutarate is of particular interest because it contains an internal ester bond. This bond can post-fermentative be subjected to alkaline hydrolysis, giving rise to medium-chain sophorolipids with a C12 fatty acid moiety, which will probably have other physico-chemical properties as compared to the native sophorolipids, such as improved water solubility and foam forming capacities.

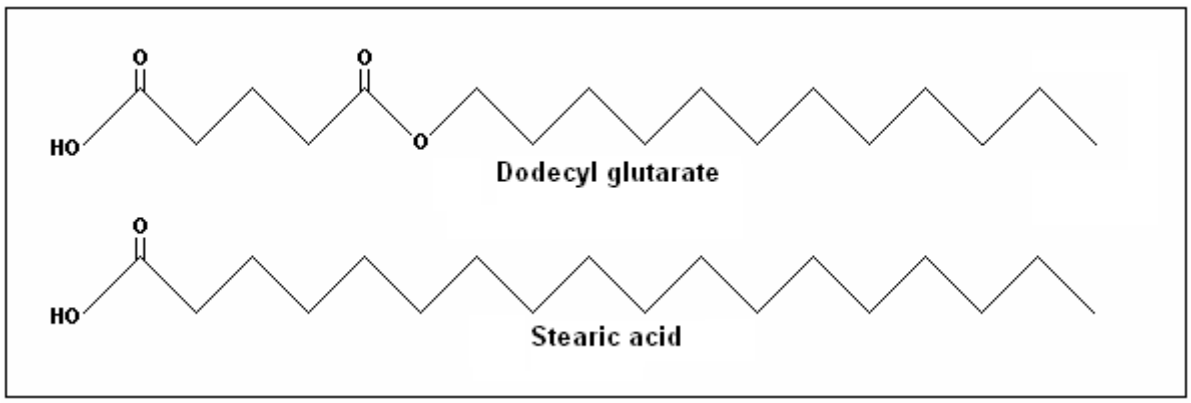


**Figure II-12**: Chemical structure of dodecyl glutarate compared with the structure of stearic acid.

#### 3.3.1.2 Shake flask experiment

In a first small scale experiment the capacity of *C. bombicola* to handle dodecyl glutarate as a hydrophobic substrate for sophorolipid production is tested. 48 hours after inoculation 10 g/L dodecyl glutarate is added. 3, 4 and 5 days after this first supplementation, 7.5, 5 and 7.5 g/L were added respectively (a total of 30 g/L). This gradual feeding is done because of the possible toxicity towards yeast cells and because of possible instability of the substrate at low pH and 30 °C during long periods of incubation. The CFU's indeed drop gradually once the feeding is started; the total difference between the start of the feeding and the end of the fermentation is 0.9 log units CFU/mL. The toxic effect could be caused by the substrate itself or by its degradation products 1-dodecanol and glutaric acid. Previous experiments with 1-dodecanol revealed that 5 g/L 1-dodecanol could cause a drop in the CFU number per mL with 0.6 log units in two days (this amount of dodecanol corresponds with 8.1 g/L dodecyl glutarate) and that 20 g/L, corresponding to 32.2 g/L dodecyl glutarate, causes a reduction of 2 log units after 8 days of incubation. Those findings illustrate the toxic effect of this potential degradation product. Also glutaric acid is thought to be toxic towards yeast cells. This hypothesis was tested by adding 0, 10 or 100 % of glutaric acid 48 hours after inoculation. 100 % of glutaric acid equals the amount that is released after the hypothetical total hydrolysis of 30 g/L dodecyl glutarate. It turned out that even high concentrations of glutaric acid had no adverse effect on the cell viability. The toxicity is consequently caused by the substrate itself, 1-dodecanol or the produced new-to-nature sophorolipids.

The substrate's melting temperature is higher than 30 °C, so it is present in the culture medium in the particulate solid state. This can make the uptake by the cells more complicated, especially in shake flasks. Probably partially for this reason, there was 13 g/L substrate left at the end of the fermentation period. Nevertheless, 26 g/L sophorolipids were formed. HPLC-analysis confirmed the incorporation of dodecyl glutarate by the presence of two new peaks in the hydrophilic area (Figure II-13). Their exact molecular structure remains unclear, but it is believed that the first peak is caused by the acidic dodecyl glutarate sophorolipids, while the second one originates from the lactonic form. A deeper structural analysis is given in Section 3.3.1.4, but GC-MS analysis already put forward that the large majority of the sophorolipids contain dodecyl glutarate as their fatty acid moiety.

The shake flask experiments illustrate that dodecyl glutarate can be incorporated into sophorolipids by *C. bombicola*. The incorporation of such substrates with an internal ester bond opens great perspectives for post-fermentative modifications. It was therefore decided to perform a fermentor scale synthesis of those special sophorolipids, in this way more parameters can be monitored.

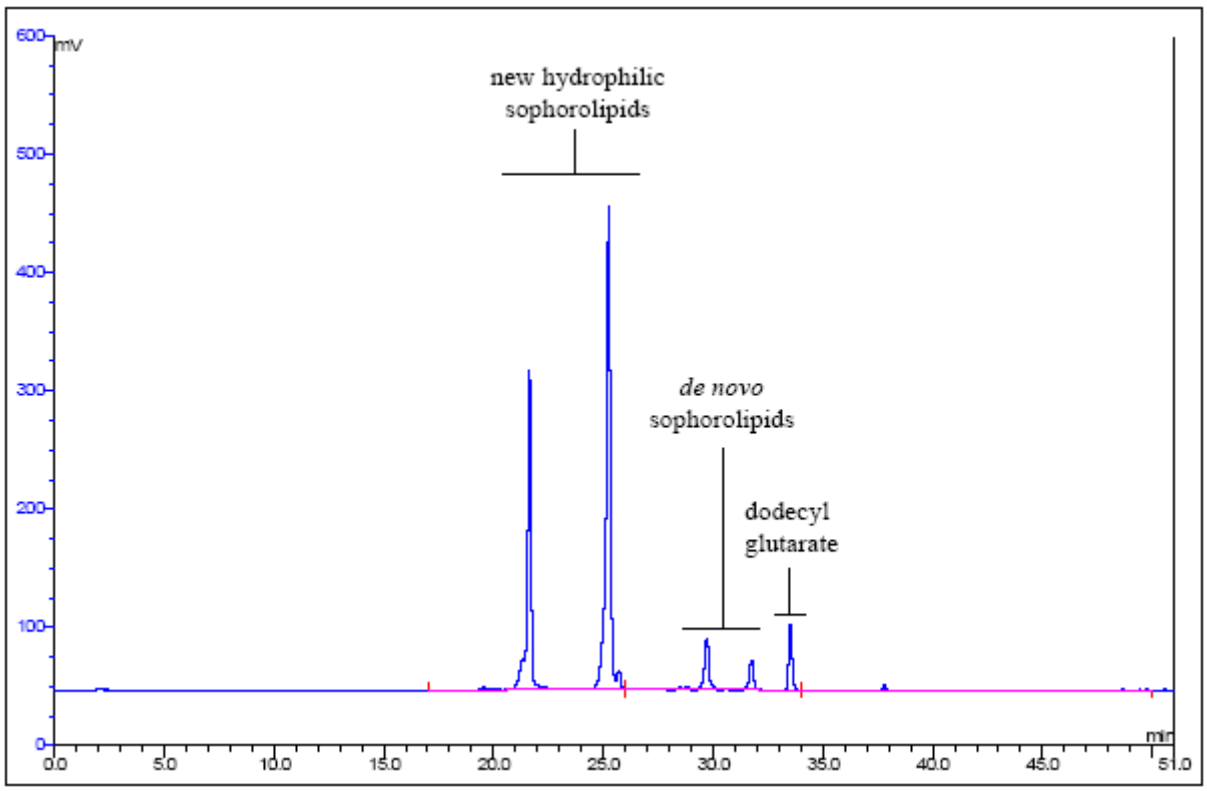


**Figure II-13:** HPLC-chromatogram of sophorolipids obtained from a shake flask experiment with dodecyl glutarate as hydrophobic carbon source. Sophorolipids were extracted with ethyl acetate and 2.5 % acetic acid (daily sampling) as described in Section 2.2.3 of this chapter.

### 3.3.1.3 Fermentor scale experiment

A fermentor scale experiment with 3 L working volume was set up in order to produce more sophorolipids and to able to monitor the synthesis of the unusual sophorolipids more in detail. Because of the bad solubility, slight toxicity and possible degradation of dodecyl glutarate, this hydrophilic carbon source was added gradually as shown in Figure II-14. Feeding was started after 48 hours and 15 to 25 g substrate was added daily, making up a total of 145.5 g.

The addition of substrate was stopped two days before end the fermentation to allow complete incorporation into sophorolipids. The initially available glucose was almost consumed after 144 hours; at this time an additional 50 g/L was supplemented.

As expected, sophorolipid production started in the exponential phase. During this phase, an increasing level of sophorolipids was detected. A total of 239.2 g sophorolipids or 79.7 g/L was recovered from the final fermentation medium by adding 145.5 g or 0.485 mol dodecyl glutarate. When making the simplification that all obtained sophorolipids have a dodecyl glutarate hydrophobic part, are in the lactonic form and have the two glucose molecules acetylated (MW: 706 g/L), a yield of 0.70 mol/mol is obtained. However, this is a slight overestimation since part of the sophorolipids is formed *de novo*, but it still is clearly much higher than the value obtained with 1,12-dodecanediol.

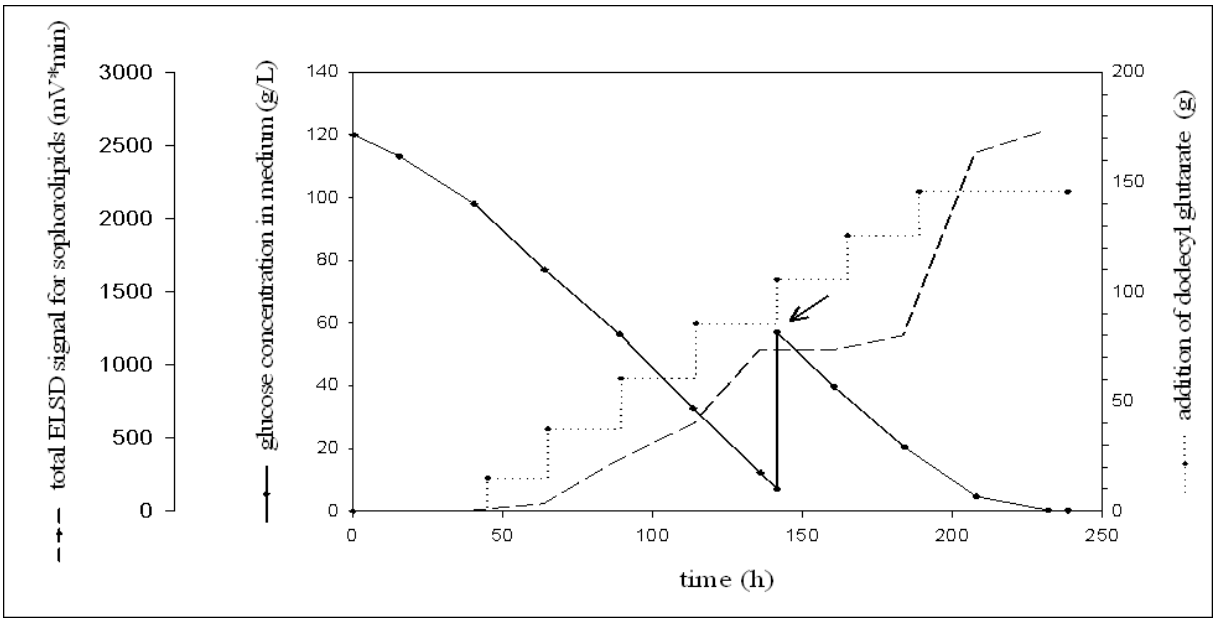


**Figure II-14:** Values for the glucose concentration in the medium (supplementation of additional glucose is marked with an arrow) and the addition of the hydrophobic carbon source for the fermentor scale production of sophorolipids on dodecyl glutarate. The amount of dodecyl glutarate is not given in g/L since the volume of the culture broth decreased during fermentation because of evaporation and sampling. The ELSD signal form the HPLC-analysis is also plotted; all peaks originating from sophorolipids were added up. Please notice that this method does not allow a correct quantification, but rather gives an indication of the sophorolipid production profile.

The pH could be maintained at 3.5, indicating normal fermentive behaviour (Figure II-15). $pO_2$ decreased strongly upon inoculation; after 20 hours it increased again to reach a more or less stable value after 30 hours. Each time, upon addition of dodecyl glutarate, the $pO_2$ value goes down, but after 24 hours it increases and stabilizes again, indicating that the substrate is consumed.

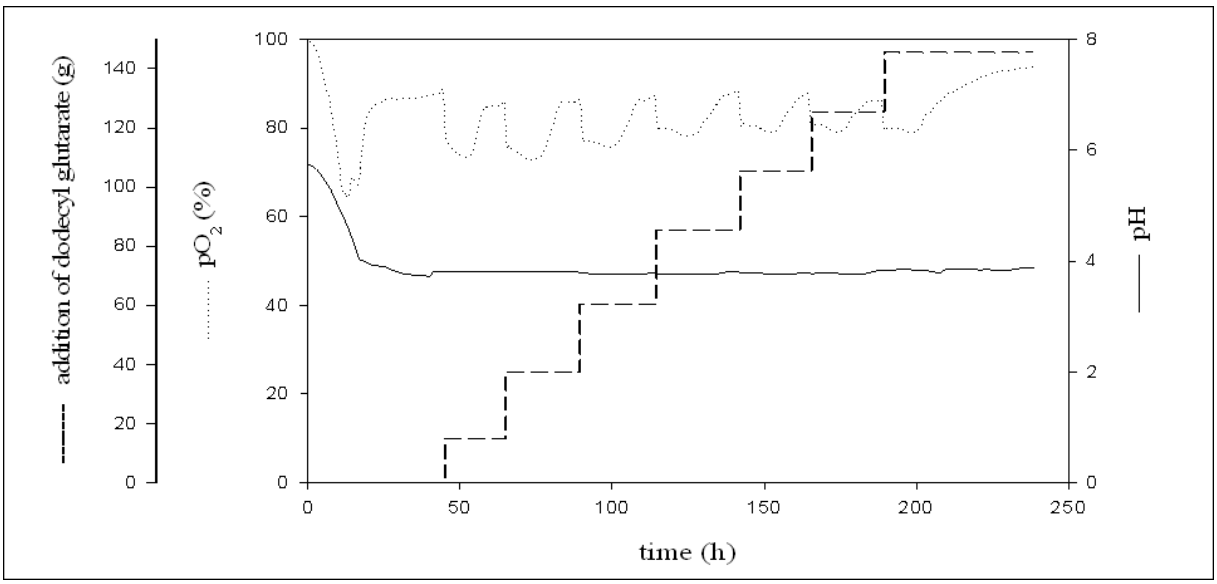


**Figure II-15**: pH and $pO_2$ profile during the fermentor scale production of sophorolipids on dodecyl glutarate. The addition of hydrophobic carbon source is also plotted to see its effect on $pO_2$.

After 50 hours of incubation, the values of CDW, CFU and optical density reach more or less a constant value, corresponding with the stationary phase (Figure II-16). Although the overall sophorolipid production (and glucose consumption) is normal, the values for the cell dry weight are quite low when compared to those obtained on fermentations with rapeseed oil. The maximum value here is 15.5 g/L, while over 23 g/L can be reached when rapeseed oil is used. This negative effect on growth extent is probable caused by the toxicity of dodecyl glutarate.

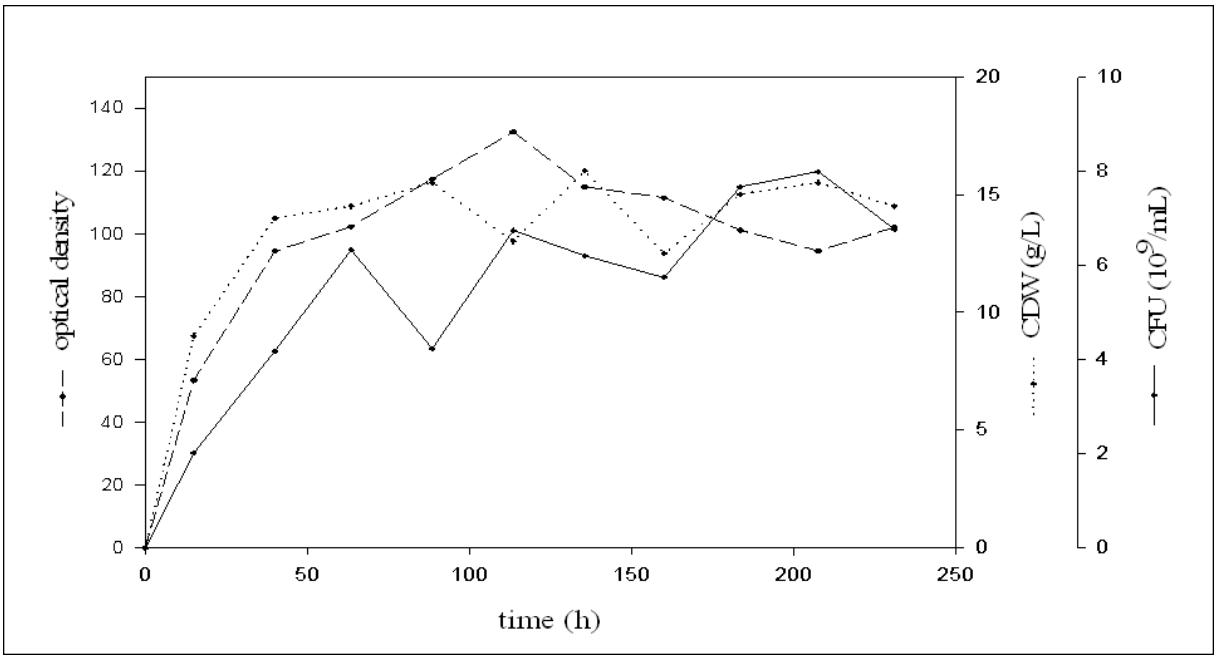


**Figure II-16:** Values for CDW, CFU and OD of the *C. bombicola* cells during fermentor scale production of sophorolipids on dodecyl glutarate.

### 3.3.1.4 Alkaline hydrolysis of the sophorolipids and analysis of the initial and hydrolysed products

The same two hydrophilic peaks detected in the shake flask experiment were also found in the fermentor scale experiment (Figure II-17). Whithin the same hydrophilic area, other minor peaks occurred as well. It is possible that those components were already present in the shake flask experiment, but could not be detected because of a too low concentration. To get certainty on the structure of the new components, an HPLC-MS analysis was conducted on the initial sophorolipid mixture. This confirmed that the first peak originated from a di-acetylated acidic dodecyl glutarate sophorolipid, while the second major peak relates to its lactonic form. The minor components were identified as mono-acetylated forms of the previous ones. One can notice that two peaks are present for the mono-acetylated form; this difference is nevertheless not thought to be caused by the position of the acetyl groups. When looking at alkaline hydrolysed sophorolipids, with a consequent lack of acetyl groups, one still can observe two peaks. Therefore it is believed that the difference in hydrophobicity is caused by the position of the fatty acid hydroxyl group: sophorolipids with subterminal hydroxylation are slightly more hydrophobic.

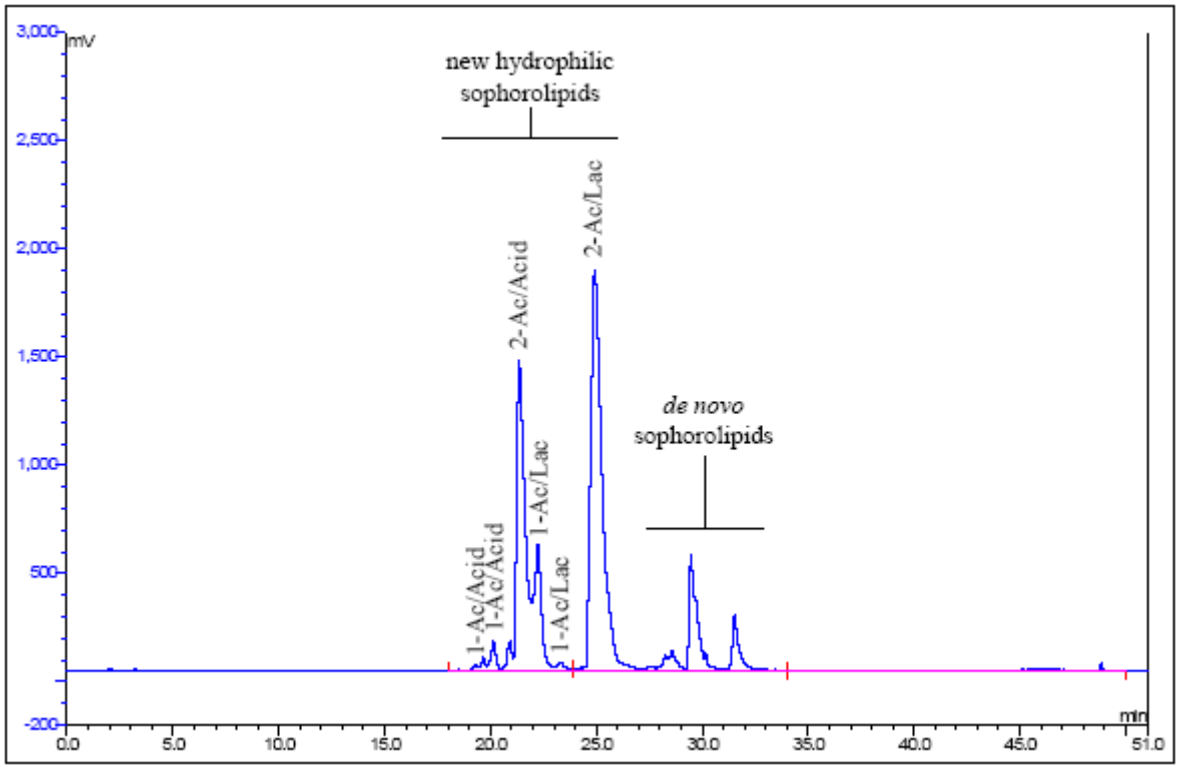


**Figure II-17:** HPLC-chromatogram of sophorolipids obtained from a fermentor scale fermentation with dodecyl glutarate as hydrophobic carbon source. Sophorolipids with dodecyl glutarate incorporation are named by the following principle: 1-Ac: mono-acetylated, 2-Ac: di-acetylated, Acid: acidic form and Lac: lactonic form.

Sophorolipid molecules are not stable in an environment with a pH higher than 7.5; the lactonic bond is subjected to irreversible hydrolysis and acetyl molecules are removed from the glucose units. Increased pH will also cause hydrolysis of the internal ester bond of the dodecyl glutarate fatty acid part, giving rise to a C12 sophorolipid. Here, the hydrophobic moiety will not be a fatty acid, but a fatty alcohol. Figure II-18 shows that the complexity of the peak pattern of the new dodecyl glutarate sophorolipids is reduced by the removal of lactonic bonds and acetyl groups, leading to one major component: a non-acetylated acidic dodecanol sophorolipid; its structure was verified by HPLC-MS analysis (0-Ac/Acid C12oh-2G). This glycolipid is quite hydrophilic because of the short fatty alcohol part. In the same hydrophilic area, a number of very small but unexpected peaks were detected. The MS results could only be explained by the fact that those components were also non-acetylated acidic dodecanol glycolipids, but with a different number of glucose units. This hypothesis is supported by their elution order: additional glucose units render the molecule more hydrophilic. The incorporation of additional glucose molecules can only take place if the

dodecyl glutarate tail is already hydrolysed; glucose can not be coupled to the carboxylic end of the glutarate by *C. bombicola*. It is indeed possible that some minor degradation of the substrate occurred, either before or after incorporation into sophorolipid molecules, in this way giving rise to a free hydroxyl group which can be used by glucosyltransferase I.

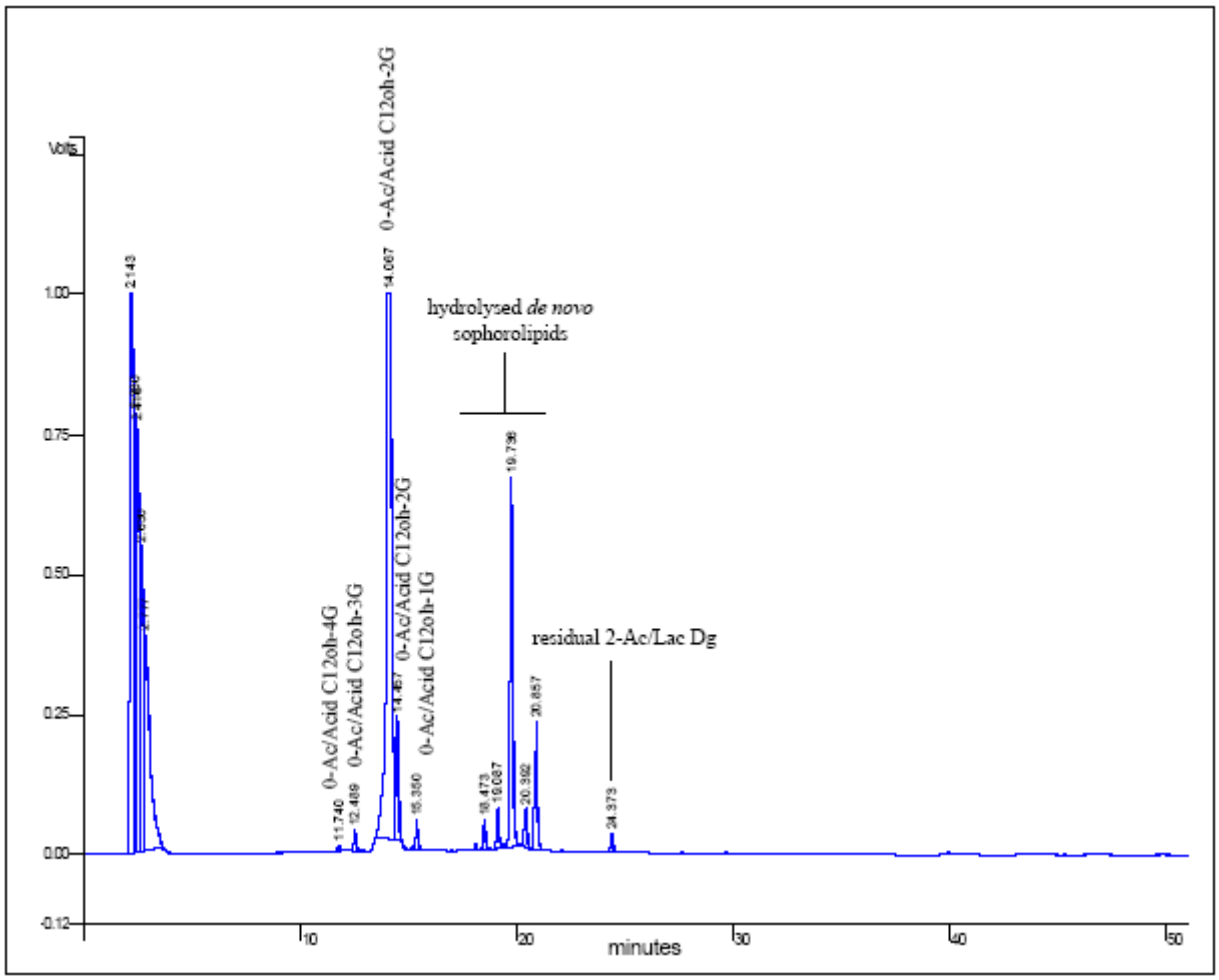


**Figure II-18:** HPLC-chromatogram of alkaline hydrolysed dodecyl glutarate sophorolipids. Sophorolipids with dodecyl glutarate incorporation are named by the following principle: 2-Ac: di-acetylated, 0-Ac: non-acetylated, Acid: acidic form, Lac: lactonic form, Dg: dodecyl glutarate fatty acid tail, C12oh: dodecanol fatty acid tail, 1G: one glucose molecule, 2G: two glucose molecules, 3G: three glucose molecules and 4G: four glucose molecules.

The occurrence of molecules with only one glucose molecule can be explained by a different affinity of the glucosyltransferase II enzyme towards the unconventional dodecyl glutarate substrate. This could also give an explanation for the glycolipids with three glucose units. On the other hand, those molecules could be the result of degradation due to the high pH, but the

absence of this phenomenon for the native hydrolyzed sophorolipids, makes this hypothesis less plausible.

The yeast *C. bombicola* is capable of incorporating the unusual substrate dodecyl glutarate into its sophorolipid molecules thanks to the molecule's structural similarity with stearic acid. Dodecyl glutarate is incorporated at high efficiencies, but because of its toxicity, controlled feeding is required. Substrates with an internal ester bond are of particular interest because the resulting sophorolipids can be subjected to post-fermentative modifications yielding medium-chain sophorolipids. The above experiments with dodecyl glutarate prove that this strategy works and for this reason other substrates with a different position of the ester bond were tested as described in the following sections.

## 3.3.2 Dodecyl and tetradecyl malonate

### 3.3.2.1 Introduction

The substrate "dodecyl and tetradecyl malonate" (Figure II-19) is in fact a mixture of dodecyl malonate and tetradecyl malonate, with respectively 69 and 31 % of both components as determined by GC-MS analysis on the methylated substrate. The substrate dodecyl and tetradecyl malonate will from hereonwards be referred to as "malonate ester".

The molecules are quite similar to dodecyl glutarate and are therefore believed to be readily incorporated into sophorolipid molecules by *C. bombicola*. Tetradecyl malonate has the same chain length of stearic acid, while dodecyl malonate has a chain length comparable to palmitic acid. Nevertheless, as palmitic acid is also incorporated into sophorolipid molecules, it is expected that the same is true for dodecyl malonate.

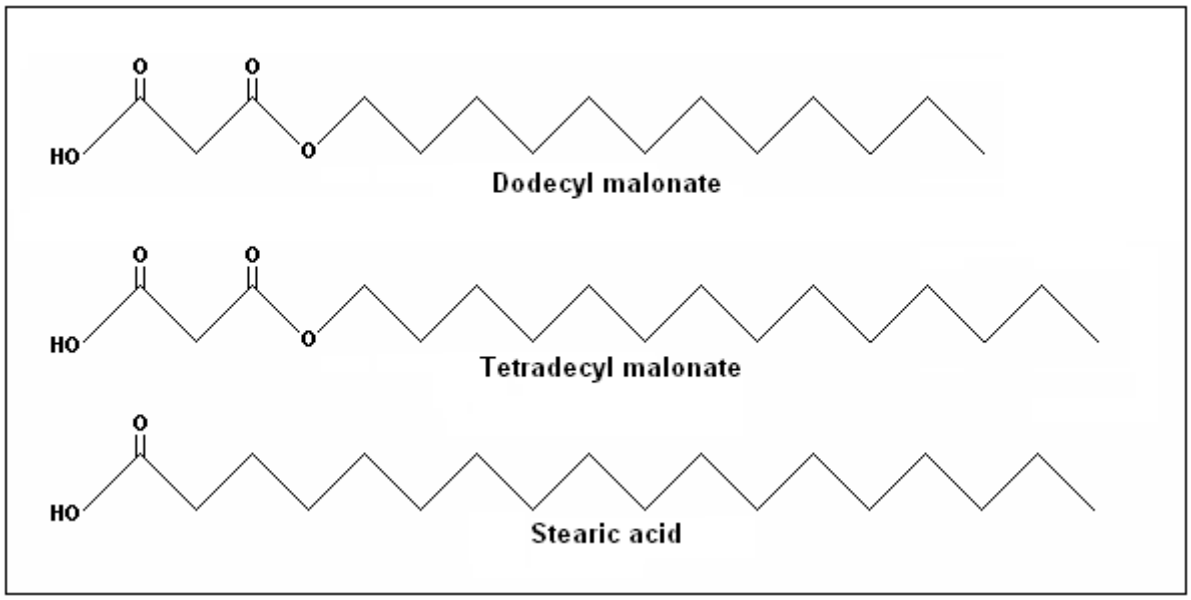


**Figure II-19**: Chemical structure dodecyl malonate and tetradecyl malonate compared with the structure of stearic acid.

### 3.3.2.2 Fermentation process

The use of malonate ester as a hydrophobic carbon source was tested in a fermentor scale experiment. A working volume of 1.5 L was used. The fermentation had to be stopped after

166 hours because of strong foam formation. Foaming was induced upon addition of malonate ester and for this reason the stirring velocity was temporally lowered to 700 rpm after 115 hours. During the incubation time of 166 hours, a total amount of 49 mL substrate was added (Figure II-20), corresponding with 48.6 g ($\rho$= 0.992 g/mL) and after 161 hours, an additional amount of glucose was added.

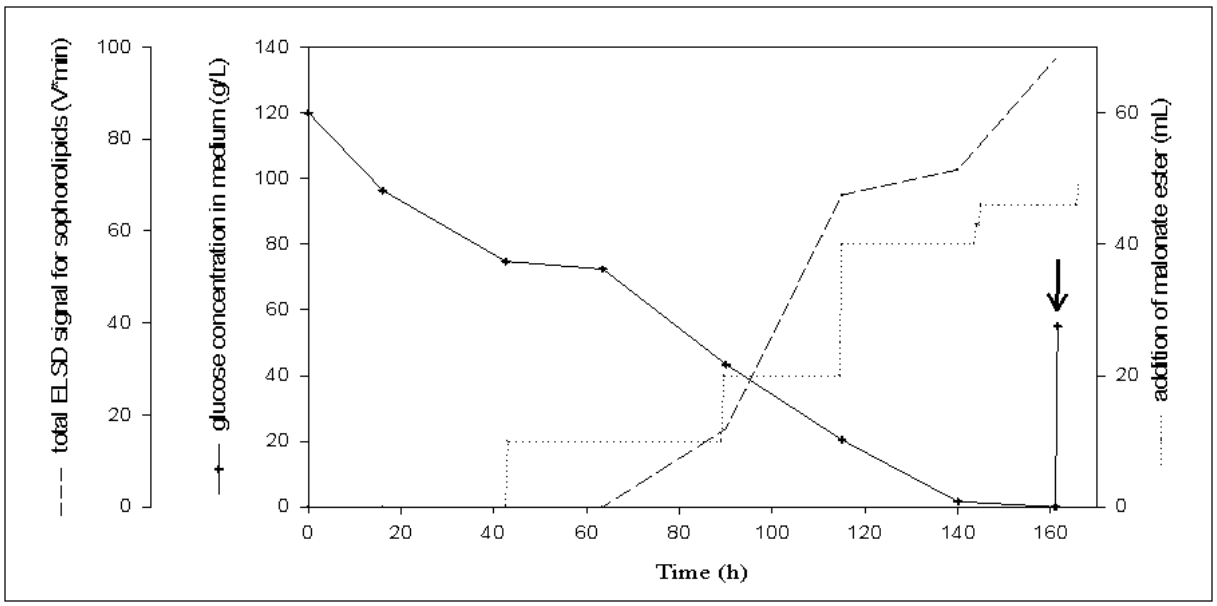


**Figure II-20:** Values for the glucose concentration in the medium (supplementation of additional glucose is marked with an arrow) and the addition of the hydrophobic carbon source for the fermentor scale production of sophorolipids on malonate ester. The amount of malonate ester is not given in mL/L since the volume of the culture broth decreased during fermentation because of evaporation and sampling. The ELSD signal form the HPLC-analysis is also plotted; all peaks originating from sophorolipids were added up. Please notice that this method does not allow a correct quantification, but rather gives an indication of the sophorolipid production profile.

The feeding of malonate ester has an effect on the $pO_2$ values; after each supplementation $pO_2$ decreases and then slowly increased again (Figure II-21).

The formation profile of biomass is more or less the same as for the fermentation with dodecyl glutarate, a maximal CDW value of 17 g/L was reached (data not shown).

Although the fermentation lasted only 166 hours, 55.4 g/L sophorolipids were formed by adding 48.6 g or 0.173 mol substrate (MW: 280.7 g/mol). The amount of sophorolipids for 1.5 L corresponds to 0.120 mol when the assumption is made that all obtained sophorolipids have

a malonate ester from, are in the lactonic form and have the two glucose molecules acetylated (MW: 686.7 g/mol), resulting a yield of 0.70 mol sophorolipids per mol substrate. However, this stays an overestimated ratio since the *de novo* formed sophorolipids are not taken into account.

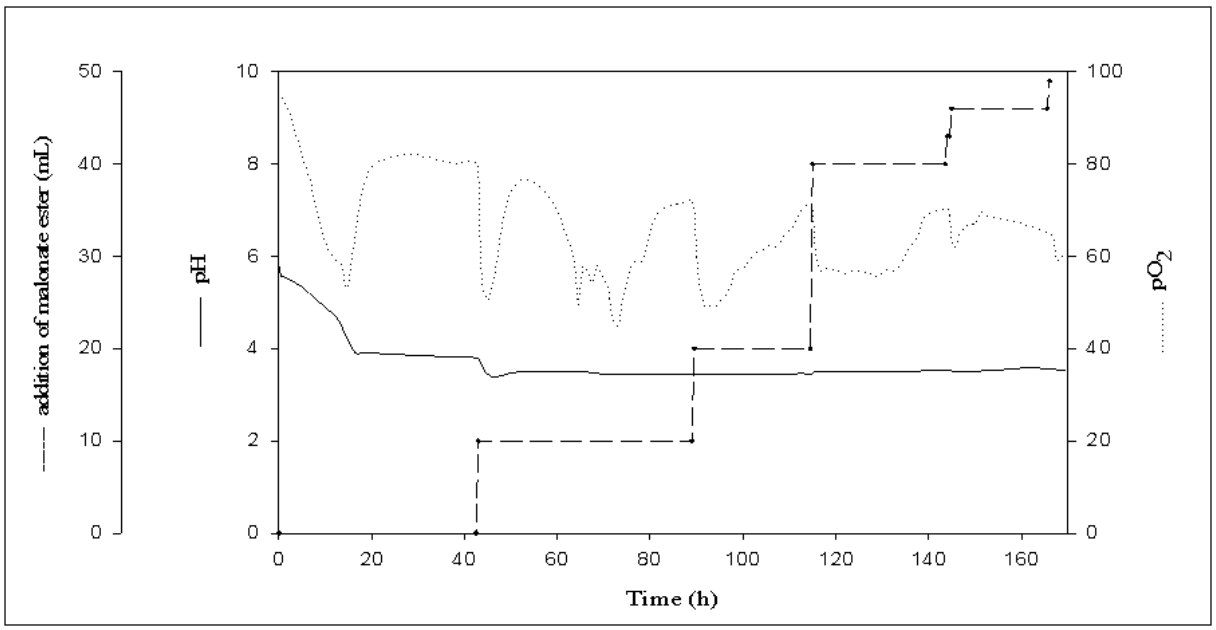


**Figure II-21**: pH and $pO_2$ (%) profile during the production of sophorolipids on malonate ester. The addition of hydrophobic carbon source is also plotted to see its effect on $pO_2$.

### 3.3.2.3 Analysis of the initial and hydrolysed malonate ester sophorolipids

There is a discrepancy between the HLPC results of the sophorolipids extracted with ethyl acetate (daily sampling) and those extracted with ethanol and water (end extraction). This phenomenon was also observed for the 1,12-dodecandiol sophorolipids (Section 3.2.2.1), and has its consequences for the more hydrophilic sophorolipids eluting of the HPLC column the first 20 min (Figure II-22); those compounds are not extracted by ethyl acetate. The values of the daily taken sophorolipid samples as shown in Figure II-20 are therefore an underestimation of the real amount of synthesized sophorolipids.

Figure II-22B indicates that a cluster of new peaks arise in the hydrophilic area. The pattern is quite complicated due to the substrate, which is in fact a mixture of the two components dodecyl malonate and tetradecyl malonate. The substrates themselves are eluted from the

column after 31.1 and 35.7 minutes and do not interfere with the sophorolipid analysis. The two major peaks with a retention time of about 17.7 and 19.3 minutes are believed to originate from the di-actylated acidic dodecyl malonate and tetradecyl malonate sophorolipids respectively. The other smaller peaks in the same area are probable caused by differences in the acetylation and hydroxylation pattern. In contrast to the fermentation run on docecyl glutarate, the major components here are acidic ones. This can be explained by the fact that the fermentation process was preliminary stopped; as also observed in other fermentations, lactonization of the sophorolipid molecules does not tend to occur rapidly at the onset of the

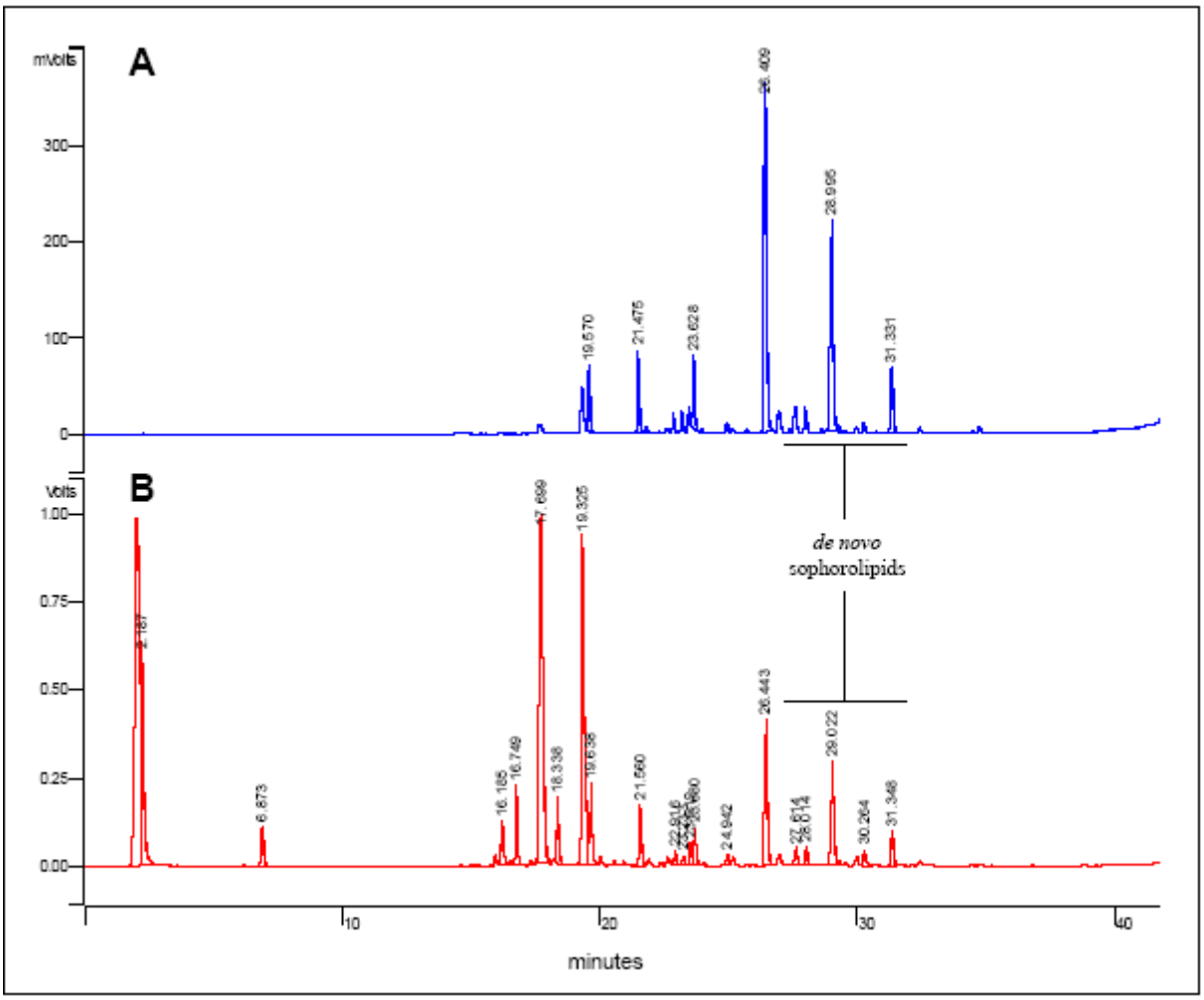


**Figure II-22:** HPLC-chromatogram of fermentor scale produced sophorolipids with malonate ester as hydrophobic carbon source. **(A)** Sophorolipids extracted with ethyl acetate and 2.5 % acetic acid (daily sampling) as described in Section 2.2.3 of this chapter. **(B)** Sophorolipids extracted with water and ethanol (end extraction) as described in Section 2.3 of this chapter.

production period. On the other hand, another interesting peak is the one appearing after 26.4 minutes; and it is believed that this peak corresponds to the lactonic di-acetylated tetradecyl malonate sophorolipid.

As expected, by alkaline hydrolysis of the sophorolipid mixture, non-acetylated acidic dodecanol and tetradecanol sophorolipids were detected as the major components in the hydrolysed mixture (Figure II-23, 0-Ac/Acid C12oh-2G and 0-Ac/Acid C14oh-2G). Similar to the hydrolysed dodecyl glutarate sophorolipids, MS analysis identified a number of small

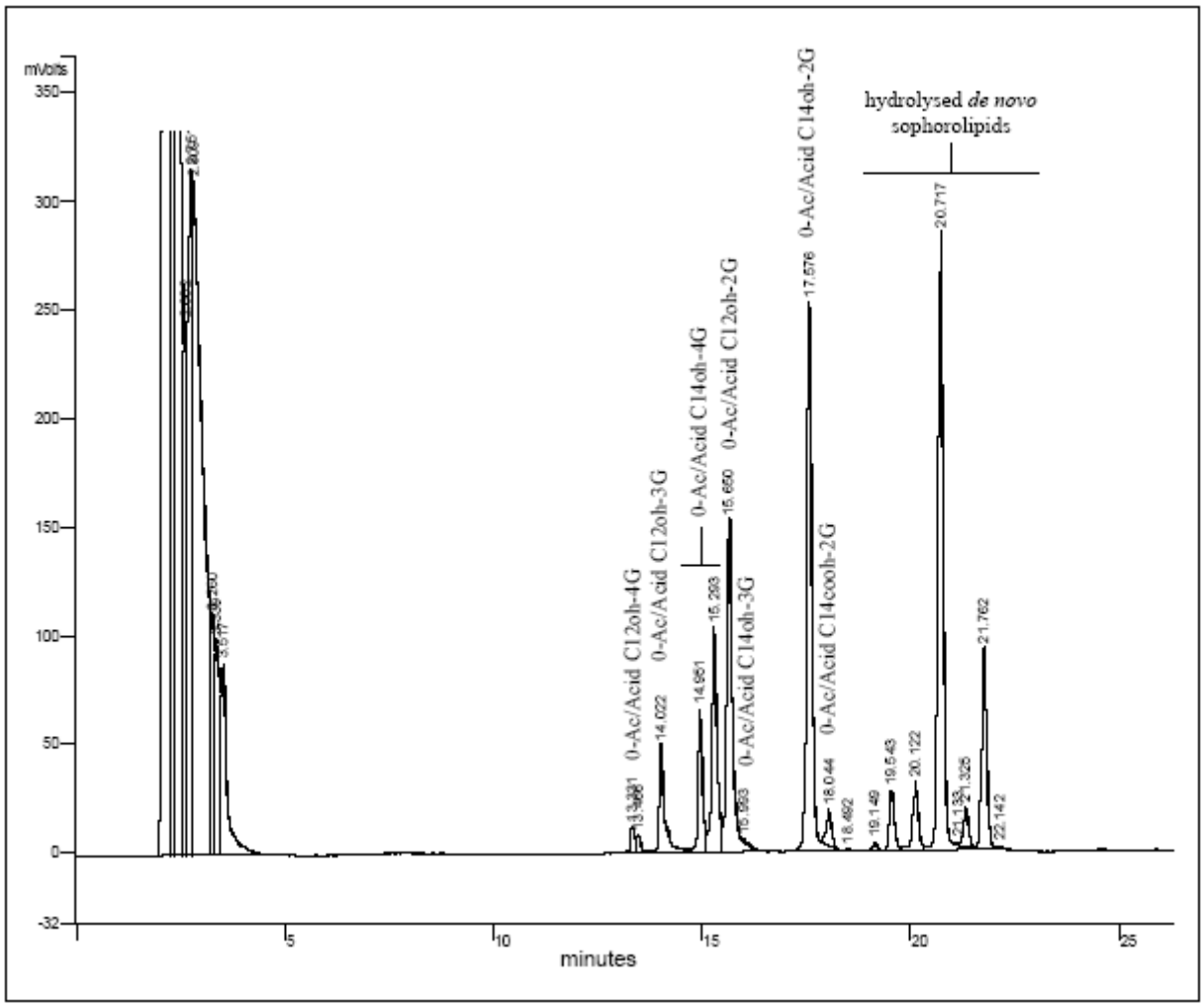


**Figure II-23:** HPLC-chromatogram of alkaline hydrolysed malonate ester sophorolipids. Sophorolipids with malonate ester incorporation are named by the following principle: 0-Ac: non-acetylated, Acid: acidic form, Lac: lactonic form, C12oh: dodecanol lipid tail, C14oh tetradecanol lipid tail, C14cooh: tetradecanoic acid lipid tail, 2G: two glucose molecules, 3G: three glucose molecules and 4G: four glucose molecules.

peaks caused by glycolipids with 4 or 3 glucose units. Furthermore, a small peak originating from a C14 sophorolipid with a carboxylic end was detected (0-Ac/Acid C14cooh-2G). This indicates the incorporation of the malonate ester degradation product tetradecanol after further oxidation by the yeast cells till hydroxy-tetradecanoic acid.
Another interesting trend is better representation of tetradecyl malonate derived sophorolipids. As possible explanation is the fact that the molecular structure of myristyl malonate closely resembles those of stearic acid, while dodecyl malonate is two carbon atoms shorter.

It can be concluded that malonate ester just like dodecyl glutarate gives rise to medium-chain sophorolipids after alkaline hydrolysis. The obtained mixture is yet more complex due to the characteristics of the substrate, but it also contains a non-acetylated acidic dodecanol sophorolipid and the estimated yields are more or less the same. One drawback related to this substrate is its foam forming capacity. This makes fermentations run with a high feeding rate a bit more challenging and difficult to control.

### 3.3.3 Pentenyl dodecanoate

#### 3.3.3.1 Introduction

In contrast to the previously described substrates with an internal ester bond, pentenyl dodecanoate (Figure II-24) does not posses a carboxylic end and consequently shows less structural resemblance with stearic acid. Nevertheless, the substrate is structurally similar to an alkane, a substrate which can also be used for sophorolipid production.

An important implication of this lack of carboxylic end is the substrate's loss of directionality for incorporation into the sophorolipid molecule; hydroxylation prior to the incorporation can take place at both ends of the molecule.

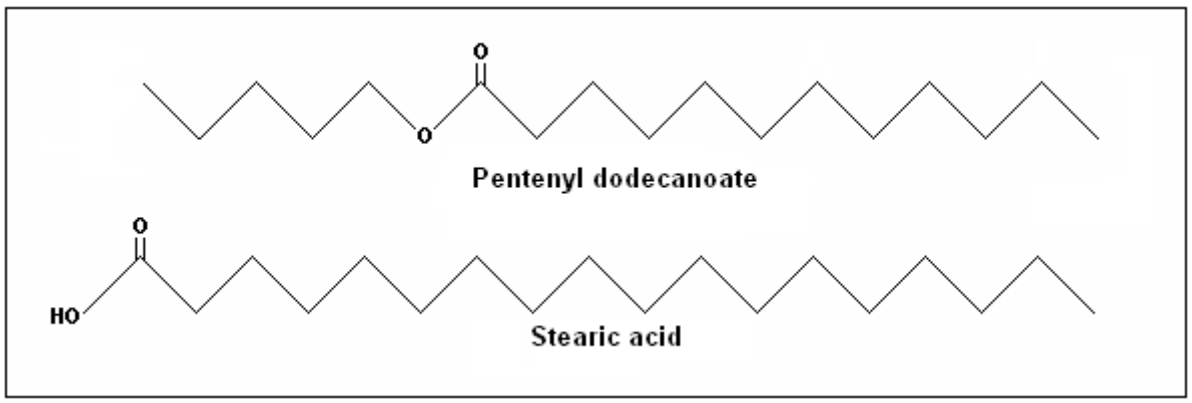


**Figure II-24**: Chemical structure of pentenyl dodecanoate compared with the structure of stearic acid.

#### 3.3.3.2 Fermentation process

The fermentation run was performed at 1.5 litre scale. Feeding of the liquid pentenyl dodecanoate was started 48 hours after inoculation (Figure II-25). Since the glucose consumption rate remained high (glucose was three times additional supplemented) and since the pH did not increase (which both are indications for metabolically active cells), the fermentation period was extended to 464 hours. Figure II-26 indeed confirms good cell viability: the values for CFU remain constant even after more than 200 hours. It can also be seen that during the second half of the fermentation process, CDW and optical density values slightly increase. The CDW even reaches a maximal value of 21 g/L, which is quite high for a fermentation run on an unconventional substrate.

$pO_2$ values could not be registered due to the deposition of sophorolipids on the probe.

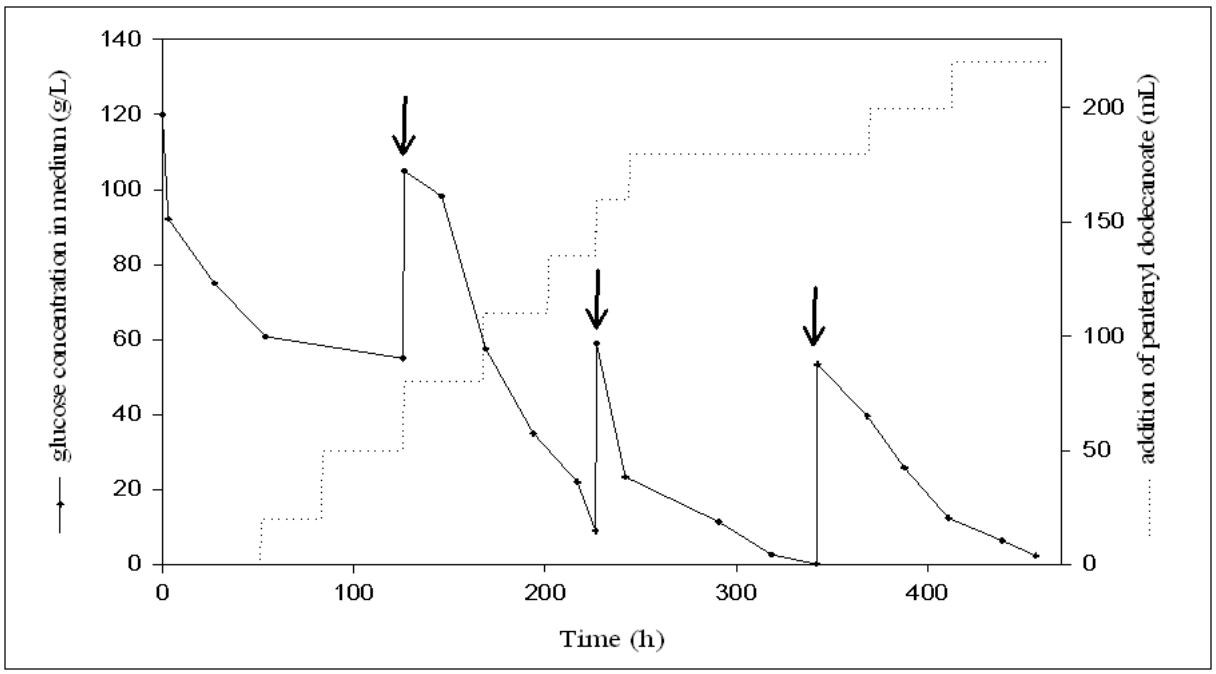


**Figure II-25:** Values for the glucose concentration in the medium (supplementations of additional glucose are marked with arrows) and the addition of the hydrophobic carbon source for the production of sophorolipids on pentenyl dodecanoate. The amount of pentenyl dodecanoate is not given in mL/L since the volume of the culture broth decreased during fermentation because of evaporation and sampling.

During this long fermentation period of 464 hours, a total amount 220 mL or 200 g ester ($\rho$=0.9101 g/mL) was added by controlled feeding. Despite the high amount of supplemented hydrophobic carbon source, only 124 g sophorolipids were recovered. This means that 0.740 mol substrate (MW 270 g/mol) only yields 0.183 mol sophorolipids, if assuming that all obtained sophorolipids derived from pentenyl dodecanoate and are di-acetylated, lactonic molecules (MW: 676 g/mol). This yield of 0.25 mol/mol is however an overestimation since the obtained 124 g sophorolipids also include the *de novo* formed molecules.

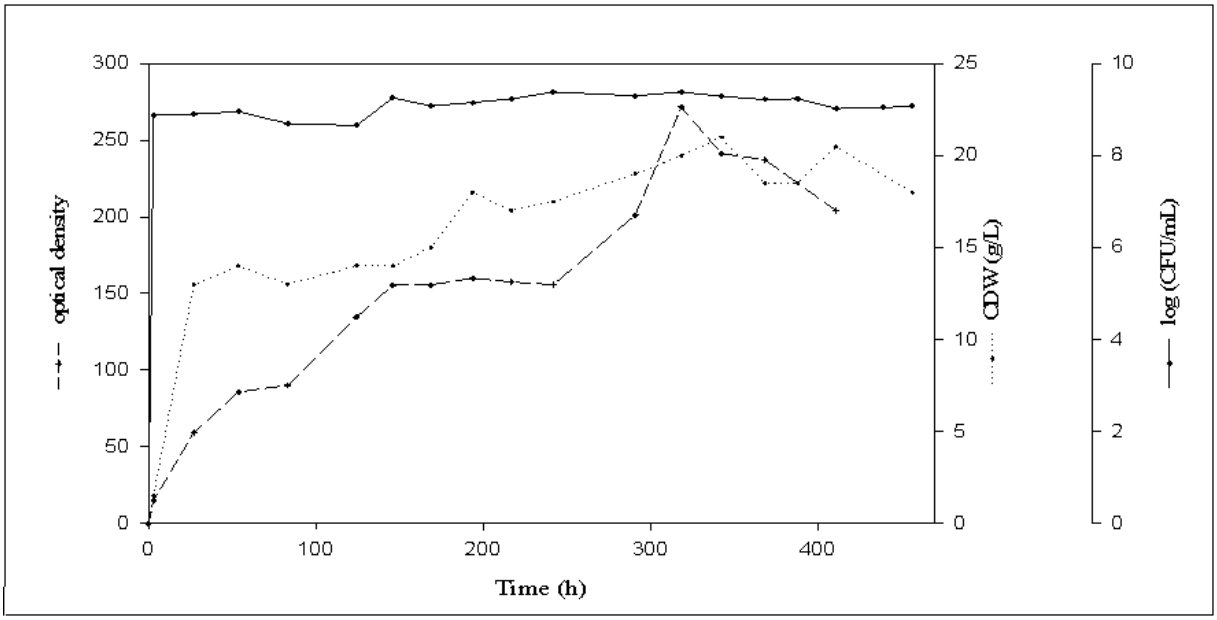


**Figure II-26:** Values for CDW, CFU and OD of the *C. bombicola* cells during the production of sophorolipids on pentenyl dodecanoate.

### 3.3.3.3 Analysis of the initial and hydrolysed pentenyl dodecanoate sophorolipids

The relative amount of hydrophilic sophorolipids remained quite small (Figure II-27). Only at the very last sampling point, the hydrophilic value strongly increased, while the one for the hydrophobic sophorolipids decreased. When looking at the HPLC-chromatograms of the two last samples, it can be seen that a clear change occurs in the hydrophilic peak pattern: new peaks arise in the very hydrophilic area (Figure II-28).

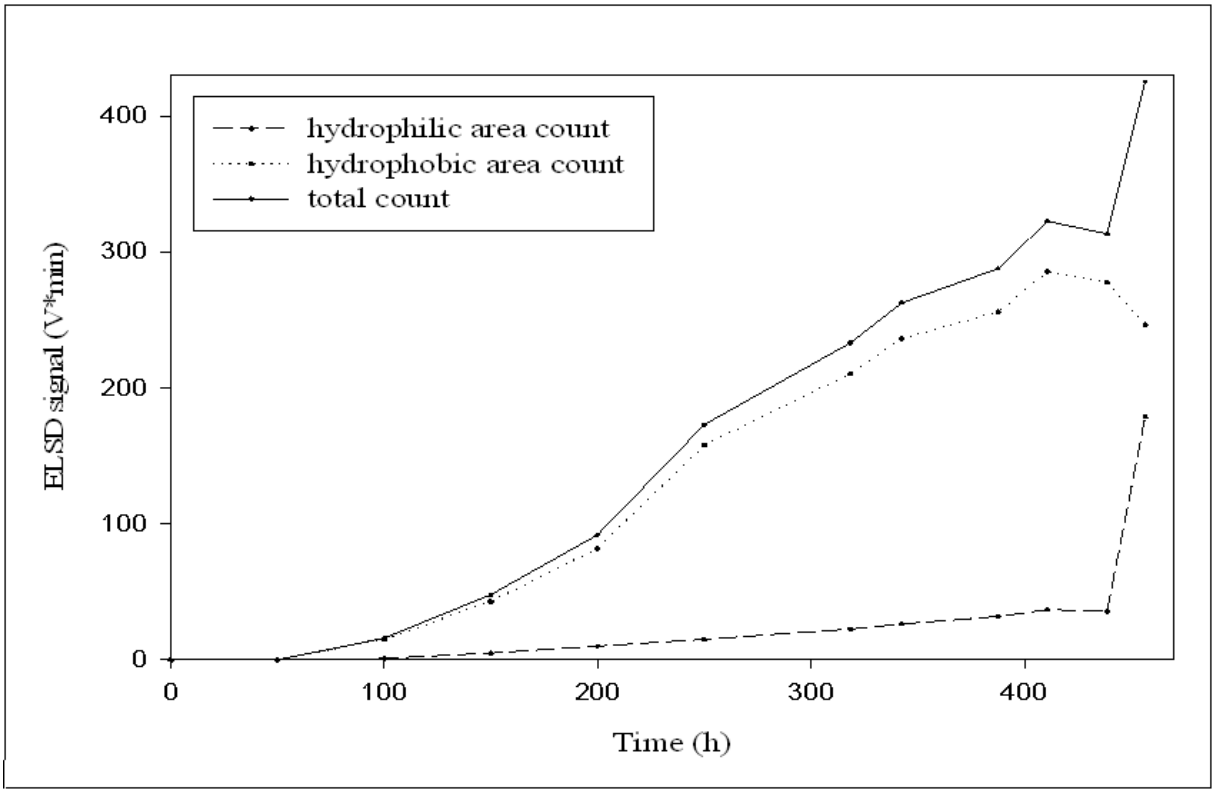


**Figure II-27:** Signal for sophorolipids analysed by HPLC and ELSD. All peaks from the hydrophilic, hydrophobic or total area were added up. Please notice that this method does not allow a correct quantification, but rather gives an indication of the sophorolipid production profile.

The overall profile of this last sample is identical to the one for the end-extracted sophorolipids. It is quite likely that hydrolysis or degradation took place in the last stage of the fermentation period. LC-MS analysis indeed confirmed that the very hydrophilic peaks originated from glycolipids, which only contained the pentenyl part of the ester (C5 sophorolipids). Upon degradation of the substrate, pentanol and dodecanoic acid are formed. As dodecanoic acid is a too short molecule for efficient hydroxylation by *C. bombicola* cytochrome P450 monooxygenase enzymes, it will not be incorporated into the sophorolipid molecules (see also Section 3.1.2 and 3.2.1 from this chapter). Nevertheless, pentanol possesses a terminal hydroxyl group which makes it a potential substrate for incorporation. Furthermore, a molecule with a hydroxylated C5-tail instead of an alkyl is detected at retention time 10.2 (2-AC/Acid C5oh). It is believed that those molecules originate from sophorolipids which have incorporate pentenyl dodecanoate by the pentenyl side and which are than submitted to hydrolysis of the internal ester bond.

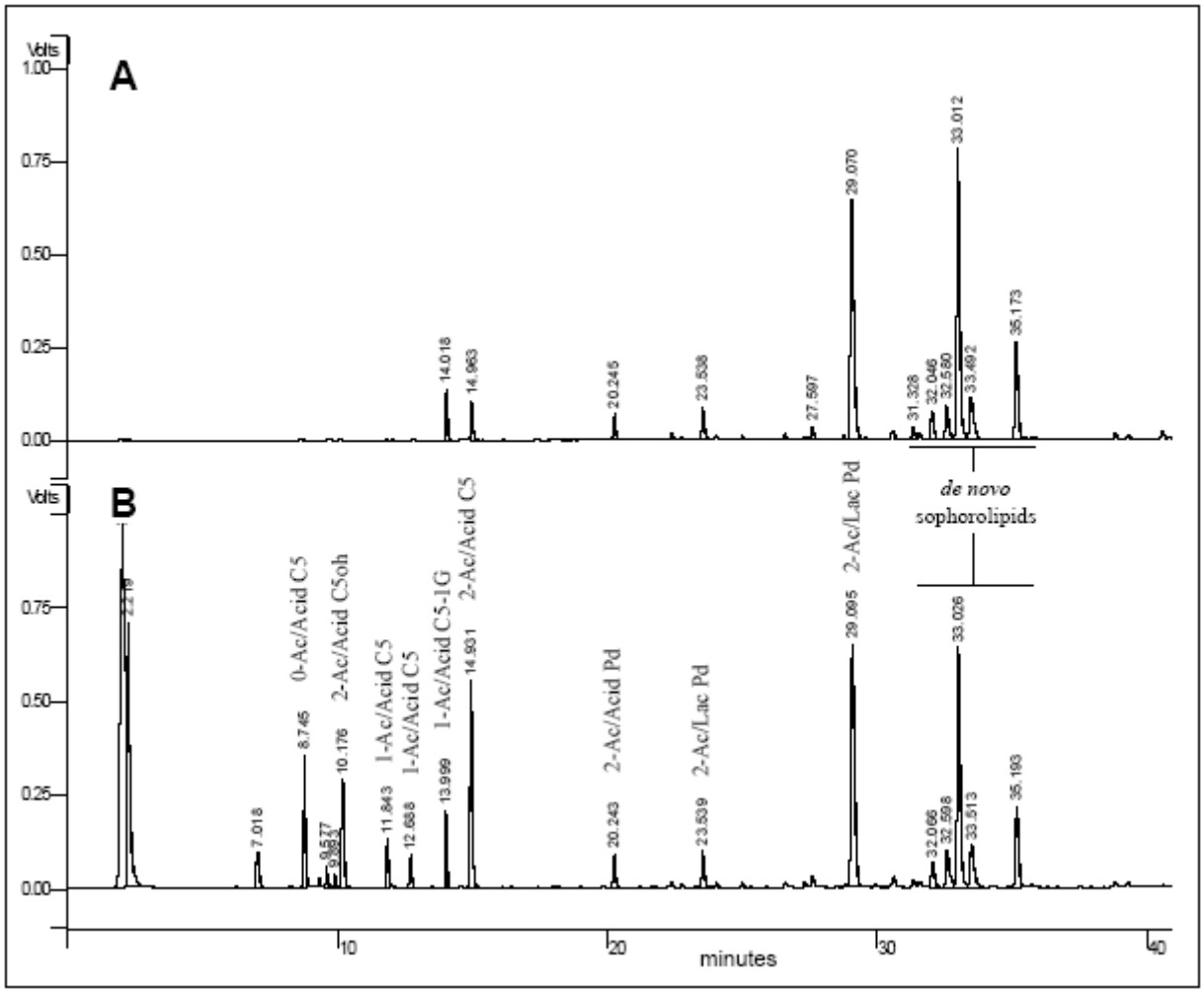


**Figure II-28:** Comparison between **(A)** the last and **(B)** second last sophorolipid sample. Sophorolipids with pentenyl dodecanoate incorporation are named by the following principle: 2-Ac: di-acetylated, 1-Ac: mono-acetylated, 0-Ac: non-acetylated, Acid: acidic form, Lac: lactonic form, Pd: pentenyl glutarate lipid tail, C5: pentenyl lipid tail, C5oh: pentanol lipid tail, 1G: only one glucose molecule.

The three new peaks which elute of the HPLC column later than 20 minutes all comprise the complete pentenyl dodecanoate substrate (Pd sophorolipids). Some of them occur in the lactonized form, which means that both substrate ends must be hydroxylated, while one of them is further oxidized till a carboxyl group. LC-MS analysis could not give further information about the exact orientation of the pentenyl dodecanoate.

Post-fermentative hydrolysis of the sophorolipid mixture resulted in the formation of sophorolipids with a carboxylated C12 fatty acid tail (Figure II-29, 0-Ac/Acid C12cooh), while for the previous substrates dodecyl glutarate and malonate ester, sophorolipids with a

hydroxylated fatty acid tail were formed due to the different orientation of the internal ester bond. The carboxylated molecules are formed upon hydrolysis of sophorolipids which have incorporated pentenyl dodecanoate by the dodecanoate side. The other orientation of the substrate into the sophorolipid molecule gives rise to non-acetylated C5 tailed sophorolipid with a hydroxyl end upon hydrolysis (0-Ac/Acid C5oh). This peak was clearly detected by LC-MS, but ELSD detection turned out to be less suitable for this molecule. Furthermore, non-acetylated C5 tailed sophorolipid with an alkyl end were detected (0-Ac/Acid C5), originating from the sophorolipids which incorporated the degradation product pentanol as described before. Finally, also the molecules with just one glucose molecule were found (0-Ac/Acid C5-1G).

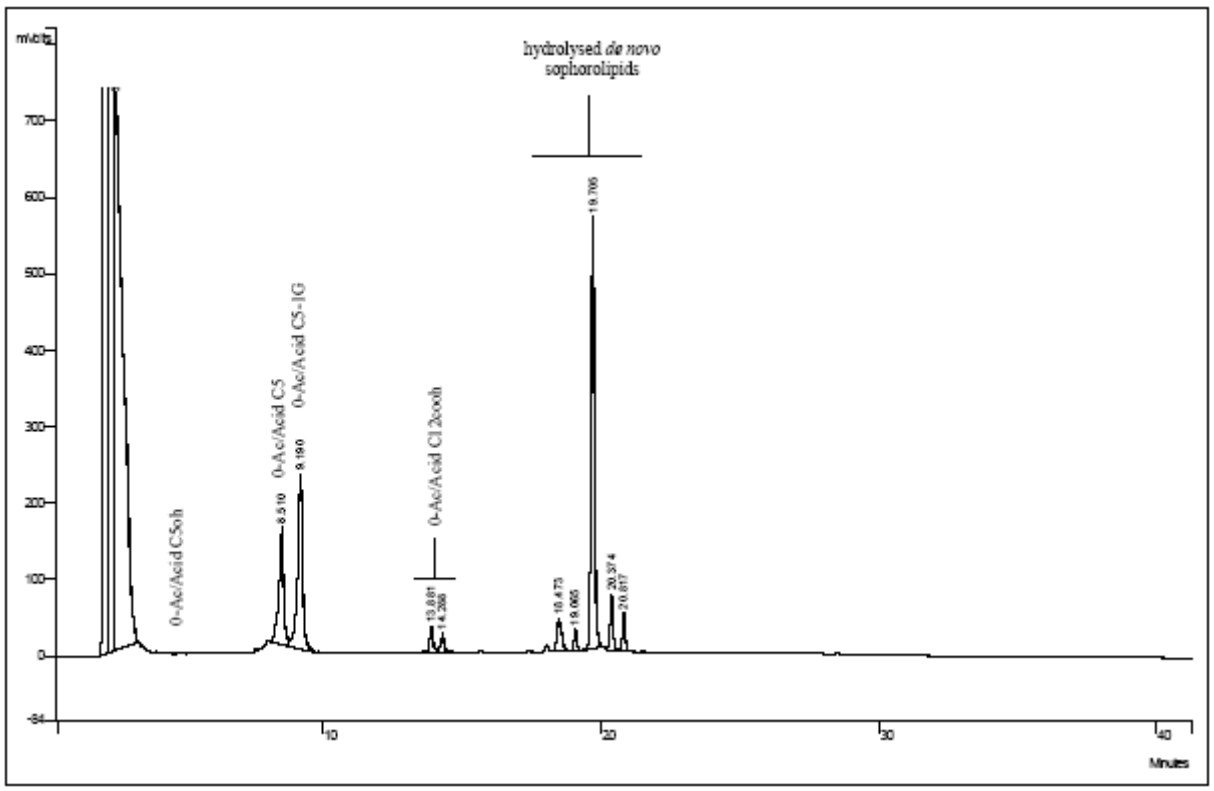


**Figure II-29:** HPLC-chromatogram of alkaline hydrolysed pentenyl dodecanoate sophorolipids. Sophorolipids with pentenyl dodecanoate incorporation are named by the following principle 0-Ac: non-acetylated, Acid: acidic form, C5: pentane lipid tail, C5oh: pentanol lipid tail, C12cooh: dodecanoic acid lipid tail, 1G: one glucose molecule.

## 3.3.4 Dodecyl pentanoate

### 3.3.4.1 Introduction

Dodecyl pentanoate is structurally related to the previously tested dodecyl glutarate: the position of the ester bond is the same, but there is no terminal carboxyl group (Figure II-30). This makes it less comparable with stearic acid, but should allow incorporation into two directions like observed for pentenyl dodecanoate.

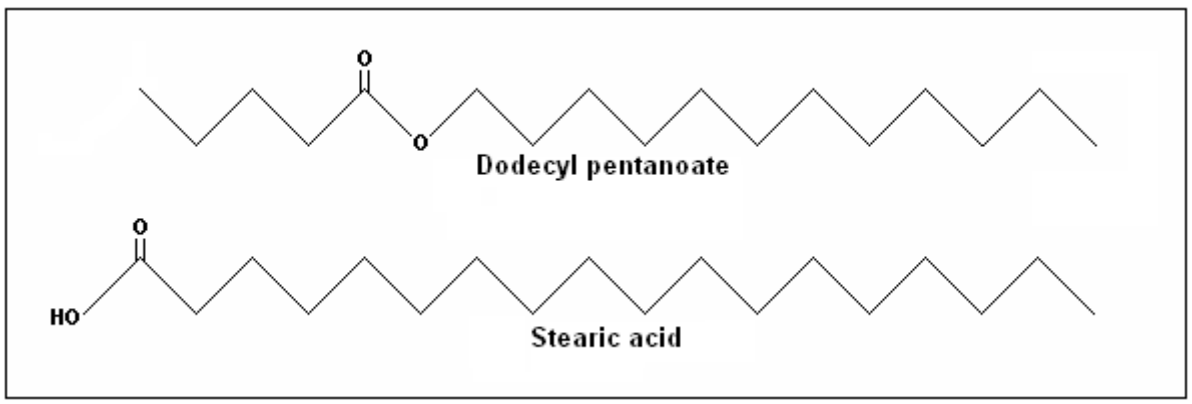


**Figure II-30**: Chemical structure of dodecyl pentanoate compared with the structure of stearic acid.

### 3.3.4.2 Fermentation process

The fermentation on dodecyl pentanoate was run on a 1.5 litre working volume scale. The total running time of the fermentation was 360 hours, which is extremely long. The fermentation time was extended because the cells were negatively affected by the first addition of dodecyl pentanoate 47 hours after inoculation. From this moment onwards, one can see a decrease in CDW, CFU and OD, which lasts till about 120 hours after inoculation (Figure II-31). The reduced cell viability and activity is also reflected in the slow glucose consumption rate (Figure II-32) and the high values for dissolved oxygen during this period (Figure II-33). After this adaptation phase, cell viability seemed to recover and the feeding of dodecyl pentanoate was restarted. One can indeed see that glucose is readily consumed and that sophorolipids start to form.

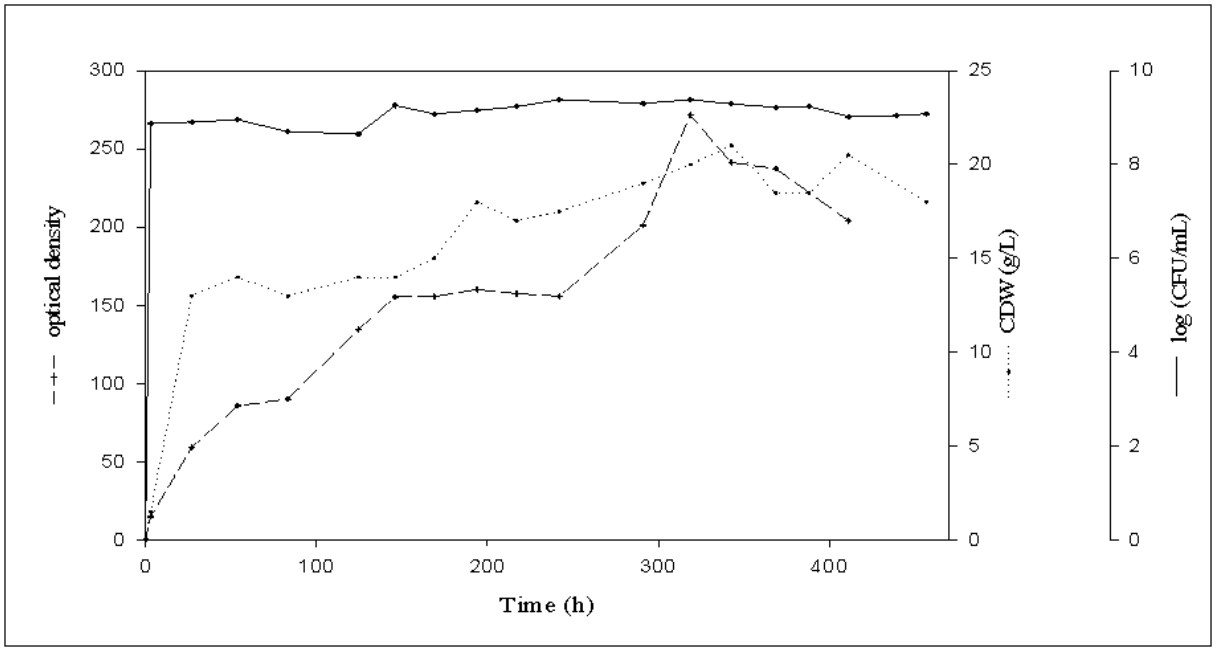


**Figure II-31:** Values for CDW, CFU and OD of the *C. bombicola* cells during the production of sophorolipids on dodecyl pentanoate.

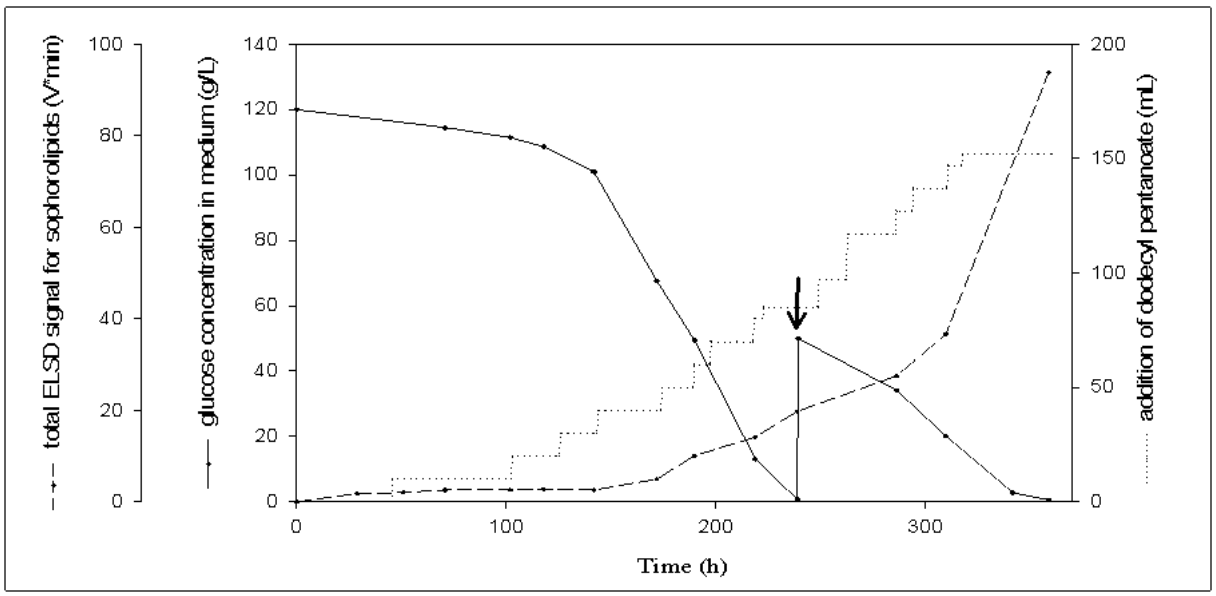


**Figure II-32:** Values for the glucose concentration in the medium (supplementation of additional glucose is marked with an arrow) and the addition of the hydrophobic carbon source for the fermentor scale production of sophorolipids on dodecyl pentanoate. The amount of dodecyl pentanoate is not given in mL since the volume of the culture broth decreased during fermentation because of evaporation and sampling. The ELSD signal form the HPLC-analysis is also plotted; all peaks originating from sophorolipids were added up. Please notice that this method does not allow a correct quantification, but rather gives an indication of the sophorolipid production profile.

A total amount of 152 mL or 138.3 g ester was added (ρ=0.91 g/mL), while only 68.6 g sophorolipids were obtained. Furthermore, a substantial amount of residual ester was extracted from the culture broth. Consequently, the calculated yield for diacetylated lactonic dodecyl pentanoate sophorolipids (MW: 676 g/mol) is 0.100 mol sophorolipids per 0.512 mol substrate or 0.20 mol/mol, which is quite low.

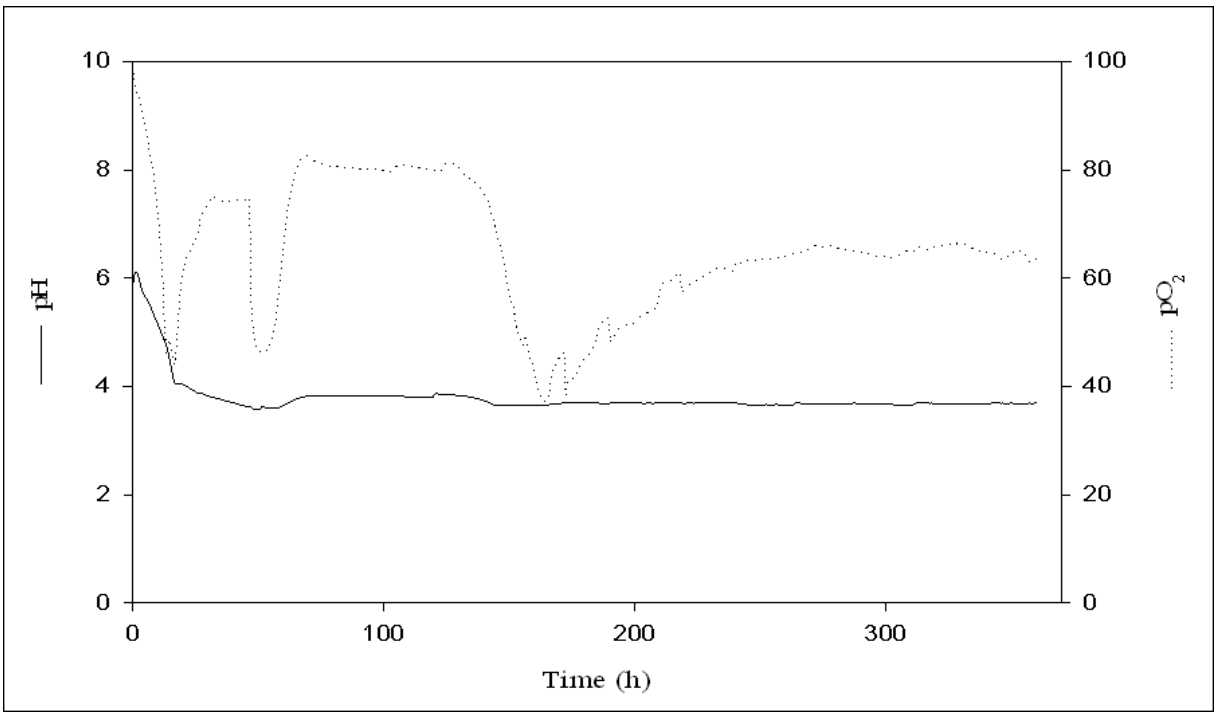


**Figure II-33**: pH and $pO_2$ (%) profile during the production of sophorolipids on dodecyl pentanoate.

### 3.3.4.3 Analysis of the initial and hydrolysed dodecyl pentanoate sophorolipids

When looking at the HLPC analysis of the obtained sophorolipids, it is clear that the mixture is dominated by native *de novo* sophorolipids, but one can also see peaks arising in the hydrophilic area (Figure II-34). Analoguos to the pentenyl dodecanoate sophorolipids, it is believed that the peaks with retention time 20.8 and 23.8 minutes originate from di-acetylated acidic and lactonic dodecyl pentanoate sophorolipids, respectively. The smaller peak at 12 minutes is thought to be caused by degradation of those products resulting in sophorolipids with a pentanoic acid tail.

.

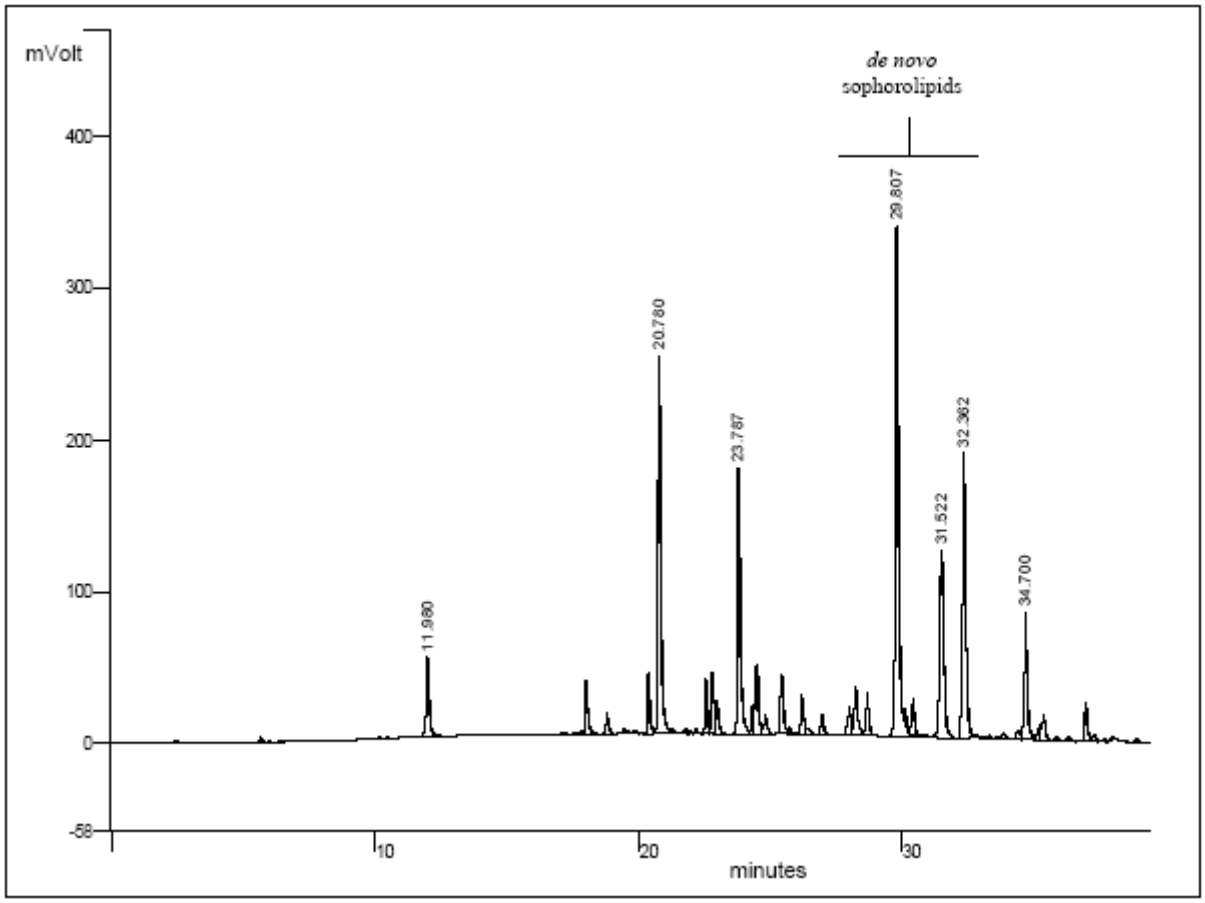


**Figure II-34:** HPLC-chromatogram of sophorolipids obtained from a fermentation with dodecyl pentanoate as hydrophobic carbon source.

LC-MS analysis of the hydrolysed sophorolipid mixture indeed reveals the presence of sophorolipids with a C12 fatty acid tail with a terminal hydroxyl group (Figure II-35, 0-Ac/Acid C12oh-2G); the same structure is also found in the hydrolysed dodecylglutarate and malonate ester sophorolipids. Furthermore, a sophorolipid with a C12 fatty acid tail with an alkyl end was also found (0-Ac/Acid C12-2G). This molecule could only arise after degradation of the substrate into pentanoic acid and dodecanol. This later component can be integrated directly into sophorolipid molecules by its hydroxyl group. This C12 sophorolipid did thus not originated from the alkaline hydrolysis. The hydrophobicity of the terminal alkyl group gave the molecule a far less hydrophilic character as compared to its hydroxylated counterpart (retention time of 15.3 and 21.3 respectively).

As expected, also the very hydrophilic sophorolipid molecule comprising a pentanoic acid moiety was formed (0-Ac/Acid C5-2G).

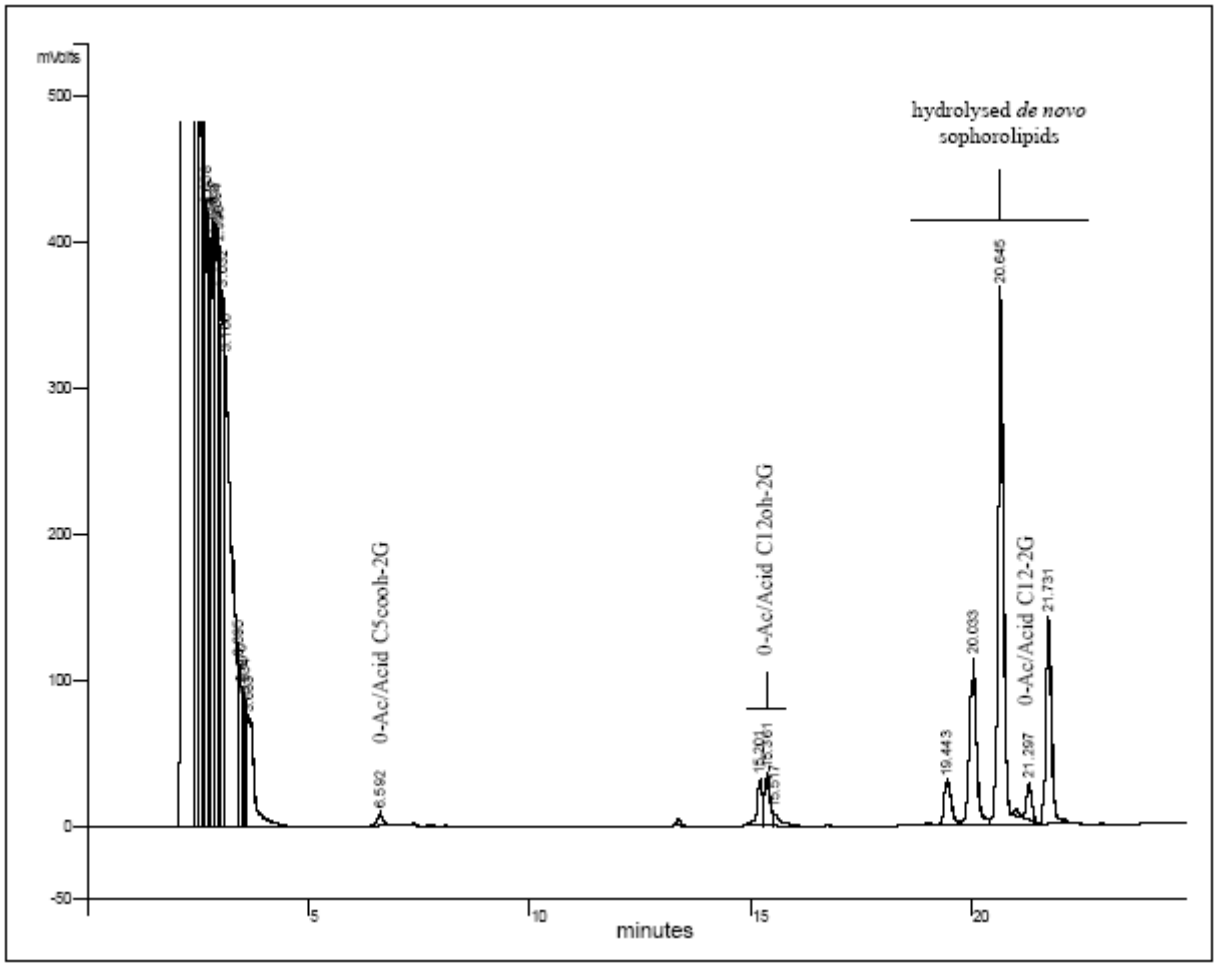


**Figure II-35:** HPLC-chromatogram of alkaline hydrolysed dodecyl pentanoate sophorolipids. Sophorolipids with dodecyl pentanoate incorporation are named by the following principle 0-Ac: non-acetylated, Acid: acidic form, C5: pentanoic acid lipid tail, C12: dodecane lipid tail, C12oh: dodecanol lipid tail and 2G: two glucose molecules.

# 4 Conclusion

The naturally occurring sophorolipids synthesized by *C. bombicola* possess - despite their overall heterogeneity - little variation in the length of the lipid tail. The range is limited to C16-C18 saturated or unsaturated fatty acids and is governed by the specificity of one or more cytochrome P450 monooxygenases. These enzymes hydroxylate the fatty acids at the terminal (ω) or subterminal (ω-1) position and transfer them in this way into suitable substrates for glucosyltransferase I, which will form an ester bond between the introduced hydroxyl group and the first glucose molecule (Figure I-9). However, the incorporation of fatty acids differing from the conventional C16-C18 range could broaden up the application potential of sophorolipids. The incorporation of medium-chain fatty acids should render the molecules more hydrophilic and consequently improve their water solubility.

In order to verify these strict boundaries of hydrophobic carbon source incorporation, fermentations with coconut oil were run. In contrast to other vegetable oils, this oil is very rich in medium-chain fatty acids. However, none of those medium-chain fatty acids was detected in the lipid tail of the obtained sophorolipids. On the other hand, when oils with longer fatty acids, such as meadowfoam oil were used, also here no incorporation was observed. It can therefore be concluded that the synthesis of sophorolipids with a shorter or longer fatty acid tail than the natural occurring ones will not be achieved by simply utilizing those shorter or longer fatty acids as the hydrophobic carbon source.

The specificity of the P450 enzymes towards C16-C18 fatty acids can be circumvented by either using already hydroxylated substrates and so bypassing their controlling action, or by using substrates with a C16-C18 like structure. Regarding this first strategy, it was demonstrated that a ω-hydroxylated medium-chain fatty acid (C12) was incorporated while the corresponding normal fatty acid, dodecanoic acid, was not. Brakemeier *et al* (1998a and b) already reported the incorporation of medium-chain primary and secondary alcohols. In this chapter, it was proven that also 1,12-dodecanediol could be used as substrate for the successful synthesis of medium-chain sophorolipids. Because of the symmetric character of this molecule, sophorose units could be introduced at both sites, resulting in a wide heterogeneity among the obtained glycolipids. The calculated yield ranged between 0.36 and

0.47 mol sophorolipids per mol substrate, which could mean that part of the substrate is metabolized via β-oxidation (Table II-6). The described experiments demonstrate that (terminal) hydroxylated substrates are integrated into sophorolipid molecules regardless their chain length.

**Table II-6:** Maximal yields for the unconventional substrates described in Chapter II.

| Substrate | Feature | Yield (mol/mol) |
|---|---|---|
| 12-Hydroxydodecanoic acid | hydroxylated | 0.29* |
| 1,12-Dodecanediol | hydroxylated | 0.47 |
| Dodecyl glutarate | carboxylated ester | 0.70 |
| Dodecyl and myristyl malonate | carboxylated ester | 0.70 |
| Pentenyl dodecanoate | alkyl ester | 0.25 |
| Dodecyl pentanoate | alkyl ester | 0.20 |

*In contrast to the other results, obtained in shake flask cultures.

It turns out that the specificity of the cytochrome P450 monooxygenases is quite selective towards the fatty acid chain length, but not towards the exact structure of the offered substrate. For example, the presence of unsaturated bonds does not hinder their activity; even arachidonic acid (C20:4) is hydroxylated (Phadtare *et al.*, 2004). The incorporation of unconventional stearic acid-like substrates which can undergo post-fermentative modification to give rise to shorter chained sophorolipids, could open up perspectives for the synthesis of new-to-nature glycolipids. Substrates with internal ester bonds are particularly useful for this purpose; the bond can be cleaved by alkaline hydrolysis. This process also de-acetylates the glucose molecules and opens lactonic bonds eventually present, resulting in a hydrolysed mixture with reduced complexity as compared to the initial sophorolipids. Substrates with a carboxylic end such as dodecyl glutarate and malonate ester are preferred over substrates with alkyl ends such as pentenly dodecanoate and dodecyl pentanoate. The carboxyl group only allows incorporation into one direction; in this way controlling the chemical nature of the hydrolysed sophorolipids. Substrates without such carboxylic end, such as pentenyl dodedecanoate and dodecyl pentanoate, can be hydroxylated by cytochrome P450 monooxygenase at no matter which site, making two ways of integration possible. Another

reason why carboxylated substrates are favoured for sophorolipid production is the fact that clearly better yields are obtained (Table II-2). For both carboxylated esters, the theoretical maximal yield is 0.70 mol sophorolipids per mol substrate, while this is only 0.25 or 0.20 for the non-carboxylated ones. This latter substrate is possibly metabolized via the β-oxidation route. It is worth mentioning that the yields on vegetable oils such as rapeseed oil, which is one of the best performing hydrophobic carbon sources, are also situated around 0.70 mol/mol.

A phenomenon which is not observed for the native sophorolipids is the synthesis of glycolipids with an odd number of glucose molecules. Therefore, it is thought that glucosyltransferase II, the enzyme which binds the second glucose unit to the previously formed glucolipid, is in some sort of way hindered by the internal ester bond and consequently displays a lower affinity for this glucolipid. The same event is to a minor extent observed for pentanol in the fermentation with pentenyl dodecanoate. Also here it is believed that the altered nature of the glucolipid causes a decreased affinity of the enzyme.

Substrates with an internal ester bond can be used to synthesise almost any medium- or short-chain sophorolipid. By applying substrates with a terminal carboxyl group, the incorporation can be orchestrated and by varying the position of the ester bond, the length of the lipid acid tail after hydrolysis can be fixed. By altering the orientation of the ester bond, *i.e.* towards or away from the terminal carboxyl group, one can obtain either carboxylated or hydroxylated hydrolysed sophorolipids.

The use of these special substrates with an internal ester bond for glycolipid production is meanwhile patented by the company Ecover Belgium N.V. (Develter and Fleurackers, 2007).

# Chapter III: Development of a transformation and selection system for *Candida bombicola*

Part of this chapter has been published as:

Van Bogaert INA, De Maeseneire SL, De Schamphelaire W, Develter D, Soetaert W, Vandamme EJ (2007) Cloning, characterisation and functionality of the orotidine-5′-phosphate decarboxylase gene (*URA3*) of the glycolipid producing yeast *Candida bombicola*. Yeast 24, 201–208

Van Bogaert INA, De Maeseneire SL, Develter D, Soetaert W, Vandamme EJ (2008). Development of a transformation and selection system for the glycolipid producing yeast *Candida bombicola*. Yeast 25, 273-278

# 1 General introduction

Sophorolipids are surface-active compounds synthesized by the non-pathogenic yeast *Candida bombicola*. Those glycolipids find several interesting industrial, cosmetic and pharmaceutical applications. Over the last decades, much effort has been spent to optimize the fermentation process in order to improve the yield. However, as far as we know, hardly any attention has been given to the genetics of the producing yeast strain itself and there are no published results available on genetic engineering of *C. bombicola*. Nevertheless, this can be a useful tool for the study of the sophorolipid synthesis pathway and open up perspectives for improved production. A first step in the genetic engineering of this yeast species is the development of a suitable transformation and selection method.

At first, a suitable transformation method was selected and further optimized for *C. bombicola*. The transformation method was tested by using dominant drug resistance genes (Section 2). Since such a selection system is expensive and not very convenient for large scale industrial fermentations, most yeast transformation systems are based on auxotrophic markers. More specifically, systems based on the orotidine-5'-phosphate decarboxylase (*URA3*) gene have been widely used for *Saccharomyces cerevisiae* (Rose *et al.*, 1984), other yeasts (Francois *et al.*, 2004; Ngan *et al.*, 2000) and filamentous fungi (Fierro *et al.*, 2004).

In order to develop such a system, the *URA3* gene of *Candida bombicola* was isolated (GenBank accession number DQ916828) and its functionality was proven by complementation of an *URA3*-negative *Saccharomyces cerevisiae* strain (Section 3).

Next step in the development of the transformation and selection system is the isolation of a stable ura3-auxotrophic mutant which can be transformed back to prototrophy with the species-own *URA3* gene (Section 4). With some luck, those mutants can be found by selection on media containing 5-fluoroorotic acid (5-FOA) which is converted into a toxic analogue by organisms with a functional URA3, but is non-toxic to mutants lacking this activity. Several mutants were tested and evaluated and one with promising characteristics was used for transformation experiments. Successful transformation was confirmed by a PCR-based method discriminating between the wild type and mutated *URA3* gene.

# 2 Transformation of *C. bombicola*

## 2.1 Introduction

The advent of yeast molecular genetics can be situated in 1978, with the succesful transformation of *Saccharomyces cerevisiae*, using auxotrophic markers and the development of *Escherichia coli* shuttle vectors for this organism (Beggs, 1978). Since that date, numerous vectors with diverse selection markers, promoters and other features are created and available for *S. cerevisiae*. Also for other yeasts used as model organism in molecular and cell biology and/or host for heterologous expression such as *Pichia pastoris* and *Schizosaccharomyces pombe*, or medically important yeasts such as *Candida albicans*, a whole battery of vectors and tools for genetic engineering is meanwhile available.

The genetic "domestication" of a certain wild type organism starts by finding a way to introduce foreign DNA into the host cells. For yeasts, the most common techniques are electroporation, chemical transformation by lithium acetate or lithium chlorate and protoplast transformation. In the work presented here, chemical transformation with lithium acetate was chosen because this technique requires no special equipment and because the critical phase of protoplast formation is circumvented. Since only the wild type strain was available, dominant drug resistance markers were selected to test and optimize the transformation method. In yeasts and fungi, the most commonly used markers are the *E. coli* hygromycin B resistance gene, the *Streptoalloteichus hindustanus* bleomycin/phleomycin gene which renders the cells resistant to zeocin and the neomycin/kanamycin/G418 resistance which allows yeast cells to survive in medium containing G418 (Ruiz-Dìez, 2002).

## 2.2 Materials and Methods

### 2.2.1 Strains, plasmids and culture conditions

*C. bombicola* ATCC 22214 was used in all transformation experiments and was grown in yeast peptone dextrose (YPD) medium (1 % yeast extract, 2 % peptone, 2 % glucose). Selection of transformants was carried out on YPD plates (2 % agar) with the appropriate amount of filtersterile antibiotic added after autoclavation of the other components. G418 and hygromycin were obtained from Sigma, zeocin from Invitrogen. *C. bombicola* was incubated at 30 °C (and 200 rpm for liquid cultures). *Saccharomyces cerevisiae* FY1679-01B (*MATa*;

*ura3-52*; *LEU2*; *TRP1*; *HIS3*; *GAL2*) from Euroscarf (Germany) was used as a reference strain to test the quality of the plates.

*Escherichia coli* DH5α was used to maintain plasmids and was grown in Luria-Bertani (LB) medium (1 % trypton, 0.5 % yeast extract and 0.5 % sodium chloride) supplemented with 0.01 % ampicillin if necessary. Liquid *E. coli* cultures were incubated at 37 °C and 200 rpm.

The plasmids pEX2kanMX4 (LMBP 5010, Figure III-1) and pREP4FLAG (LMBP 4343, Figure III-2) were obtained from the Laboratory of Molecular Biology, Ghent University (LMBP).

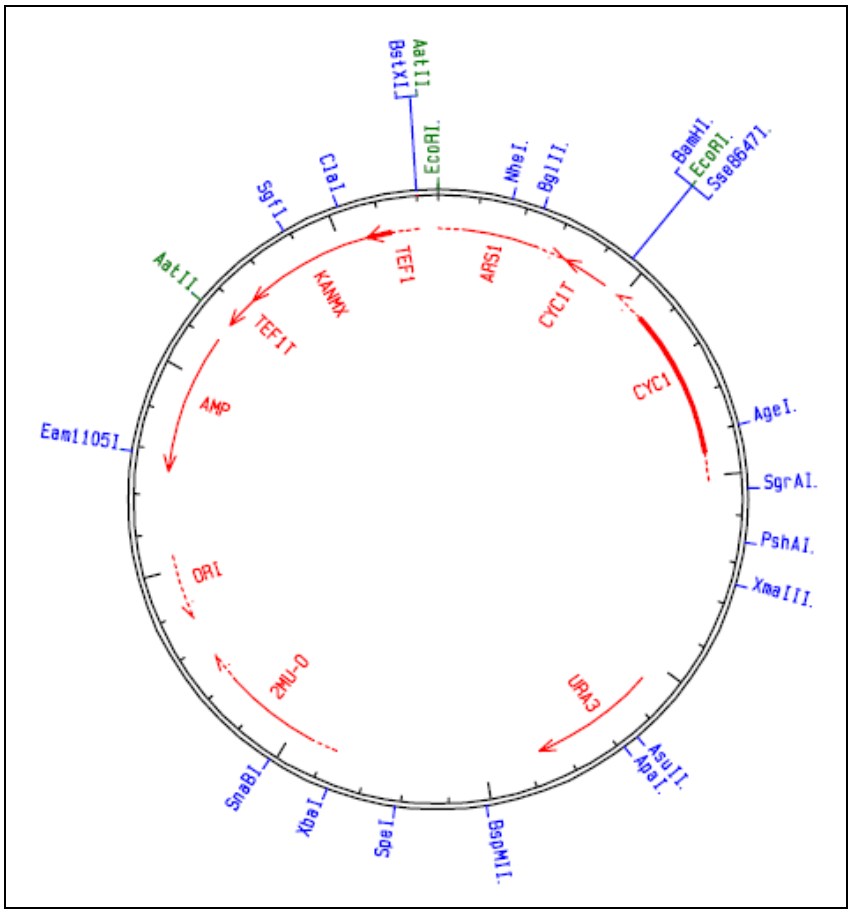


**Figure III-1:** Circular map of pEX2kanMX4 or LMBP 5010 (10 609 bp).

Plasmid DNA was isolated with the QIAprep Spin Miniprep Kit (Qiagen).

Linear DNA for transformation was obtained by standard PCR with the primers hygro4343For and hygro4343Rev (Table III-1) and pREP4FLAG as template. The PCR reaction mixture was prepared for transformation by using the GenElute™ PCR CleanUp Kit (Sigma).

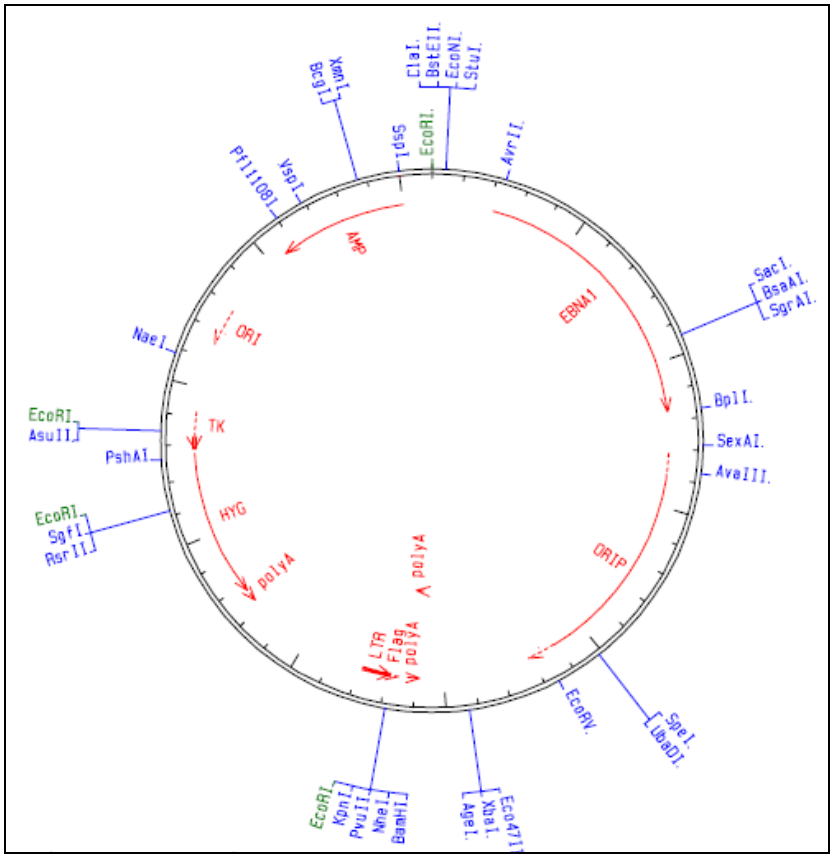


**Figure III-2:** Circular map of pREP4FLAG or LMBP 4343 (10 189 bp).

**Table III-1:** Primers used for testing the transformation method for *C. bombicola*. All primers were obtained from Sigma Genosys.

| Name | Feature | Sequence |
|---|---|---|
| hygro4343For | creation linear fragment for transformation | 5'-AAC GCC AGC AAG ACG TAG C-3' |
| hygro4343Rev | creation linear fragment for transformation | 5'-GGA AGA AAC GCG GGC GTA TT-3' |
| hygroInsertCheckFor | checking presence of hygromycin gene | 5'-TTC GAC AGC GTC TCC GAC CTG AT-3' |
| hygroInsertCheckPromFor | checking presence of hygromycin gene | 5'-GAA TTC GAA CAC GCA GAT G-3' |
| hygroInsertCheckRev | checking presence of hygromycin gene | 5'-GCG ATT TGT GTA CGC CCG ACA GT-3' |

### 2.2.2 Transformation protocol for *Saccharomyces cerevisiae*

An overnight grown yeast culture (YPD) was diluted to an optical density (OD) of 0.2 in 300 mL fresh YPD medium and grown for an additional 3 to 5 hours. The cells were harvested by centrifugation at 1 000 g for 5 min at room temperature and resolved in 50 mL sterile water. After centrifugation, the pellet was resolved in 1.5 mL Solution I (100 mM LiAc, 10 mM Tris-HCl, 1 mM EDTA, pH 7.5). For each transformation, 100 µL cells were transferred in a

sterile eppendorf tube and add 100 µg carrier DNA and 0.1 µg plasmid or 1 µg linear DNA was added. Carrier DNA (Salmon Sperm DNA) was denaturated by heating it at 95 °C for 5 min and immediately putting it on ice afterwards. 600 mL Solution II (40 % PEG 4000, 100 mM LiAc, 10 mM Tris-HCl, 1 mM EDTA, pH 7.5) was added. The transformation mixtures were vortexed and incubated at 30 °C and 250 rpm for 30 min. 70 µL 100 % DMSO was added and the cells were heat shocked at 42 °C for 15 min. The transformation mixtures were centrifuged at 17 000 g for 5 min. The cell pellet was resolved in 0.5 mL TE-buffer (10 mM Tris-Cl, 1 mM EDTA, pH 7.5 adjusted with HCl) and plated on selective plates.

### 2.2.3 Transformation protocol for *Candida bombicola*

The protocol is similar the one described for *S. cerevisiae*, but has the following modifications: only 50 mM LiAc is used instead of 100 mM, the incubation period at 30 °C lasts for 90 min instead of 30 and no DMSO is added.

### 2.2.4 Testing transformants by yeast colony PCR

Integration of the hygromycin resistance gene was tested by yeast colony PCR with the primer pairs hygroInsertCheckFor - hygroInsertCheckRev, and hygroInsertCheckPromFor - hygroInsertCheckRev (Table III-1). The temperature program consists of an initial denaturation of 94 °C for 7 min, a 30 fold repeat of 94 °C for 30 s, a specific annealing temperature for 30 s, 72 °C for 2.5 min with 5 s increasing time and a final 7 min elongation at 72 °C. The specific annealing temperature is 62 °C for the first primer pair and 57 °C for the second one.

## 2.3 Results and Discussion

### 2.3.1 Development of the selective media

#### 2.3.1.1 G418

Before testing the actual transformation method, one must know which drug resistant markers can be applied and in what concentration they should be added to the selective media in order to inhibit cell growth even at high cell densities. The first antibiotic tested is G418, an aminoglycoside related in structure to gentamicin, neomycin and kanamycin. It blocks polypeptide synthesis by interfering with ribosomal function through binding on the 80S subunit. The recommended concentrations for use as a selection agent are 100-800 µg/mL. For *S. cerevisiae* for example, a concentration of 200 µg/mL is used. Resistance is conferred by one of two dominant genes of bacterial origin which can be expressed in eukaryotic cells: aminoglycoside phosphotransferase 3' I and II (Barnun *et al.*, 1983).

*C. bombicola* turned out to be quite resistant to G418: there is still minimal growth at a concentration of 800 µg/mL. There were no colonies observed on agar plates with 1200 or 1600 µg/mL, even at high cell densities (> $10^9$ cells/mL). To check the quality of the plates, *S. cerevisiae* was also tested. This species did not show any growth at 200 µg/mL G418.

Initially, the transformation protocol for *C. bombicola* was tested by transforming the cells with pEX2kanMX4 (Figure III-1). This plasmid harbours a bacterial kanamycin resistance gene (Tn 903) under control of the transcription elongation factor 1α (*TEF1α*) promoter and terminator from the filamentous fungus *Eremothecium gossypii*, rendering yeast species resistant to G418.

When the transformation protocol was tested with plates containing 1200 µg/mL G418, colonies appeared on plates of the negative control transformed with sterile MQ-water. When the plates were retested with untransformed cells, no growth was observed. It therefore was concluded that nothing was wrong with the agar plates, but that the treatment of the cells during transformation or their specific growth phase renders the cells less sensitive to G418.

Although there were fewer colonies observed on the negative control plates than on the plates with transformed yeasts, which could led us to conclude that true transformants were created, we decided to choose another antibiotic selection system to optimize the transformation, especially since the required concentration G418 was already extremely high.

### 2.3.1.2 Zeocin

Another yeast species with high G418 tolerance is *Pichia farinosa*. Concentrations of 1000 µg/mL did not inhibit cell growth. *P. farinosa* on the other hand, is unable to grow on YPD plates containing 200 µg/mL zeocin (Wang *et al.*, 2006). More or less the same trend was observed for *Candida glabrata* (Alderton *et al.*, 2006). Zeocin is an antibiotic of the bleomycin/phleomycin group. Once the drug enters the cell, it binds and breaks down DNA. Resistance to zeocin can be achieved by expression of the *Streptoalloteichus hindustanus* bleomycin resistance gene, the corresponding protein inhibits zeocin action by binding it stoichiometrically. A concentration of 50 – 300 µg/mL zeocin should be sufficient to inhibit growth of yeasts in YPD or minimal medium agar plates. The pH must range between 6.5 and 8.0 and the optimum has to be determined for each species.

First, the zeocin concentration was tested. *C. bombicola* was able to grow on YPD agar plates containing 50 and 150 µg/mL zeocin. At 300 µg/mL colonies still appeared, but not at very low cell densities. However, *S. cerevisiae* was completely inhibited at 300 µg/mL. The pH of those plates was left unchanged at the initial 6.48. In a next step, the optimal pH was determined (at 300 µg/mL zeocin). The results are shown in Table III-2. Before the actual test, it was verified if *C. bombicola* was affected in its viability by a pH of 8.0. This was not the case. It is clear that the optimal pH for selection of *C. bombicola* transformants is 8.0. Yet, when the incubation lasts longer than 3 days, also at this pH colonies appear, although only at very high cell densities. For this reason higher amounts of the antibiotic were tested at pH 8.0, but the results were comparable to those obtained for 300 µg/mL (Table III-3). A possible explanation for this phenomenon is the degradation of the antibiotic caused by the long term exposure to 30 °C, or the ability of *C. bombicola* to develop some sort of resistance to zeocin when exposed to it long enough.

**Table III-2:** Effect of pH of the YPD agar plates on zeocin toxicity for *C. bombicola*.

| **Incubation time** | **Minimal cell density (cells/mL) for which growth is observed** | | | |
|---|---|---|---|---|
| **(days)** | **pH 6.5** | **pH 7.0** | **pH 7.5** | **pH 8.0** |
| 3 | $10^4$ | $10^6$ | $10^8$ | - |
| 4 | $10^3$ | $10^5$ | $10^6$ | $10^9$ |
| 5 | $10^3$ | $10^5$ | $10^5$ | $10^9$ |
| 6 | $10^3$ | $10^5$ | $10^5$ | $10^5$ |

**Table III-3:** Evaluation of high concentrations zeocin at long incubation times.

| Incubation time | Minimal cell density (cells/mL) for which growth is observed | | |
|---|---|---|---|
| (days) | 300 µg/mL | 500 µg/mL | 700 µg/mL |
| 5 | - | - | - |
| 9 | $10^6$ | $10^7$ | $10^6$ |

It was concluded that 300 µg/mL zeocine at pH 8 should be used, but that one must keep in mind that false positives colonies can be present among the ones appearing after longer incubation times. It is not yet clear how long it takes for *C. bombicola* transformed with the zeocin resistance gene to grow and form a colony on a selective plate after the transformation procedure.

#### 2.3.1.3 Hygromycin

Hygromycin belongs to the group of the aminoglycoside antibiotics. Its mode of action is the inhibition of protein synthesis, by inducing the misreading of the mRNA. This antibiotic has been used to select drug-resistant stable transformants after transfer of the hygromycin phosphotransferase gene in a variety of prokaryotic and eukaryotic organisms. For use as a selective agent, the suggested concentration is about 200 µg/mL for lower eukaryotes (Gonzalez *et al.*, 1978).

Because of the high tolerance of *C. bombicola* to the previously tested antibiotics, concentrations of 200, 500 and 1000 µg/mL were tested. No growth was observed with the concentrations of 500 and 1000 µg/mL, even after 11 days of incubation. At a concentration of 200 µg/mL, colonies appeared after 6 days, but only at the highest cell concentration of $10^9$ cells/mL. After 11 days, growth was observed starting from cell densities of $10^7$ cells/mL. So a concentration of 200 µg/mL could be used if the plates are not incubated too long. However, it was decided to work with 500 µg/mL because this totally inhibits growth of the yeast cells and still is effective even after long incubation times.

### 2.3.2 Development of the transformation protocol

We decided to transform *C. bombicola* by means of the Lithium Acetate method, and the initial protocol was one used for *S. cerevisiae* (see Materials and Methods section 2.2.2)

As described in the above Section 2.3.1.1, at first the transformation protocol was tested by transforming the cells with pEX2kanMX4, a plasmid rendering the *C. bombicola* cells resistant to G418. We stopped using G418 since false positives were observed, even at very high G418 concentrations. Further experiments were conducted with hygromycin and the vector pREP4FLAG (LMBP 4343, Figure III-2) harbouring a hygromycin phosphotransferase gene under control of the *Herpes simplex* virus tyrosine kinase (*TK*) promoter. Viral promoters are believed to be active in a wide range of hosts.

After the first transformation attempt, it became clear that the transformation protocol as it was, was extremely toxic for the *C. bombicola* cells: few colonies were observed on the non-selective YPD plates. Nevertheless, when the concentration of DMSO was decreased, better viability was observed. DMSO stayed however toxic, even when only 1 µL instead of 100 µL was added. For this reason, it was decided to omit the DMSO step from the protocol.

Furthermore, the initial protocol works with 100 mM LiAc. For certain yeasts such as *Yarrowia lipolytica*, this concentration is too high and 50 mM is used instead (Finlayson *et al.*, 1991). This also gave better results for *C. bombicola*. Additonally, the incubation time at 30 °C was prolonged to 90 minutes; additional augmentation did not influence the transformation efficiency. The duration of the heat shock was also evaluated: 15 minutes was kept as the optimal period. The final transformation protocol for *C. bombicola* can be seen in the Materials and Methods section 2.2.3.

### 2.3.3 Verification of *C. bombicola* transformants

The pREP4FLAG vector is actually designed for mammalian host cells. Consequently, it has no autonomous replication centre (ARS) or other elements for plasmid amplification and maintenance in yeast (Figure III-2). The observed transformants so probably display some sort of (partial) integration of the plasmid. However, stable integration of the foreign DNA into the genome is preferred: the insert is rarely lost and the selection pressure does not need to be maintained constantly. For this reason, we also transformed *C. bombicola* with a linear fragment of pREP4FLAG. The primers hygro4343For and hygro4343Rev (Table III-1) were used to amplify the hygromycin resistance gene flanked by the *TK* promoter and terminator (1533 bp). The first transformants only appeared after eight days of incubation. After 15 days

of incubation, a colony was observed on the negative control plates, so incubation was stopped.

In order to check if the colonies appearing on the selective plates were actually transformed with the linear fragment, the primer pairs hygroInsertCheckFor and hygroInsertCheckRev, and hygroInsertCheckPromFor and hygroInsertCheckRev (Table III-1) were used. The first combination amplifies 875bp of the coding sequence of the antibiotic resistance gene, the second pair also targets the *TK* promoter and should yield a fragment of 1079 bp. All 7 obtained colonies were tested by a yeast colony PCR and the results are shown in Figure III-3.

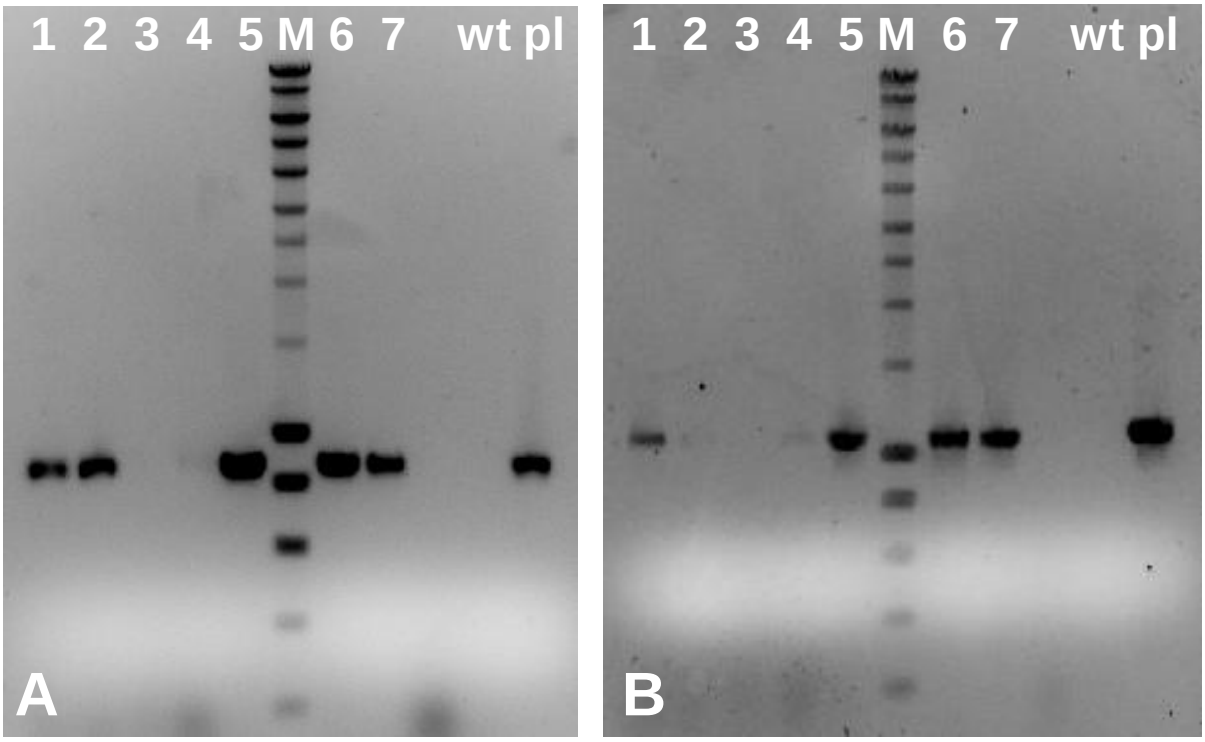


**Figure III-3:** Yeast colony PCR on the 7 possible transformants (1-7), untransformed *C. bombicola* cells (wt) and pREP4FLAG (pl) with (A) the primers hygroInsertCheckFor and hygroInsertCheckRev and (B) the primers hygroInsertCheckPromFor and hygroInsertCheckRev.

5 out of 7 colonies actually contained the hygromycin resistance gene. For only one of them, there was no indication that the *TK* promoter was also integrated, so the gene was probably expressed due to the action of an endogenous nearby promoter.

Those results prove that *C. bombicola* was successful transformed with foreign DNA and that this DNA can be integrated into the genome, very likely by illegitimate recombination since the hygromycin phosphotransferase gene and the *TK* promoter and terminator are not present

in the yeast genome. Nevertheless, the number of obtained transformants is low (7 colonies for two transformations with 1.63 µg linear DNA, yields 2.2 colonies/µg DNA), but it can be expected that this number is higher when species own sequences are used and homologous recombination becomes possible (see Section 4 of this chapter).

# 3 Cloning of the orotidine-5'-phosphate decarboxylase gene of *C. bombicola*

## 3.1 Introduction

A lot of research has been performed on the optimization of the sophorolipid fermentation process of *C. bombicola*, but as far as we know, hardly any work has been published on the genetics of the producing yeast strain itself. However, genetic engineering of this yeast species could open up perspectives for higher yields and modification of the glycolipid mixture produced. In this context, a prerequisite is the development of an efficient transformation and selection system for *C. bombicola*. One can use dominant drug resistance markers, but such a selection system is expensive and not very convenient for large scale industrial fermentations. Thus, most yeast transformation systems are based on auxotrophic markers. More specifically, systems based on the *URA3* gene are widely used in *Saccharomyces cerevisiae* (Rose *et al.*, 1984), other yeasts (Francois *et al.*, 2004; Ngan *et al.*, 2000) and fungi (Fierro *et al.*, 2004; Goosen *et al.*, 1987). The *URA3* gene product, orotidine-5'-phosphate decarboxylase (OMPD), takes part in the *de novo* pyrimidine biosynthetic pathway. Consequently, *ura3* mutants are uracil or uridine auxotrophic.

The next pages describe the cloning and sequence analysis of the *URA3* gene of *C. bombicola* ATCC 22214. The functionality of the gene was demonstrated by complementation of a defined *URA3*-negative *S. cerevisiae* strain. The isolation of the *URA3* gene is the first step in the development of a *C. bombicola* transformation and selection system.

## 3.2 Materials and Methods

### 3.2.1 Strains, plasmids and culture conditions

*C. bombicola* ATCC 22214 was used for the preparation of genomic DNA. *S. cerevisiae* FY1679-01B (*MATa*; *ura3-52*; *LEU2*; *TRP1*; *HIS3*; *GAL2*) from Euroscarf (Germany) was used as a host for the complementation test. *Escherichia coli* DH5α was used in all cloning experiments.

All PCR products were cloned into the pGEM-T® vector (Promega).

*C. bombicola* was cultured in medium containing 10 % glucose, 1 % yeast extract and 0.1 % urea. *S. cerevisiae* FY1679-01B was grown in yeast peptone dextrose (YPD) medium (1 % yeast extract, 2 % peptone, 2 % glucose) or in synthetic dextrose (SD) medium [ 0.67 % yeast nitrogen base without amino acids (DIFCO) and 2 % glucose] for the complementation test. Liquid yeast cultures were incubated at 30 °C and 200 rpm. *E. coli* was grown in Luria Bertani (LB) medium (1 % trypton, 0.5 % yeast extract and 0.5 % sodium chloride) supplemented with 100 mg/L ampicillin and 40 mg/L X-gal if necessary. Liquid *E. coli* cultures were incubated at 37 °C and 200 rpm.

### 3.2.2 DNA isolation and sequencing

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

### 3.2.3 Transformation

*S. cerevisiae* cells were transformed by the method described in Section 2.2.2. *E. coli* cells were transformed as described by Sambrook and Russell (2001).

### 3.2.4 Degenerate PCR

Part of the *URA3* gene of *C. bombicola* was amplified by use of the degenerate primers ura3For1, ura3Rev1, ura3For2 and ura3Rev2 (Table III-4) based on the conserved sequences of the *URA3* genes of the *Saccharomycotina*. PCR amplification was carried out with an initial denaturation of 94 °C for 4 min, a 40 fold repeat of 94 °C for 30 s, 50 °C for 1 min, 72 °C for 2 min with 5 s increasing time and a final 7 min elongation at 72 °C.

**Table III-4:** Primers used for the isolation and cloning of the *C. bombicola URA3* gene. All primers were obtained from Sigma Genosys.

| Name | Feature | Sequence |
|---|---|---|
| ura3For1 | degenerate primer | 5'-TTY GAR GAY MGN AAR TTY GC-3' |
| ura3Rev1 | degenerate primer | 5'-ACH GTY CKR TAY TGY TGD CC-3' |
| ura3For2 | degenerate primer | 5'-AAR TTY GCH GAY ATH GG-3' |
| ura3Rev2 | degenerate primer | 5'-CCH ACH CCD GGN GTC A-3' |
| ura3U1 | primary GSP for upstream amplification | 5'-TCG AAC TCA GTT CTG CAA GCA TGA CAA-3' |
| ura3Un | nested GSP for upstream amplification | 5'-CAG CCT CTT CTA GTC CGC TCA CAA TTC C-3' |
| ura3D1 | primary GSP for downstream amplification | 5'-ACA CTG GCT CAC GGC GAA TAC TCG CAA-3' |
| ura3Dn | nested GSP for downstream amplification | 5'-GCG ACA GTA GAC ATC GCT CGC AGT AAC-3' |
| ura3TotFor | primer for isolation of 2061 bp fragment | 5'-GCT GGT CTG ACG GGC GGA TAG TAC A-3' |
| ura3TotRev | primer for isolation of 2061 bp fragment | 5'-CAT CAT CGT CAC TAT ACA CAT CGT CAT-3' |

### 3.2.5 Genome Walking

The two unknown genomic DNA sequences upstream and downstream of the degenerate PCR fragment were identified by genome walking, which was carried out according to the manual of the BD GenomeWalker™ Universal Kit (BD Biosciences). The strategy used for the isolation of the complete gene is shown in Figure III-4. For both the upstream and downstream sequence, two PCR amplifications – a primary PCR followed by a nested PCR – were performed. Four gene-specific primers (GSP) were designed and are listed in Table III-4. The PCR reaction mixture and cycles were optimized for use with the Expand Long Template PCR System (Roche Diagnostics), as described by De Maeseneire *et al.* (2006).

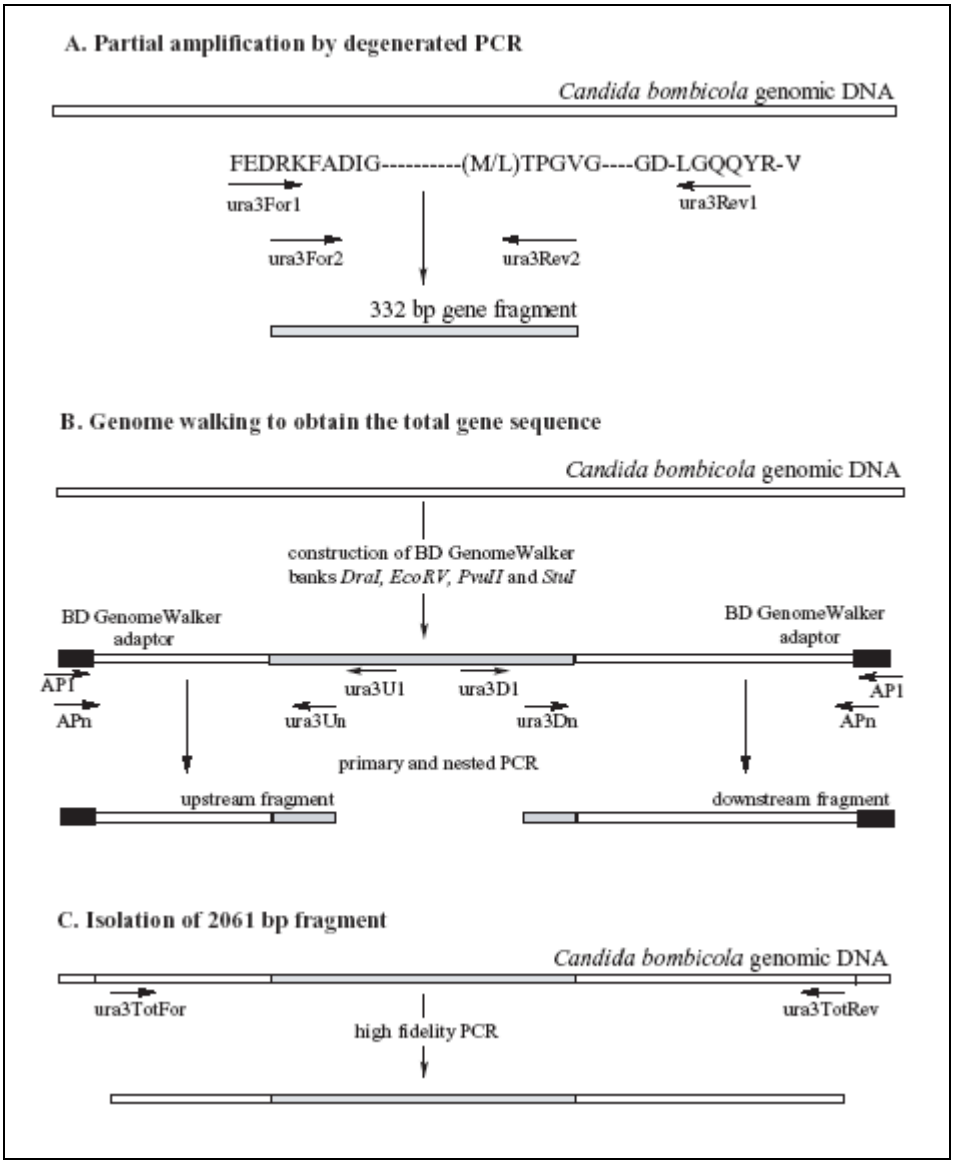


**Figure III-4:** Schematic representation of the strategy followed to obtain the complete sequence of the *URA3* gene of *Candida bombicola*. (A) PCR with degenerate primers (Table 1) designed on the conserved regions of yeast *URA3* gave rise to a 332 bp fragment. (B) Four genomic libraries were constructed by cutting genomic DNA with four different restriction enzymes and ligating these fragments to the BD GenomeWalker™ adaptor. The unknown regions were amplified with the BD GenomeWalker™ adaptor primers AP1 and APn (for primary and nested PCR respectively) and the gene specific primers as listed in Table III-4. (C) The complete *Candida bombicola URA3* gene was physically isolated by using the primers ura3TotFor and ura3TotRev.

### 3.2.6 Cloning and sequence analysis

The isolation of the complete *URA3* gene with its upstream and downstream flanking regions was carried out with the two specific primers ura3TotFor and ura3TotRev (Table III-4). The fragment was amplified from genomic DNA with the High Fidelity PCR Master kit (Roche

Diagnostics) using the following temperature program: initial denaturation at 94 °C for 2 min, a 10 fold repeat of 94 °C for 10 s, 50 °C for 30 s, 72 °C for 2 min, a 15 fold repeat of 94 °C for 15 s, 50 °C for 30 s, 72 °C for 2 min with 5 s increasing time and a final 7 min elongation at 72 °C. The obtained PCR-fragment was cloned in pGEM-T®, and the resulting vector was called pCbUra3.

Sequences were analysed with the Clone Manager Professional Suite Software (Version 6.0). The BLAST program (Altschul *et al.*, 1997) was used for similarity searches in databases available on the NCBI website (http://www.ncbi.nlm.nih.gov). Multiple sequence alignments were made with the CLUSTAL W program (Higgins *et al.*, 1992). Subsequently, a phylogenetic tree was constructed in BioEdit using the Protein Maximum Likelihood (ProML) algorithm.

### 3.2.7 Nucleotide sequence accession number

The nucleotide sequence described in this paper has been deposited at the GenBank nucleotide database under the accession number DQ916828.

## 3.3 Results and Discussion

### 3.3.1 Isolation of the *C. bombicola URA3* gene

The *URA3* genes of several yeast species of the subphylum *Saccharomycotina* were matched together in a multiple alignment, revealing several conserved regions. Those regions correspond to the conserved amino acid sequences described by Rodriguez *et al.* (1998). The first degenerate primer pair ura3For1 and ura3Rev1 was designed on the highly conserved FEDRKFADIG (amino acids 87 till 96, Figure III-5) and GD-LGQQYR-V regions (amino acids 206 till 216) (Francois *et al.*, 2004).

```
   1  cgacggcccg rgctggtctg acgggcggat agtacaggct ttgccaaaag cctataaggc taaagaaagt
  71  aaacaagtga ggttgaacca tgatggcagt gttcgaattc tgatcaatga rgtacactgc gaagggaatc
   1                          Hypothetical protein →  M   X   Y   T   A   K   G   I

 141  cccgaaacgg cgaacaaaaa gaacatcaga ggaggaacgc cctcgcaatc ccgaacatac cagtttcgca
   9    P   E   T   A   N   K   K   N   I   R   G   G   T   P   S   Q   S   R   T   Y   Q   F   R

 211  gaacctgggg tatcaactgg atgcaccagc atactgttcc cactgttgcc aatgctgtag acgctccatt
  32   R   T   W   G   I   N   W   M   H   Q   H   T   V   P   T   V   A   N   A   V   D   A   P

 281  gttgtcagtc attttagcat tttacagtaa ccaactccaa aaaacagccc gctctgctgg gaagacttcg
  55  L   L   S   V   I   L   A   F   Y   S   N   Q   L   Q   K   T   A   R   S   A   G   K   T   S

 351  caattattta tccactactg ctgcggttat atacttctcg atctcagtct cggttataat tgccgcttga
  79    Q   L   F   I   H   Y   C   C   G   Y   I   L   L   D   L   S   L   G   Y   N   C   R   L

 421  cagcctggag aaattcggat actccacgtg ataattgcca tagggcataa ttttcgaaac agctcgcaac
 102   T   A   W   R   N   S   D   T   P   R   D   N   C   H   R   A   *

 491  gatctcggct agttttcccc ttttttgacc catatcgacg ctgagactca ctcacttgat gcctaccgtt
 561  agggtaaatt tttcaagcct gcagaatatc gcgggacgca gtctcctgca cgcgcgtgac ttcatcttac
 631  ttacatcaaa cagcccgatt aatttgaaaa gtcctagctg atcgagggca cgggcactac tgtagagaaa
 701  taatatgaag ctgagctatg aggagcgccg agagaggctg ccggctgtag cagcccggct attcgacatc
   1  URA3: M   K   L   S   Y   E   E   R   R   E   R   L   P   A   V   A   A   R   L   F   D   I

 771  attgtgagca agcaaacaaa tctttgcgca agcttggatg tgcgaactac ctctgagtta ctgagtatcc
  23    I   V   S   K   Q   T   N   L   C   A   S   L   D   V   R   T   T   S   E   L   L   S   I

 841  tggaccgcat tggaccttac atttgtatgg ttaagaccca cattgacata attgacgact tcgaatacga
  46   L   D   R   I   G   P   Y   I   C   M   V   K   T   H   I   D   I   I   D   D   F   E   Y

 911  cacaactgtc agcggtttga aacagctttc aacgaagcac aattttctca tttttgaaga ccgaaagttc
  69  D   T   T   V   S   G   L   K   Q   L   S   T   K   H   N   F   L   I  F   E   D   R   K   F

 981  gcagacatcg gttccactgt taaggcccaa tatgcaggtg gagtgtttaa gatcgctcaa tgggctgata
  93   A   D   I   G  S   T   V   K   A   Q   Y   A   G   G   V   F   K   I   A   Q   W   A   D

1051  taacaaatgc tcacggtgtt cctgggccgg gaattgtgag cggactagaa gaggctgcga aggaaactac
 116   I   T   N   A   H   G   V   P   G   P   G   I   V   S   G   L   E   E   A   A   K   E   T

1121  ggatgaacct cgcggccttg tcatgcttgc agaactgagt tcgaagggca cactggctca cggcgaatac
 139  T   D   E   P   R   G   L   V   M   L   A   E   L   S   S   K   G   T   L   A   H   G   E   Y

1191  tcgcaagcga cagtagacat cgctcgcagt aaccgcgcat ttgtgtttgg tttcatcgct cagcaaaaag
 163    S   Q   A   T   V   D   I   A   R   S   N   R   A   F   V   F   G   F   I   A   Q   Q   K

1261  tcggaaagcc agaggaagac tgggtcatta tgactcctgg ggtgggcctg gacgacaaag gtgatggatt
 186   V   G   K   P   E   E   D   W   V   I   M   T   P   G   V   G   L   D   D   K  G   D   G

1331  ggggcagcag tatcgtactg tggacgacgt catagagacc ggcacagacg ttattatcgt cggacgcggg
 209 L   G   Q   Q   Y   R   T   V  D   D   V   I   E   T   G   T  D   V   I   I   V   G   R   G

1401  ctctatagca agggacgaga tcctgtgcac gaagctcagc gttaccaaaa ggcgggctgg aatgcatatc
 233   L  Y   S   K   G   R   D   P   V   H   E   A   Q   R   Y   Q   K   A   G   W   N   A   Y

1471  tgagaaaagt tcagtcaaga tgattttctc aaacagttcc ttcaatgcaa cttgcacatg aatacctata
 256   L   R   K   V   Q   S   R   *

1541  aaatctgatt aaattaccat aaaaggtaca gattaaaata tatatgcctt caatggcatc cttcgcgatt
1611  ctgattcgtc agcacacttc aaccttccta ctatgagtga cagtgatgat gatctgctgg cattggccga
   1                  Putative Rtf1p -->  M   S   D   S   D   D   D   L   L   A   L   A

1681  cgttggctcc gactccgaag aggaaatctc gcrgccgtcg ccgccaagca atgaggtcgt caatccctat
  13  D   V   G   S   D   S   E   E   E   I   S   X   P   S   P   P   S   N   E   V   V   N   P   Y

1751  cctctagaag gcaaatatct cgatgctgaa gacagggcga agttggacgc gctgccagag attgagcgag
  37    P   L   E   G   K   Y   L   D   A   E   D   R   A   K   L   D   A   L   P   E   I   E   R

1821  aagagatctt gtatgaccga gctcaggaga tgcagcggta cgaggagaga aggtatcttg ctcagcgaag
  60   E   E   I   L   Y   D   R   A   Q   E   M   Q   R   Y   E   E   R   R   Y   L   A   Q   R

1891  gaagcagatg acgcgggttg ctgacgagga cgaagccccc yccgccaagc gtcaacgggg tacaacaggc
  83  R   K   Q   M   T   R   V   A   D   E   D   E   A   P   X   A   K   R   Q   R   G   T   T   G
```

**Figure III-5:** Nucleotide and deduced amino acid sequence of the *URA3* gene of *C. bombicola*. The deduced amino acid sequences of the putative and partial protein are also indicated. Possible promoter elements are underlined and sequences presumed to be involved in polyadenylation are marked in bold. The conserved amino acid sequences of the OMPD as described by Francois *et al.* (2004) are marked as bold and underlined cases.

An amplicon of about 400 bp was detected, but the signal was very weak and attempts to isolate and clone the PCR product or to submit it to re-amplification did not succeed. For this reason, the second degenerate primer pair ura3For2 and ura3Rev2 was designed in such a way that they also could be used as a nested primer pair on the weak amplification product. Ura3For2 is located in the same conserved area as ura3For1, but ura3Rev2 is based on the conserved (M/L)TPGVG region (amino acids 196 till 201). The degenerate PCR with ura3For2 and ura3Rev2 was performed both on genomic *C. bombicola* DNA and the weak amplification product of ura3For1 and ura3Rev1. The result is shown in Figure III-6. The use of the genomic DNA as template gave rise to a weak amplicon of about 330 bp (the expected length), but also stronger, non-specific amplification was observed. However, when the amplicon of ura3For1 and ura3Rev1 was used as starting DNA, a major nested PCR product was observed at the expected length. The DNA sequence of this later fragment was blasted against the NCBI databases and showed a 75 % amino acid identity with the orotidine-5'-phosphate decarboxylases of *Candida glycerinogenes* and *Pichia fabiani*.

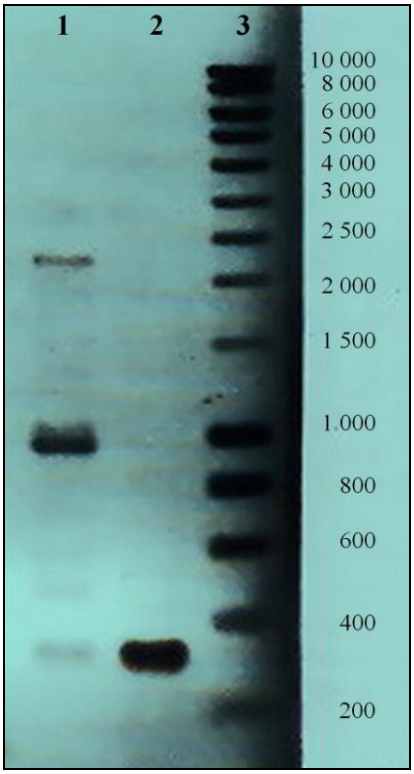


**Figure III-6:** Degenerate PCR with the ura3For2 and ura3Rev2 primers on genomic DNA (1) and on the PCR product of the ura3For1 and ura3Rev1 primers (2). SmartLadder (Eurogentec) marker in lane (3), the numbers on the right-hand side correspond to the length of its fragments.

The regions adjacent to the obtained fragment were cloned using the Universal GenomeWalker kit™. As a result, a DNA sequence of 2074 bp was obtained, carrying the total coding sequence of 789 bp and its flanking upstream and downstream regions of 704 and 581 bp respectively.

### 3.3.2 Characterisation of the *URA3* sequence

As for most yeast *URA3* genes, there were no indications for the presence of an intron. The *URA3* open reading frame is translated in a protein of 262 amino acids with a calculated size of 28.97 kDa and an estimated pI value of 5.8. The protein is most identical to the orotidine-5'-phosphate decarboxylase of *Candida glycerinogenes* (69 % amino acid identity), and also for the majority of the other yeast species high identity scores, ranging from 65 to 68 %, were observed.

There was no conventional TATA-box found in the 5' region. This is however not unusual for yeast genes as only 20 % of the promoters are supposed to be regulated by TATA-box elements (Basehoar *et al.*, 2004). At position -57 till -45, an AT-rich sequence is found, which possibly takes part in transcription initiation. Highly expressed genes of *S. cerevisiae* show no G-residues in the 7 bases preceding the start codon (Hamilton *et al.*, 1987) and the same is observed for this *C. bombicola URA3* gene.

The consensus sequence of the polyadenylation signal found in most eukaryotic genes is AATAAA (Guo and Sherman, 1996), although for yeasts several variations are possible (Kozak, 1991). In the 3' region of the *URA3* gene of *C. bombicola*, three putative sequences were observed 45 bp (ATAAAA), 66 bp (again ATAAAA) and 82 bp (AAAATA) downstream of the TGA stop codon.

The 32 other known yeast amino acid sequences of orotidine-5'-phosphate decarboxylase found in the GenBank database were used to construct a phylogenetic tree based on the protein maximum likelihood principle (Figure III-7). The sequence of the basidiomycete *Ustilago maydis* was used as outgroup and the tree was rooted against it. The Ln Likelihood of the tree is -8165 and all knobs and branches were statistically significant ($p < 0.05$). The OMPD of *Candida bombicola* clustered together with the OMPD of *Candida glycerinogenes*, but considering the length of the branches, there still are quite some differences among them.

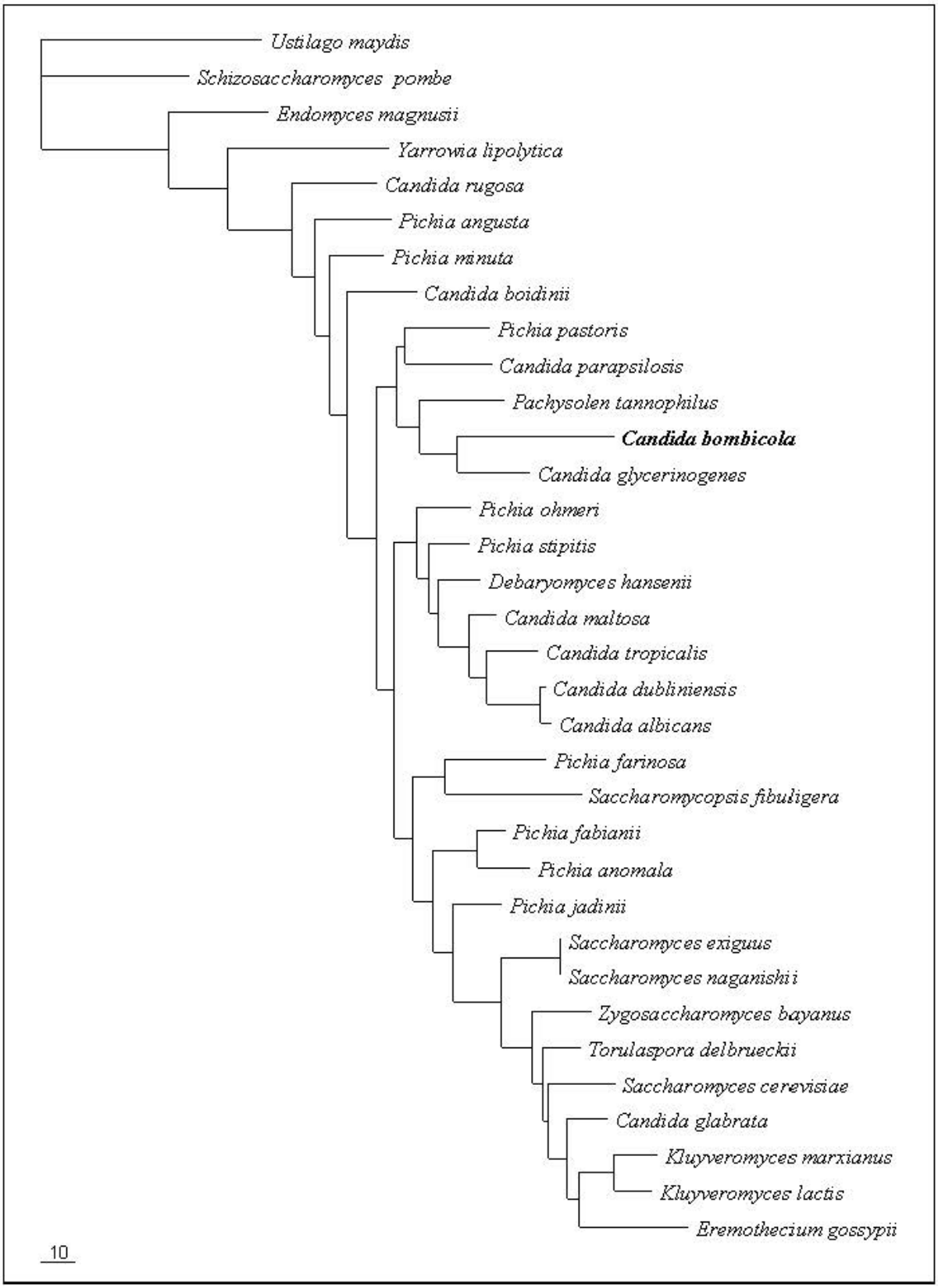


**Figure III-7:** Phylogenetic tree for yeast orotidine-5'-phosphate decarboxylase rooted against the *Ustilago maydis* OMPD. The marker bar below denotes the integer branch length.

### 3.3.3 Other putative genes flanking *URA3*

In the 581 bp region downstream the *URA3* coding sequence, a possible translation start and an open reading frame encoding the first 144 amino acids of the putative RNA polymerase II transcription elongation factor protein (RTF1P) was found. This is a transcription factor involved in TATA site selection and elongation by RNA polymerase II. A clear TATA-box element was found 64 bp upstream the putative start codon. The TATATATG sequence corresponds well with the proposed consensus sequence TATA(T/A)A(T/A)(A/G) of Basehoar *et al.* (2004). Furthermore, a CAAT sequence is present 52 bp upstream the ATG-codon.

Also in the 5' upstream region, an additional open reading frame was detected. The start and ending are located at position 117 and 470, giving rise to a hypothetical protein of 117 amino acids. The function of this protein is unclear, but it is up to 60 % identical (amino acid based) to other hypothetical proteins of several yeast species.

### 3.3.4 Complementation of an *URA3* deficient *S. cerevisiae* strain

In order to verify the functionality of the proposed *URA3* gene product, the pCbUra3 vector harbouring 2061 bp of the known sequence was constructed as described in the Materials and Methods section 3.2.6. pCbUra3 was used to transform the *URA3* deficient *S. cerevisiae* FY1679-01B (*MATa*; *ura3-52*; *LEU2*; *TRP1*; *HIS3*; *GAL2*). As a negative control, the same yeast strain was transformed with the pGEM-T® vector. Only the transformants with the plasmid pCbUra3 could grow on SD-medium without uracil, this led us to conclude that the *URA3* gene of *C. bombicola* encodes the orotidine-5'-phosphate decarboxylase.

# 4 Construction of an ura3$^-$ *C. bombicola* and a transformation system based on the *URA3* auxotrophic marker

## 4.1 Introduction

Previous Section 3 describes the isolation of the *URA3* gene of *C. bombicola*. This is one of the steps in creating a transformation and selection system based on the ura3 auxotrophic marker. The next step is screening for and selection of an *URA3* auxotrophic *C. bombicola* strain.

Those mutants can be selected on media containing 5-fluoroorotic acid (5-FOA). 5-FOA is toxic for wild type cells; they have a functional orotidine 5'-phosphate decarboxylase (URA3) in the *de novo* pyrimidine synthetic pathway and convert 5-FOA to one or more toxic intermediates, while 5-FOA does not affect cells with a dysfunctional *URA3* gene (Boeke *et al*., 1984). In most cases, spontaneous mutants can arise, but sometimes mutagenic agents such as N-methyl-N'-nitro-N-nitrosoguanidine (Sakai *et al.*, 1991), UV-treatment or techniques for directed disruption of the gene with a mutated sequence are required (Goosen *et al.*, 1987; Hara *et al.*, 2001). The latter techniques can be helpful when the occurrence of spontaneous mutants is low, when they are not stable (*e.g.* the mutation is the result of a single point mutation and revertation to the wild type occurs at relatively high frequence), or when working with di- or polyploid strains (Hara *et al.*, 2001).

Once a stable mutant is obtained, it can be transformed back to prototrophy with the wild type *URA3* gene either introduced on a plasmid with an ARS or by integrating the gene into the genome of the auxotrophic mutant.

## 4.2 Materials and Methods

### 4.2.1 Strains and culture conditions

*C. bombicola* ATCC 22214 was used as the parent strain and *Escherichia coli* DH5α was used for plasmid maintenance.

*C. bombicola* was cultured on yeast peptone dextrose (YPD) medium (1 % yeast extract, 2 % peptone, 2 % glucose) or on synthetic dextrose (SD) medium (0.67 % yeast nitrogen base without amino acids (DIFCO) and 2 % glucose). To obtain better growth of the mutants, 0.03

% uracil was added to the YPD media. The SD medium used for the selection of auxotrophic mutants was supplemented with 0.03 % uracil and 0.1 % 5-fluoroorotic acid (5-FOA). Liquid yeast shake flask cultures were incubated at 30 °C and 200 rpm. *E. coli* was grown as described in Section 2.2.2.

### 4.2.2 PCR and sequence analysis

The *URA3* genes of the selected mutants were amplified by yeast colony PCR (see 2.2.4) with a specific annealing temperature of 55 °C. The primers ura3genFor and ura3genRev (Table III-5) were used for this purpose.

All DNA sequences were determined at the VIB Genetic Service Facility (Belgium).

**Table III-5:** Primers used for the amplification of the *C. bombicola URA3* gene. All primers were obtained from Sigma Genosys.

| Name | Feature | Sequence |
|---|---|---|
| ura3genFor | amplification CDS *URA3* | 5'-GTT GAA CCA TGA TGG CAG TGT TCG-3' |
| ura3genRev | amplification CDS *URA3* | 5'-TGC TTC CTT CGC TGA GCA AGA TAC-3' |
| ura3TotFor | amplification total *URA3* | 5'-GCT GGT CTG ACG GGC GGA TAG TAC A-3' |
| ura3TotRev | amplification total *URA3* | 5'-CAT CAT CGT CAC TAT ACA CAT CGT CAT-3' |
| ura3wtRev | binds strictly on wild type *URA3* | 5'-CCG ACT TTT TGC TGA GCG A-3' |
| ura3G9Rev | binds strictly on the G9 *URA3* | 5'-TCC GAC TTT TTG CTG AAA C-3' |

### 4.2.3 Plasmids and transformation

The construction of vector pCbUra3 containing the functional *URA3* gene of *C. bombicola* is described in 3.2.6. This vector was used for transformation experiments with circular plasmids. Plasmid DNA was isolated with the QIAprep Spin Miniprep Kit (Qiagen). Linear DNA for transformation was obtained by standard PCR with the primers ura3TotFor and ura3TotRev (Table III-5) and pCbUra3 as template. The PCR reaction mixture was prepared for transformation by using the GenElute™ PCR CleanUp Kit (Sigma).

*C. bombicola* cells were transformed as described in 2.3.3.

The yeast colony PCR discriminating for the wild type and mutated *URA3* gene was carried out with primers ura3TotFor and ura3wtRev for amplification of the wild type gene and

ura3TotFor and ura3G9Rev for the G9 mutated gene (Table III-5). The temperature program is identical to the one in 2.2.4, but the specific annealing temperature is 66 °C for the wild type primers and 60 °C for the G9 primers.

## 4.3 Results and Discussion

### 4.3.1 Isolation and selection of ura3-auxotrophic mutants

In order to develop an *URA3* based transformation system, the *URA3* gene of *C. bombicola* (GenBank accession number DQ916828) was isolated and tested for functionality (see Section 3 of this chapter). The gene encodes for an enzyme of 262 amino acids and shows high homology with the known orotidine-5'-phosphate decarboxylases of several other yeast species. The functionality of the gene was proven by complementation of an ura3-negative *Saccharomyces cerevisiae* strain.

Next step in the development of the transformation system is the isolation of a stable ura3-auxotrophic mutant. Although there are no direct data available, it is believed that *C. bombicola* is a haploid yeast species. Mating experiments with *Starmerella bombicola*, the teleomorph of *C. bombicola*, tend to support this hypothesis (Rosa and Lachance, 1998). For this reason, it is quite likely that spontaneous ura3-auxotropic mutants can be found. As stated in the introduction part, those mutants can be selected on media containing 5-FOA. Similar to most yeast species, the growth of wild type *C. bombicola* was inhibited at a concentration of 1 g/L 5-FOA. *C. bombicola* cells from an old agar plate were suspended and washed twice with physiological solution and plated onto SD supplemented with uracil and 5-FOA. It took sometimes more than two weeks before spontaneous mutants arose. A total of 33 colonies were picked up and their phenotype was further evaluated by growing them on YPD, SD and SD supplemented with uracil and 5-FOA. 17 mutants with the correct phenotype and normal growth on YPD were further tested as to their ability to maintain the uracil auxotrophy in rich medium: nine of them reverted to uracil prototrophy at low frequencies ($<1.3\ 10^{-8}$). The coding region of the *URA3* gene of those nine mutants was amplified by yeast colony PCR with the primers ura3genFor en ura3genRev (Table III-5) and sequenced. In six genes, no mutation could be detected in the coding region; it is nevertheless possible that a mutation arose in the regulatory regions, in regulating proteins or in other enzymes involved in the *de novo* pyrimide synthesis pathway. However, three *URA3* genes were clearly dysfunctional:

the corresponding mutants were C3, F4 and G9 (Figure III-8). To exclude the possibility of PCR or sequencing errors, the colony PCR and subsequent sequencing were repeated. The results were confirmed for all three mutants. C3 shows a point mutation at bp 760, changing the glutamine encoding CAA into a TAA stop codon and in this way creating a truncated URA3 protein of 247 amino acids instead of 262. Also the malfunction of the F4 *URA3* gene is caused by a single point mutation. This time, the highly conserved FEDRKFADIG sequence (amino acids 87-96; Figure III-8) is affected (Francois *et al*., 2004). The central lysine residue is believed to play a key role in enzyme activity and comparison between the *URA3* genes of several organisms of different phyla indeed brings the conserved region down to tree essential amino acids: -D-K--D- (Kimsey and Kaiser, 1992; Rodriguez *et al*., 1998). In F4, the AAG codon is changed to ACG, mutating the central lysine to a threonine residue. The dramatic effect of this single point mutation confirms the importance of the lysine for enzymatic activity. The *URA3* gene of G9 on the other hand contains no point mutations, but a total deletion of seven subsequent nucleotides at bp positions 542 to 549 of the coding region.

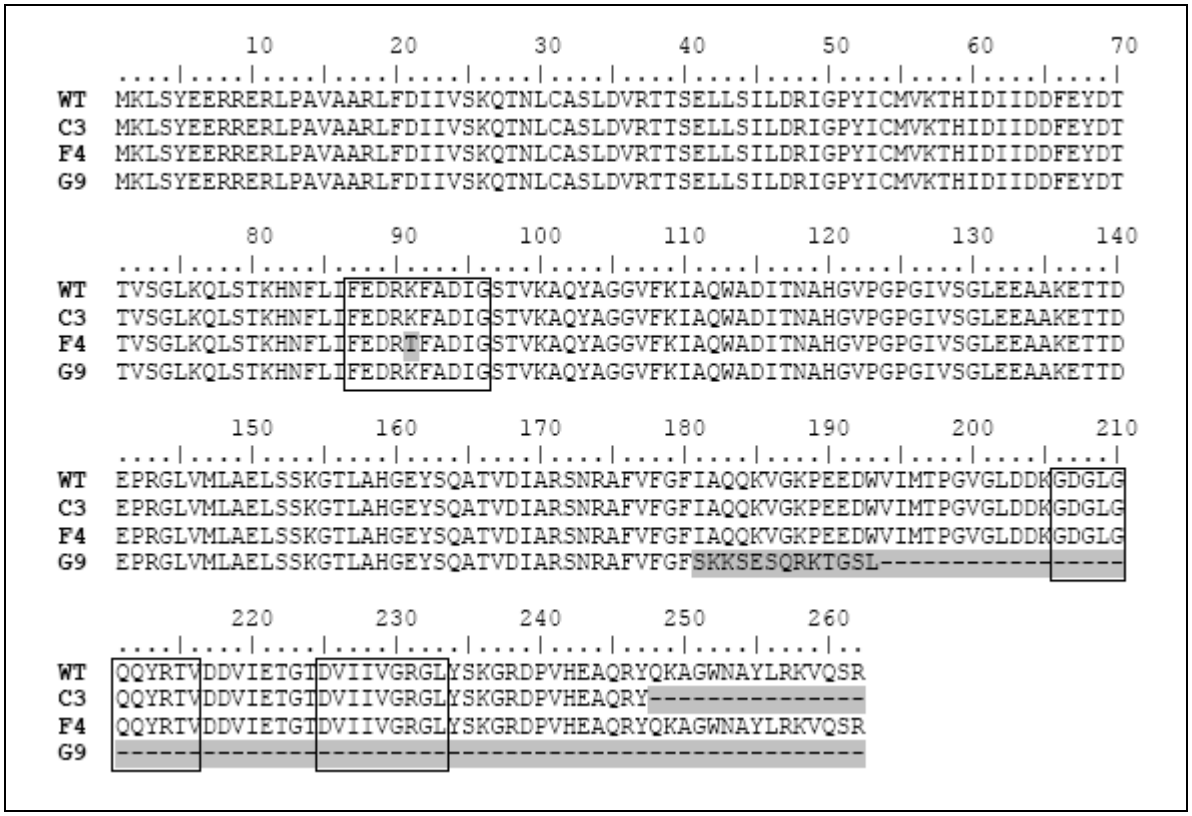


**Figure III-8:** Alignment of the *C. bombicola URA3* wild type and mutated genes. Highly conserved regions are marked by a box (Francois *et al*., 2004).

This causes a translational frame shift, creating a protein with altered amino acids starting from residue 181 which is in addition no longer than 193 amino acids. Although the revertant frequency for al mutants was very low, G9 was chosen to conduct further transformation experiments because the chances on revertation to the wild type are practically non-existing and because the discrimination between the G9 and inserted WT *URA3* gene in the genome can be monitored by PCR-based techniques.

### 4.3.2 Transformation of *C. bombicola* G9 with the *URA3* gene

Vector pCbUra3 carries the functional *URA3* gene of *C. bombicola* consisting of the coding sequence of 789 bp and a upstream and downstream region of 693 and 579 bp respectively. In initial experiments, circular plasmid DNA was used to transform *C. bombicola* G9. When 0.1 µg of DNA was used, 1 to 5 prototrophs were obtained per transformation reaction (10 to 50 transformants per µg DNA). In an attempt to increase this low number of transformants, higher amounts of plasmid DNA were used (up to 2.5 µg), but this did not result in more prototrophic colonies.

Since the plasmid does not contain any yeast autonomously replicating sequence, functional expression of the *URA3* gene must occur after integration into the genome. In several yeast species, this event is more likely to occur with linear DNA fragments (Sakai *et al*., 1991; Yang *et al*., 1994). A linear *URA3* fragment of 2055 bp was created by PCR with the primers ura3intFor and ura3intRev, and pCbUra3 as a template. When 2.5 µg of linear DNA was used, 43 to 46 colonies appeared on the minimal medium (17 to 18 transformants per µg DNA). Although the number of transformants per µg DNA is lower compared too the plasmid DNA, the absolute number of obtained transformants is one order of magnitude higher. In both experiments with circular and linear DNA, no colonies were observed when the yeast cells were transformed with an equal volume of sterile water.

The insertion of the wild type *URA3* gene into the *C. bombicola* G9 genome was checked by yeast colony PCR. The primers ura3TotFor and ura3wtRev specifically amplify part of the wild type gene, while the combination ura3TotFor and ura3G9Rev is specific for the G9 mutated gene (Table III-5). The reverse primers are designed in such a way that their 3' end is unique for either the wild type or G9 *URA3* gene (Figure III-9). The PCR-program used for amplification is described in the Materials and Methods section.

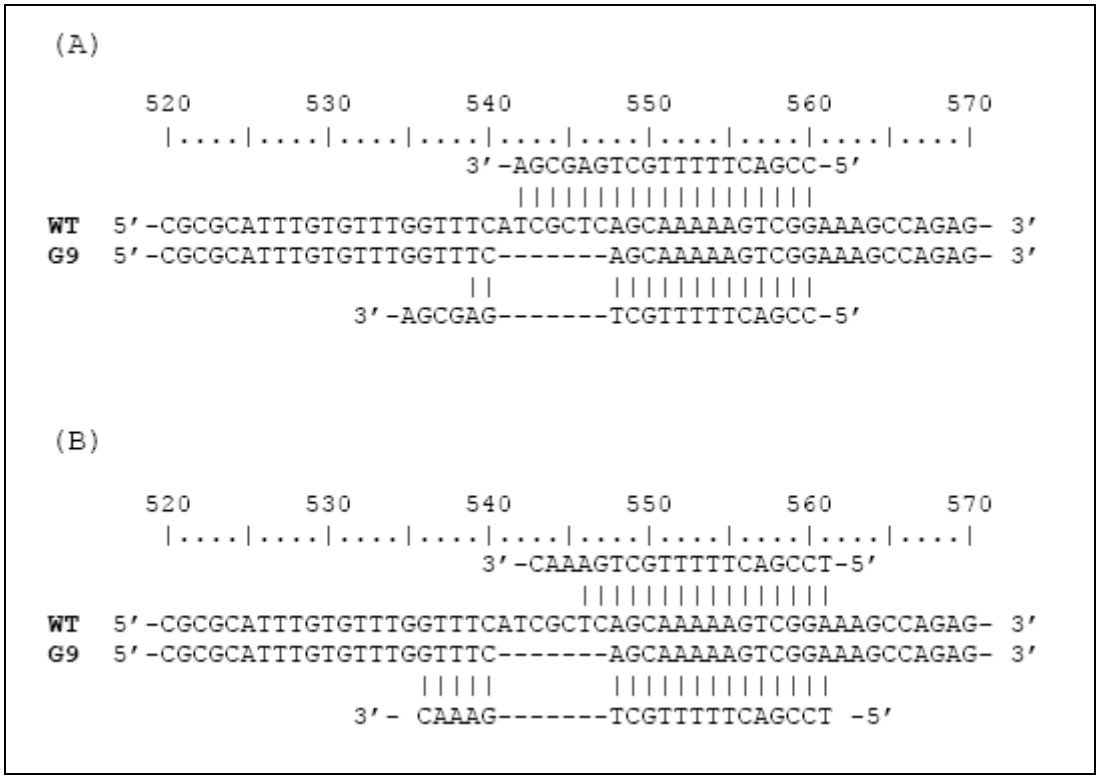


**Figure III-9:** Presentation of binding of the reverse primers used for PCR-based discrimination between the wild type and G9 *URA3* gene. (A) Annealing of the primer ura3wtRev, which amplifies the wild type gene. (B). Annealing of the primer ura3G9Rev, which amplifies the G9 gene.

68 randomly selected prototrophs obtained after transformation with linear DNA were tested (colonies coming from two transformations). Five tested colonies turned out to be false positive: they only contained the G9 gene. All other 63 transformants regained the wild type *URA3* gene. For most transformants, the mutated gene was still present; indicating that either non-homologous recombination or homologous recombination with single crossing-over took place. For 12 transformants however, no mutated gene could be detected, suggesting that a double crossing-over event occurred resulting in gene replacement. As an example, the results of the first six transformants can be seen in Figure III-10.

Those findings demonstrate that prototrophy was restored due to the integration of the wild type *URA3* gene in the *C. bombicola* G9 genome. In 19 % of the cases, this was due to a gene replacement event.

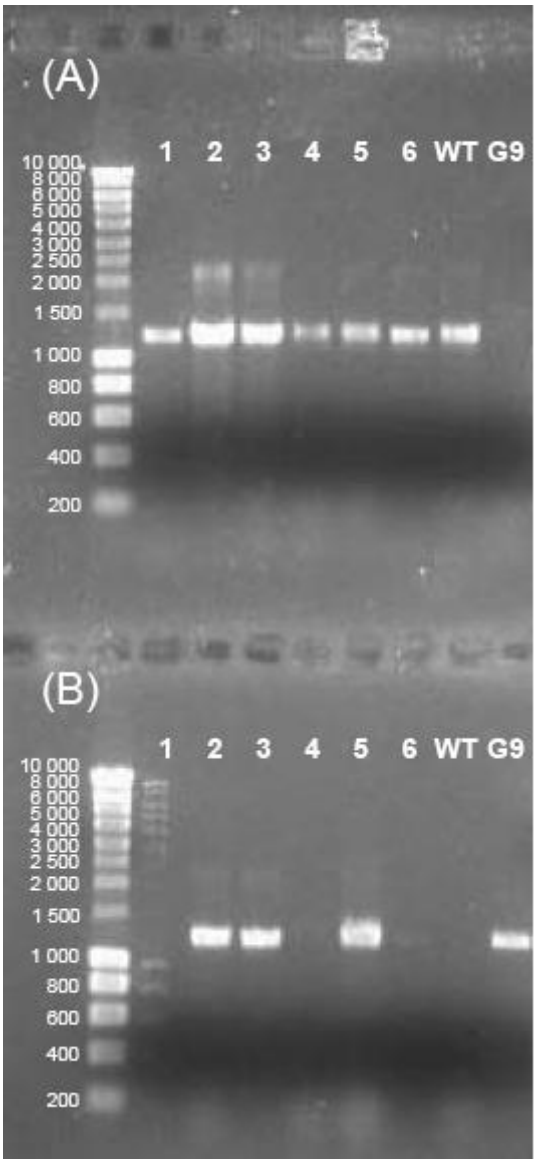


**Figure III-10:** Results of the wild type – G9 discriminating PCR. (A) Amplification with the wild type specific primers on transformants 1 to 6 and *C. bombicola* wild type and G9 as positive and negative control respectively. (B) Amplification with the G9 specific primers on transformants 1 to 6 and *C. bombicola* wild type and G9 as negative and positive control respectively.

# 5 Conclusion

The yeast *C. bombicola* produces sophorolipids in large amounts, which makes it an interesting organism to study. Most research has however been performed towards the fermentation conditions and parameters in order to improve the yield and production process, while hardly any attention has been spent to the elucidation of enzymes and/or genes involved

in sophorolipid synthesis pathway or the genetic engineering of *C. bombicola*. The genome of *C. bombicola*, for example, is completely unknown.

We decided to develop a transformation and selection system for *C. bombicola* which should allow us to perform some simple genetic engineering of this yeast and which could help us in studying the sophorolipid synthesis pathway.

A chemical transformation protocol with LiAc for *S. cerevisiae* was slightly modified for *C. bombicola*. The protocol was tested with dominant drug resistant markers. When testing the resistance of non transformed cells to the antibiotics G418, zeocin and hygromycin, it became clear that *C. bombicola* is very resistant towards those antibiotics (especially G418). For all of them, high concentrations must be used as compared to other yeasts and upon long incubation times, cells often succeed to grow after all. There is no clear explanation for this fact, but it could be that the produced sophorolipids protect the cell against the antibiotics by preventing them from entering the cell; many antibiotics are quite hydrophobic and can be absorbed in the sophorolipids surrounding the cells.

When transforming *C. bombicola* with a simple linear fragment for integration into the genome (the hygromycin resistance gene and the *TK* promoter and terminator), only 2.2 transformants per µg DNA were obtained. This low number can be explained by the fact that the fragment must be integrated by illegitimate recombination. When homologous recombination is possible, the number of transformants is almost tenfold higher. This is illustrated in Section 4 of this chapter when a linear fragment of the *URA3* gene was used to complement an ura3 auxotrophic mutant.

Since a selection system based on dominant drug resistant markers is expensive and not very convenient for large scale industrial fermentations, most yeast transformation systems are based on auxotrophic markers. Such a system would also allow us to improve the transformation efficiency because of the possibility of homologous recombination and the availability of more than one selection marker can of course become handy when multiple manipulations are desired. In yeasts and fungi, the system based on the orotidine 5'-phosphate decarboxylase (*URA3*) gene is used very often. This system offers the advantage that both positive and negative selection can be applied; a characteristic which is particularly useful for the creation and selection of ura3 auxotrophic strains and for introduction of multiple disruptions in the genome. In order to develop such a system, the *URA3* gene of *C. bombicola* was isolated (GenBank accession number DQ916828) and tested for functionality.

Next step in the development of the transformation and selection system was the isolation of a stable ura3 auxotrophic mutant which can be transformed back to prototrophy with the species-own *URA3* gene. Such mutants were found by selection on media containing 5-FOA. Several mutants were tested and evaluated on true phenotype, viability and revertation frequency. The mutant G9, missing 7 bp in the coding sequence of the *URA3* gene, was used to perform the transformation with the wild type *URA3* gene in order to restore the ura3 prototrophy. Successful transformation was confirmed by a PCR-based method discriminating between the wild type and mutated *URA3* gene. In 19 % of the transformants double cross-over resulting in gene replacement was observed.

The developed transformation system described in this chapter is a useful tool in the genetic engineering of *C. bombicola*. For example, one can knock out genes by transformation with the linear *URA3* gene flanked by regions of the target gene, and experiments regarding this strategy are described in Chapter V. Furthermore, one can (over)express certain genes in order to improve or alter sophorolipid production. However, a suitable promoter is needed to achieve this. It is possible that other yeast promoters, such as those from cytochrome c1 (*CYC1)* from *Saccharomyces cerevisiae* or from translation elongation factor 1 α (*TEF 1α*) from *Eremothecium gossypii* are also active in *C. bombicola*. On the other hand, species-own promoters generate in most cases better expression values. The housekeeping gene glyceraldehyde-3-phosphate dehydrogenase (*GAPD*) of *C. bombicola* has been cloned and sequenced, together with a 1613 bp upstream region (Chapter IV).

# Chapter IV: Cloning and characterisation of the *Candida bombicola* glyceraldehyde 3-phosphate dehydrogenase (*GAPD*) gene

Part of this chapter has been submitted for publication as:

Van Bogaert INA, De Maeseneire SL, Develter D, Soetaert W, Vandamme, EJ (2008) Cloning and characterisation of the glyceraldehyde 3-phosphate dehydrogenase gene of *Candida bombicola*. J Ind Microbiol Biotech, submitted

# 1 Introduction

Chapter III describes the development of a transformation and selection system for *C. bombicola*. One of the logical next steps in genetic engineering of this yeast species is the construction of an expression system. It was demonstrated in Chapter III that the *TEF1α* promoter of *Eremothecium gossypii* and the *TK* promoter of the *Herpes simplex* virus also display activity in *C. bombicola*. On the other hand, it is believed that homologous promoters give rise to better and higher expression levels.

Glyceraldehyde-3-phosphate dehydrogenase (GAPD, E.C. 1.2.1.12) is one of the key enzymes in the Embden-Meyerhof-Parnas or glycolysis pathway. It catalyzes the reversible oxidation and phosphorylation of glyceraldehyde-3-phosphate to produce 1,3-diphosphoglycerate. The enzyme contributes in this way to the formation of ATP and provides additional energy to the cell by reducing $NAD^+$ to NADH and $H^+$ upon its action. GAPD is a tetrameric enzyme and typically does not follow simple Michaelis-Menten kinetics (Harris and Waters, 1976). Since GAPD is a crucial enzyme functioning in a pathway which is essential in every living cell, its amino acid sequence is highly conserved. Meanwhile, *GAPD* sequences of a wide range of organisms have been determined, and are an important tool in phylogenetic analysis and protein evolution.

Due to its key function in glycolysis, GAPD can represent up to 5 % of the soluble cellular protein content of *Saccharomyces cerevisiae*, *Aspergillus nidulans* and other (higher) eukaryotic organisms (Piechaczyk *et al.*, 1984; Punt *et al.*, 1990). The abundance of the GAPD protein suggests that the *GAPD* gene is regulated by a constitutively and highly active promoter. The promoter sequences of native *GAPD* genes have been successfully applied for the expression of heterologous genes in several yeasts and filamentous fungi such as *S. cerevisiae* (Bitter and Egan, 1984), *Pichia pastoris* (Doring *et al.*, 1998) *Lentinula edodes* (Hirano *et al.*, 2000), *Mucor circinelloides* (Wolff and Arnaou, 2002) and *Flammulina velutipes* (Kuo *et al.*, 2004).

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

# 2 Materials and Methods

## 2.1 Strains, plasmids and culture conditions

*C. bombicola* ATCC 22214 was used for the preparation of genomic DNA (gDNA). *Escherichia coli* DH5α was used to maintain plasmids.

All PCR products were cloned into the pGEM-T® vector (Promega). The plasmid pREP4FLAG (LMBP 4343, Figure III-2) was obtained from the Laboratory of Molecular Biology, Ghent University (LMBP).

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

## 2.2 DNA isolation and sequencing

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

## 2.3 Transformation

*C. bombicola* cells were transformed by the lithium acetate method as described in Section 2.2.3 of Chapter III. *E. coli* cells were transformed as described by Sambrook and Russell (2001).

## 2.4 Degenerate PCR

Part of the *GAPD* gene of *C. bombicola* was amplified by the use of the degenerate primers GapdDegFor and GapdDegRev (Table IV-1) based on the conserved sequences of the *GAPD*

genes of the *Saccharomycotina*. PCR amplification was carried out with an initial denaturation of 94 °C for 4 min, a 40 fold repeat of 94 °C for 30 s, 50 °C for 1 min, 72 °C for 2 min with 5 s increasing time and a final 7 min elongation at 72 °C.

**Table IV-1 :** Primers used for the isolation of the *Candida bombicola GAPD* gene, creation of the Southern blot probe and the expression cassette. All primers were obtained from Sigma Genosys.

| Name | Feature | Sequence |
|---|---|---|
| GapdDegFor | degenerated primer | GTTYAARTAYGAYTCYACYCAYGG |
| GapdDegRev | degenerated primer | GAGTARCCRWAYTCRTTRTCRTACC |
| GapdUp1 | primary upstream GPS | ATCAGCGGAAGGAGCAGAGATCACAA |
| GapdUpN | nested upstream GPS | CCTTAGCCTTCTCGGTGGTGGTGAAGAC |
| GapdDown1 | primary downstream GPS | TCTGCGGTGACACTCACTCCTCTATCTAC |
| GapdDownN | nested downstream GPS | ACGGCAACTTCGTCAAGCTCATCTCC |
| GapdProbeFor | creation Southern blot probe | GACTATGCTGCCTACATGTTCAAG |
| GapdProbeRev | creation Southern blot probe | GAGTAGCCGTACTCGTTGTCGTACCA |
| HygroCdsFor | amplification hygromycin gene | TCTAGATTCTCGAGATGAAAAAGCCTGAACTCAC |
| HygroCdsRev | amplification hygromycin gene | TGAACAAACGACCCAACAC |
| GapdPromInfFor | amplification *GAPD* promoter | CCGCGGGATTTCTAGGACATCCGATGTGTAGTTAATCA |
| GapdPromInfRev | amplification *GAPD* promoter | GCTTTTTCATCTCGATTGTGTAGAGTTGTTTTTGTTG |
| GapdHygro783For | amplification expression cassette | CCAATGGCAGTGGCTTACCACTC |
| GapdHygro639For | amplification expression cassette | TTCGTGCGACGCGGCGTAATTC |
| GapdHygro488For | amplification expression cassette | TCAAGCACTGCGGCACTCCTA |
| GapdHygro394For | amplification expression cassette | GCTGGAGTCTCATCTGCAAGGTT |
| GapdHygro352For | amplification expression cassette | AATACGAGCAATCGAAGCCTTGG |
| GapdHygro190For | amplification expression cassette | CCGTCATTGCAGGGTGTGCGTCTA |
| GapdHygroRev | amplification expression cassette | ACCCAACACCCGTGCGTTT |

## 2.5 Genome Walking

The unknown genomic DNA sequences upstream and downstream of the degenerate PCR fragment were identified by genome walking as described in Section 3.2.5 of Chapter III. Four gene-specific primers (GSP) were designed and are listed in Table IV-1.

## 2.6 Yeast colony PCR

Yeast colony PCR was performed as described in Section 2.2.4 of Chapter III. A specific annealing temperature of 64 °C was used.

## 2.7 Sequence analysis

Sequences were analysed with the Clone Manager Professional Suite Software (Version 6.0). The BLAST program (Altschul *et al.*, 1997) was used for similarity searches in databases available on the NCBI website (http://www.ncbi.nlm.nih.gov). Multiple sequence alignments were made with the CLUSTAL W program (Higgins *et al.*, 1992). Subsequently, a phylogenetic tree was constructed in BioEdit using the Protein Maximum Likelihood (ProML) algorithm.

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

## 2.8 Southern blot

20 µg digested gDNA was blotted onto a positively charged nylon filter (Hybond N+, GE Healthcare) with the Trans-Blot® SD DNA/RNA Blotting Kit (Bio-Rad) according to the manufacturer's instructions. DNA was hybridised to a DIG labelled probe which was constructed with the DIG High Prime Labelling and Detection Kit (Roche Diagnostics). The applied hybridisation conditions allowed 18 % mismatch between probe and target. The initial unlabelled probe of 842 bp was obtained by means of the High Fidelity PCR Master kit (Roche Diagnostics) and the primers GapdProbeFor and GapdProbeRev (Table IV-1). PCR products were purified by using the GenElute™ PCR CleanUp Kit (Sigma). Restriction enzymes were obtained from Roche Diagnostics or New England Biolabs and were used as indicated by the manufacturer.

## 2.9 Construction of the GAPD-hygromycin resistance cassette

The potency of the *GAPD* promoter was tested by fusing it to a hygromycin B phosphotransferase gene. The hygromycin B phosphotransferase gene and the TK terminator were obtained by high fidelity PCR amplification with BD Advantage 2 Polymerase Mix (Clontech) and the primers HygroCdsFor and HygroCdsRev and pREP4FLAG as template. The forward primer HygroCdsFor was extended with the *Xba*I (TCTAGA) and *Xho*I (CTCGAG) restriction sites for convenient cloning of the *GAPD* promoter fragment (Table IV-1). The obtained fragment of the hygromycin B phosphotransferase gene and the TK terminator (1154 bp) was cloned into pGEM-T® (Promega). A vector with a T7 oriented insert was chosen to continue cloning. This vector was cut with the unique cutters *Xba*I and *Xho*I, located just upstream the hygromycin resistance gene, and the *GAPD* promoter of *C. bombicola* was inserted by means of the In-Fusion™ 2.0 Dry-Down PCR Cloning Kit (Clontech). This *GAPD* promoter was obtained by high fidelity PCR amplification with the BD Advantage 2 Polymerase Mix (Clontech) and the primers GapdPromInfFor and GapdPromInfRev, amplifying 1560 bp of the *GAPD* promoter. The obtained vector harbours the hygromycin resistance gene controlled by the *C. bombicola GAPD* promoter.

For obtaining linear hygromycin resistance cassettes under the control of a *GAPD* promoter with a length of 783, 639, 488, 394, 352 and 190 bp respectively, the primers GapdHygro783For, GapdHygro639For, GapdHygro488For, GapdHygro394For, GapdHygro352For and GapdHygro190For were combined with GapdHygroRev (Table IV-1).

## 2.10 Nucleotide sequence accession number

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

# 3 Results and Discussion

## 3.1 Isolation of the *C. bombicola GAPD* gene

Since *GAPD* genes are known to be highly conserved among species, it is quite easy to create good degenerate primers for the amplification of uncharacterized *GAPD* genes. The degenerate primers GapdDegFor and GapdDegRev (Table IV-1) were designed based on the conserved regions of the *GAPD* genes of the *Saccharomycotina* and generated a single fragment of approximately 820 bp, which was in the expected range of size. The nucleotide sequence of the fragment was determined and subsequent blasting against the NCBI databases revealed a 74 % amino acid identity with the glyceraldehyde-3-phosphate dehydrogenase of *Yarrowia lipolytica*.

The regions adjacent to the obtained fragment were cloned using the Universal GenomeWalker kit™. As a result, a DNA sequence of 3740 bp was obtained, carrying the total coding sequence of 1008 bp and an upstream and downstream region of 1613 and 783 bp respectively.

## 3.2 Characterisation of the *GAPD* sequence

Since the *GAPD* gene is a highly expressed gene, it can be expected that it has at least some promoter elements in its 5' region. Two possible CAAT boxes can be found 366 and 354 bp upstream the ATG start codon (Figure IV-1). No G-residues were found in the 7 bases preceding the start codon, which is a typical feature for highly expressed genes of *S. cerevisiae* (Hamilton *et al.*, 1987). Furthermore, a conventional TATA-box, TATATAAT, is present at position -78.

The polyadenylation signal in yeasts can differ from the signal found in eukaryotic genes (AATAAA), as described by Graber *et al.* (1999). The 3' region of the *C. bombicola GAPD* gene contained two possible sequences: ATAAAA 43 bp and AATATA 50 bp downstream the TAA stop codon.

```
   1  ccaatggcag tggcttacca ctccatcgag cctagtctaa ttaccataac gctaacaata
  61  ttcgacgggg ttccatgcac caaaagtcaa ggcagctaaa aaacaacatc ccagtatcgg
 121  gaccttcccc aggacagtcg agggttcgtg cgacgcggcg taattcctgg cagggcggta
 181  ttatttttca ggtgccacac gcgcattaat cgagcaagaa ggtacccttg gtagacgtgc
 241  atacatgcga gttgacgtcg taacgtcggc acgctcgtaa cgctcgcgcc ccctctcaag
 301  cactgcggca ctcctaccgt taggacttag gcttagggct atgcttgtgc gcgccgcgaa
 361  acggttagta taagcagtac gacgttgttg ctggagtctc atctgcaagg ttgagtacca
 421  atccctgccc caatacgagc aatcgaagcc ttggggaaag atgcggcggg ctagcttcag
 481  caataaatag caggcgacac acaaaaatta ggcggcaagc gcacgctcag catgccatct
 541  accagggcaa aaagcaaggc aacctctttt tcgcatcccg atttagagcc tacccgtcat
 601  tgcagggtgt gcgtctacga tataacacga tcgacatcgc gctgggtatg cttctgggta
 661  aggggtcgca acgtgtgagt tgtcagcact ggccgatacc caaagtatat aatgcgccgt
 721  tgaacggtta tagtcggtca agctcttaaa gaaagactta acaacaaaaa caactctaca
 781  caaatggctg gattcactgt cggtatcaac ggtttcggac gcattggtcg tctcgttctt
   1      M  A  G  F  T  V  G  I  N  G  F  G  R  I  G  R  L  V  L
 841  cgcaacgctc tcgagaacag caacgtcaac gtcgttgcca tcaacgatcc gttcatcgct
  20    R  N  A  L  E  N  S  N  V  N  V  V  A  I  N  D  P  F  I  A
 901  ccggactatg ctgcctacat gttcaagtac gactctactc acggccgctt caagggcgat
  40    P  D  Y  A  A  Y  M  F  K  Y  D  S  T  H  G  R  F  K  G  D
 961  gtccaggctg cctctgacgg tattgtcatc aacggcaaga aggttcttgt cttcaacgag
  60    V  Q  A  A  S  D  G  I  V  I  N  G  K  K  V  L  V  F  N  E
1021  aaggacccgg cccagatccc ctggggcaag agcggcgttt cttacgttgt tgagtccact
  80    K  D  P  A  Q  I  P  W  G  K  S  G  V  S  Y  V  V  E  S  T
1081  ggtgtcttca ccaccaccga gaaggctaag gctcacctcg ctggcggcgc caagcgtgtt
 100    G  V  F  T  T  T  E  K  A  K  A  H  L  A  G  G  A  K  R  V
1141  gtgatctctg ctccttccgc tgatgcccct atgtacgttg ttggtgtcaa ccttgagaag
 120    V  I  S  A  P  S  A  D  A  P  M  Y  V  V  G  V  N  L  E  K
1201  tactctccca aggaccagat cgtctccaac gcttcgtgca ccaccaactg ccttgctccc
 140    Y  S  P  K  D  Q  I  V  S  N  A  S  C  T  T  N  C  L  A  P
1261  ctcgccaagg tcatccacga cgagttcggt ctcgttgagg gtctcatgac tactgtccac
 160    L  A  K  V  I  H  D  E  F  G  L  V  E  G  L  M  T  T  V  H
1321  tcctacactg ccacccagaa gaccgttgac ggaccttccc acaaggactg gcgtggaggc
 180    S  Y  T  A  T  Q  K  T  V  D  G  P  S  H  K  D  W  R  G  G
1381  cgtactgccg ctgccaacat catcccgtct tcgactggtg ctgccaaggc tgtcggcaag
 200    R  T  A  A  A  N  I  I  P  S  S  T  G  A  A  K  A  V  G  K
1441  gtcatccctg agctcaacgg caagctcacc ggtatctctg tccgtgtccc tacccctgac
 220    V  I  P  E  L  N  G  K  L  T  G  I  S  V  R  V  P  T  P  D
1501  gtctctgttg tcgacctcac tgctcgtctc cagaagagcg ccaccattga ggagatcaac
 240    V  S  V  V  D  L  T  A  R  L  Q  K  S  A  T  I  E  E  I  N
1561  gctaccatca agaagtactc tgagggccag atgaagggtg tgctcggcta cactgacgag
 260    A  T  I  K  K  Y  S  E  G  Q  M  K  G  V  L  G  Y  T  D  E
1621  gacgtcgtct ccagcgactt ctgcggtgac actcactcct ctatctacga cgccaaggct
 280    D  V  V  S  S  D  F  C  G  D  T  H  S  S  I  Y  D  A  K  A
1681  tccctcgccc tcaacggcaa cttcgtcaag ctcatctcct ggtacgacaa cgagtacggc
 300    S  L  A  L  N  G  N  F  V  K  L  I  S  W  Y  D  N  E  Y  G
1741  tactctgccc gtgtcgtcga cttgcttgtc gctatcgcta agaaggacta aatttctcct
 320    Y  S  A  R  V  V  D  L  L  V  A  I  A  K  K  D  *  ← GAPD
1801  aataggctgt cagcgcatat gtgaggcgct catataaaac aatataaatc aaaacccatg
```

**Figure IV-1:** Nucleotide and deduced amino acid sequence of the *GAPD* gene of *C. bombicola*. Possible promoter elements are underlined and sequences presumed to be involved in polyadenylation are marked in bold cases. The conserved active site amino acids are marked as bold and underlined cases.

One feature often seen for highly expressed genes is a biased codon usage. This phenomenon is also observed for the *GAPD* gene of *C. bombicola*: there is a preference for use of a pyrimidine at the third position (80.35 %) and when the choice is allowed between a pyrimidine and a purine, a pyrimidine is chosen in 94.40 % of the cases. 24 of the 61 possible sense codons are unused (Table IV-2).

**Table IV-2:** Codon usage for the *C. bombicola GAPD* gene (336 codons). The absolute (#) and relative (%) numbers for each synonymous codon are given.

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

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

As to the other known genes of *C. bombicola*, there were no indications for the presence of an intron (Van Bogaert *et al.*, 2007a; Van Bogaert *et al.*, 2007b). The *GAPD* open reading frame is translated in a protein of 335 amino acids with a calculated size of 35.71 kDa and an estimated pI value of 7.67. The protein is most identical to the glyceraldehyde-3-phosphate dehydrogenase of *Yarrowia lipolytica* (74 % amino acid identity) and furthermore shows high homologies with GAPD's from both ascomycetous, basidiomycetous and zygomycetous fungi. The identities between the *C. bombicola* GAPD and those of lower and higher *eukaryota* are in general above 65 %, illustrating the high conservation grade of the GAPD sequence among organisms. Also for bacteria, identities between 59 and 68 % were observed. Figure IV-2 shows an alignment between the *C. bombicola* GAPD and representatives of different taxonomic classes. Figure IV-2 reveals that the enzyme contains the GAPD consensus substrate binding region found in all GAPD enzymes (PROSITE): [ASV]-S-C-[NT]-T-x(2)-(LIM). The cysteine residue is essential for the enzymatic activity since it functions as the binding site in the catalytic region (Fei *et al.*, 2006). This pattern is present in the *C. bombicola* protein as ASCTTNCL. Other conserved amino acids are a histidine, and two phosphate binding residues: lysine and arginine, in the *C. bombicola* sequence found at positions 179, 190 and 234 respectively, and indicated in Figure IV-2 (Van Wert and Yoder, 1992).

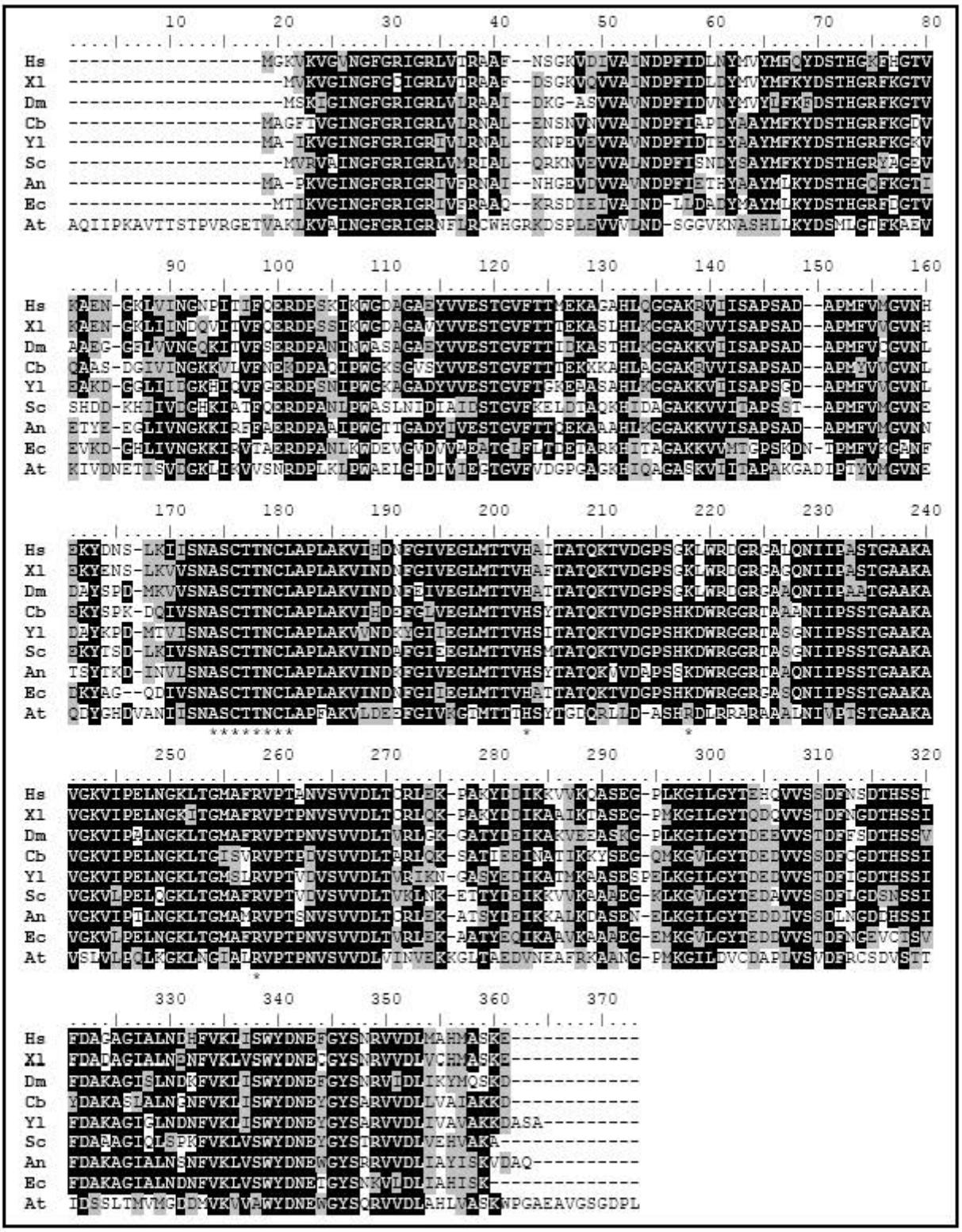

**Figure IV-2:** Alignment of GAPD sequences of organisms from different taxonomic classes: *Homo sapiens* (Hs), *Xenopus laevis* (Xl), *Drosophila melanogaster* (Dm), *C. bombicola* (Cb), *Y. lipolytica* (Yl), *S. cerevisiae* (Sc), *Aspergillus niger* (An), *E. coli* (Ec) and *Arabidopsis thaliana*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

## 3.3 Other putative genes flanking *GAPD*

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

## 3.4 Determination of the *GAPD* copy number

In most yeast genomes, only one *GAPD* gene is present. For some others, such as *S. cerevisiae*, two or three slightly different *GAPD* genes are differentially expressed, suggesting a distinct cellular role (McAlister and Holland, 1985). In order to verify this for *C. bombicola*, a Southern blot was performed. *C. bombicola* genomic DNA was cut with three different restriction enzymes: *Ava*I, *Eco*RI and *Xho*I. Only *Ava*I cuts twice into the known sequence, and binding of the probe should visualize a fragment of 1117 bp. All other enzymes do not cut the known sequence and accordingly should yield fragments larger than 3740 bp. The pattern of the Southern blot matched these predictions and no other bands were observed (Figure IV-3). These findings indicate that only a single copy of the *GAPD* gene is present in the genome of *C. bombicola* and therefore it is plausible that the cloned gene is a functional one.

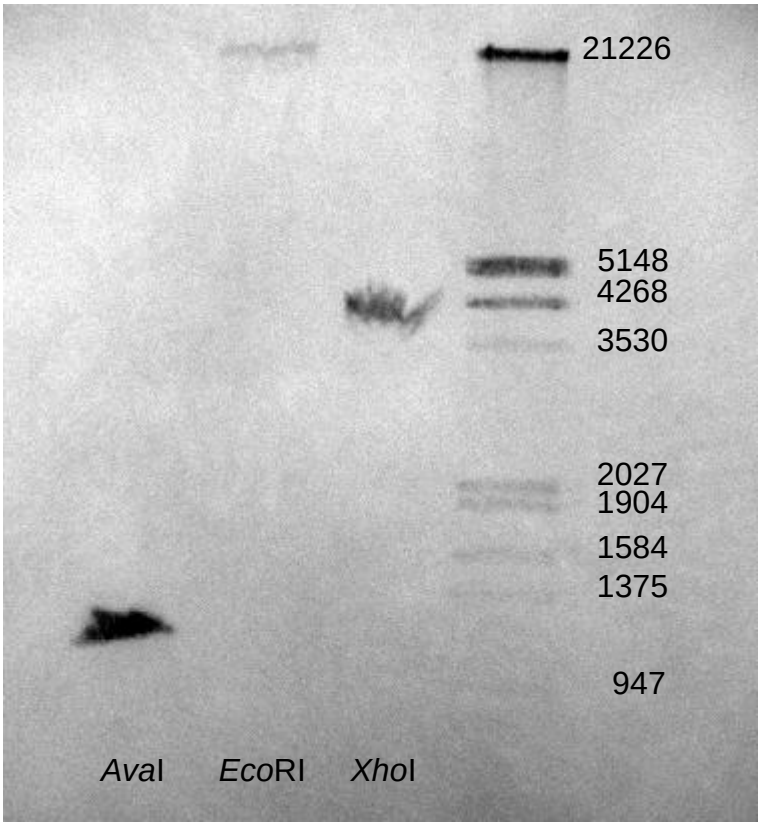


**Figure IV-3:** Southern blot with a *GAPD* probe based on the obtained sequence. *C. bombicola* gDNA was cut with *Ava*I, *Eco*RI and *Xho*I. The DIG-labelled DNA Moleclar-Weight Marker III from Roche was run alongside the gDNA. The size of its fragments are indicated on the right.

## 3.5 Evaluation of the *C. bombicola GAPD* promoter

The usefulness of the *GAPD* promoter for heterologous expression in *C. bombicola* was tested with the hygromycin resistance gene from the vector pREP4FLAG used in Chapter III (Figure III-2). An expression cassette containing the *GAPD* 5' region, the coding sequence of the hygromycin resistance gene (hygromycin phosphotransferase) and the *TK* terminator was constructed as described in the Materials and Methods section 2.9. Different promoter lengths were tested to find out whether a certain length should be respected in order to get sufficient expression. Promoter parts of 783, 639, 488, 394, 352 and 190 bp were tested. All fragments contained the TATA-box found at -78, and the fragments larger than 352 bp also possessed both CAAT-sequences located at positions -366 and -354. Consequently, the fragment of 352 bp harboured only one CAAT-sequence, while the 190 fragment had none.

Strangely enough, the expression cassettes with the longer *GAPD* promoter fragments of 783 and 639 bp did not give rise to any transformants, whereas the shorter ones, including the 190 bp fragment, still yielded transformants. This result could be explained by the fact that the larger homologous DNA fragments in the expression cassette lead to homologous recombination with the genomic *GAPD* promoter and consequently unable the *GAPD* gene. Since this gene is essential for cell viability, transformation was lethal to the cells. It has indeed been observed that homologous recombination with *C. bombicola* gDNA occurs most efficiently when the recombinant DNA is at least 500 and ideally 1000 bp long (Chapter V, Section 1.3.4). Further evidence for this hypothesis is the low transformation efficiency of about 2 colonies per µg DNA for the shorter fragments, which is comparable to those for transformations with the hygromycin resistance gene under control of the *TK* promoter as described in Chapter III, Section 2.3.3. There, no homologous sequences are present and consequently all transformants arose due to illegitimate recombination. It is thought that also for the transformants with the short *GAPD* promoter, illegitimate recombination occurred.
Another less plausible explanation could be the presence of a repressor binding site in the additional upstream regions which is removed for the shorter fragments.

The activity of the promoter fragment of 190 bp is somewhat surprising. For example, it has been shown that further upstream regions of the fungal *Lentinula edodes GAPD* promoter are essential for expression (Hirano *et al.*, 2000). Also for *S. cerevisiae*, larger promoter sequences are used for *GAPD* controlled expression. Yet, the most essential sequences of the human *GAPD* promoter are located in the region -181 to -144, and a promoter fragment of 181 bp gave an almost identical β-galactosidase activity as compared to a 489 bp fragment (Aki *et al.*, 1997).

The obtained transformants were tested for the presence of the transformation cassette into the genome by yeast colony PCR with the primers GapdHygro190For and hygroInsertCheckRev (Table IV-1 and Table III-1). All PCR reactions yielded the expected fragment of 1124 bp (Figure IV-4).

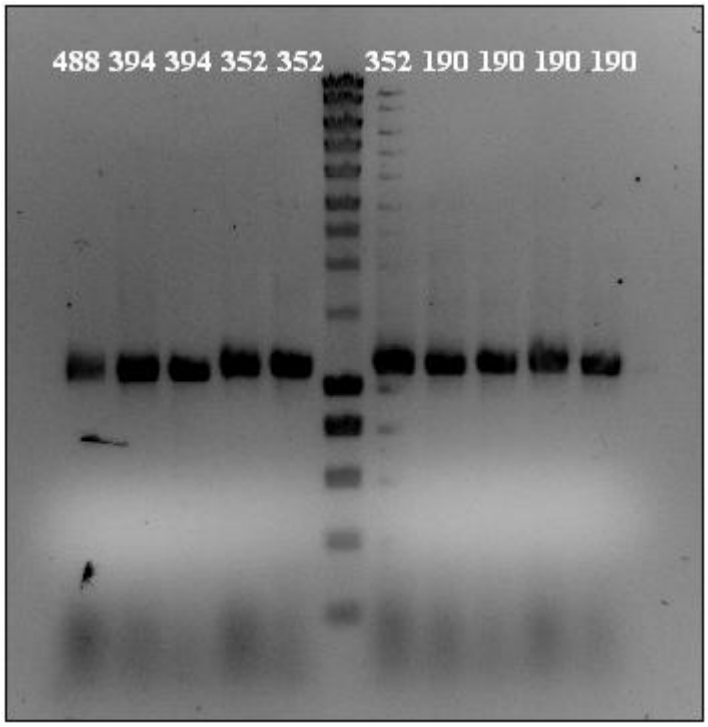


**Figure IV-4:** Results of the yeast colony PCR on the obtained transformants. The sizes of the GAPD promoter fragments are indicated above. The sizes of the marker fragments are: 10000, 8000, 6000, 5000, 4000, 3000, 2500, 2000, 1500, 1000, 800, 600, 400 and 200 bp.

A striking difference between *TK* and *GAPD* controlled hygromycin resistance is the timing of the appearance of resistant colonies. Whereas it could take even 15 days before (often very small colonies (which stayed often very small) arose in the *TK* controlled experiments, it took only 3 days before the first well growing transformants were observed in the *GAPD* controlled expression experiment. This observation confirms the hypothesis that the *GAPD* promoter is a strong constitutive promoter which can be used for heterologous expression in *C. bombicola*.

# 4 Conclusion

This short chapter describes the cloning of the complete *C. bombicola* glyceraldehyde-3-phosphate dehydrogenase gene (GenBank accession number EU315245). The obtained 3740 bp contain the 1008 bases of the coding sequence and 1613 and 783 bp of the upstream and downstream regions, respectively. The protein of 335 amino acids shows high homology to the other known *GAPD* genes and is 74 % identical to the gyceraldehyde-3-phosphate dehydrogenase of *Yarrowia lipolytica*.

The particular interest in the *C. bombicola GAPD* gene sequence originates from the potential use of its promoter for high and constitutive expression of homologous and heterologous genes. In some fungal genomes two or more *GAPD* genes are present, which are not always expressed to the same extent. This was verified for the *C. bombicola* genome by Southern blot; the results did not give any indication for the presence of multiple *GAPD* genes and it can therefore be expected that the promoter can be used for efficient and high expression. This hypothesis was further confirmed by the biased codon usage in the *GAPD* gene.
The 5' region possesses a conventional TATA-box at position -78, and two possible CAAT-sequences 366 and 354 bp upstream the ATG start codon. However, upon evaluating different promoter lengths for expressing a hygromycin resistance gene, fragments with none or only one CAAT-sequence still yielded resistant transformants. Strangely enough, the use of promoter fragments longer than 488 bp did not give rise to resistant colonies, while fragments of only 190 bp still did. This could be explained by the fact that longer promoter fragment tend to recombine with the genomic *GAPD* promoter, in this way knocking out the *GAPD* gene which is essential for cell viability. This event can be avoided in future expression experiments by flanking the gene with a functional *URA3* gene and transforming the ura3-negative G9 strain described in Chapter III, Section 4.3. This would lead to recombination at the dysfunctional *URA3* gene instead of the *GAPD* promoter. The place of integration can be verified by Southern blot.

Resistant colonies appeared much quicker as compared to transformations with a *TK* promoter controlled hygromycin resistance gene. This indicates that the homologous *GAPD* promoter is

far more active as compared to the heterologous *TK* promoter and thus can be used for the expression of heterologous and homologous genes in *C. bombicola*. The efficacy of a short *GAPD* promoter can be a convenient characteristic for the construction of compact expression cassettes or vectors for *C. bombicola*. However, the exact expression level of the different *GAPD* promoter fragments should first be investigated by *e.g.* expressing β-galactosidase and quantifying its activity.

# Chapter V: Improving medium-chain sophorolipid production

Part of this chapter will be submitted for publication as:

Van Bogaert INA, Develter D, Soetaert W, Vandamme EJ (2008) Improving medium-chain sophorolipid production by the yeast *Candida bombicola* by construction of a strain with blocked β-oxidation. J Lipid Res, in preparation

Van Bogaert INA, De Maeseneire SL, Develter D, Soetaert W, Vandamme EJ (2008) Cerulenin inhibits the de novo sophorolipid synthesis of *Candida bombicola*. Biotechnol Lett, accepted

# 1 Blocking the β-oxidation pathway

## 1.1 Introduction

When looking at sophorolipids in a simple structural way, they are composed of two glucose molecules and a long-chain fatty acid. As described in Chapter I, Section 4.5.2, sophorolipid production is optimal when those two building blocks are provided into the medium as the hydrophilic (glucose) and lipophilic (fatty acid) carbon source. However, carbon sources are also required for cell growth and energy supply. Preferentially, glucose will be used for this latter purpose. Fatty acids can also be metabolised by the cells by means of the β-oxidation pathway. For *C. bombicola* this especially happens when glucose concentrations are low, or when the fatty acids tend to accumulate because they are not easily consumed in another pathway such as sophorolipid synthesis. Therefore one could state that sophorolipid synthesis and β-oxidation are competing for the same fatty acid substrate.

When cultivating *C. bombicola* with a hydrophobic carbon source favourable for sophorolipid production such as rapeseed oil, the majority of this carbon source will end up into the biosurfactant. However, when *C. bombicola* is cultivated with a less favourable substrate for sophorolipid production such as shorter primary alcohols, part of them will be metabolized in the β-oxidation pathway. The production of sophorolipids based on those speciality substrates could probably be improved when the competing β-oxidation pathway is in some way eliminated.

Previous experiments with chemical inhibitors of the β-oxidation added to the medium did not result in better use of the substrate. Either they were not active against the yeast pathway or could not perform properly due to the long fermentation period at low pH. For this reason, it was chosen to try to block the β-oxidation pathway by knocking out one of the essential genes.

Before a fatty acid enters the actual oxidation pathway, it is converted to acyl-CoA by acyl-CoA-synthetase in the cytoplasmatic reticulum and then transported into the peroxisomes. It is believed that in yeasts β-oxidation only occurs in the peroxisomes, this in contrast to mammals where the metabolization of fatty acids occurs both in the peroxisomes and

mitochondria. The β-oxidation is a cyclic pathway with four reactions. After every cycle an acetyl-CoA molecule and an acyl-CoA lacking two carbon atoms are obtained, this until in the final cycle two acetyl-CoA molecules are released (Figure V-1).

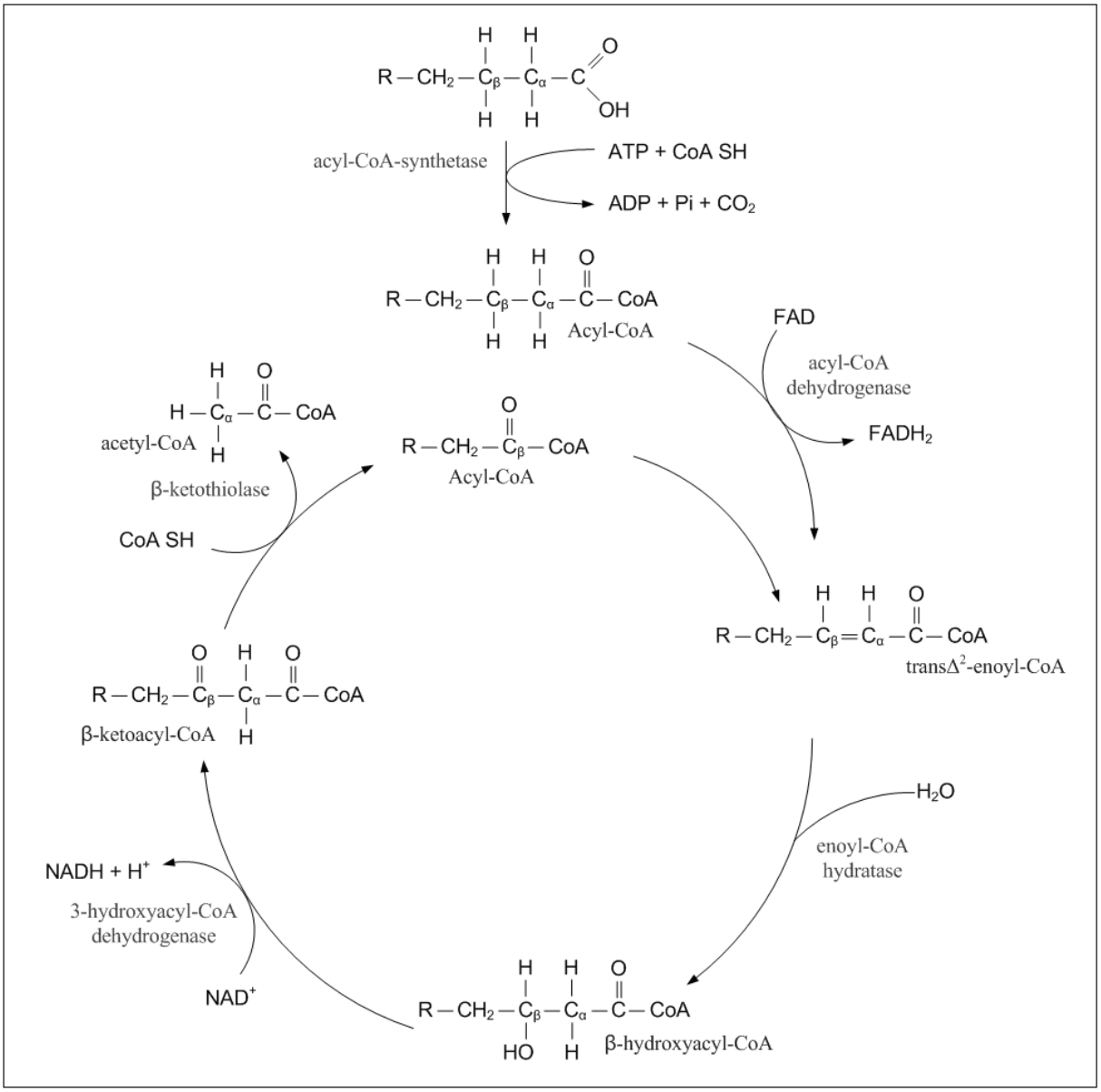


**Figure V-1:** Schematic overview the reactions taking part in the β-oxidation.

The first reaction is performed by acyl-CoA dehydrogenase. For some yeast species, multiple acyl-CoA dehydrogenase genes or *POX* genes occur, which are all functionally translated into isozymes. *C. maltosa* for example has two such genes (Masuda *et al.*, 1995) and *C. tropicalis*

three (Okazaki *et al.*, 1987). In the genome of *Yarrowia lipolytica*, even five *POX* genes were found (Wang *et al.*, 1999a) some of them show activity against short-chain fatty acids, others against long-chain fatty acids and some are active against molecules of all chain lengths. In order to completely shut down the β-oxidation pathway, a quadruple mutant had to be created (Wang *et al.*, 1999b). However, only one gene has been identified in *S. cerevisiae* or other conventional yeasts (Dmochowska *et al.*, 1990). The above findings indicate that the number of *POX* genes is variable, but that the chance on the occurrence of isozymes is much bigger for yeasts that readily metabolize alkanes or fatty acids. Also *C. bombicola* belongs to this latter group, and therefore it can be expected that in this species multiple *POX* genes are active as well. The creation of a mutant blocked in its β-oxidation by knocking out *POX* genes is consequently not very convenient since all *POX* genes must be isolated and possibly multiple deletions must be made.

The second (the hydratation) and third step (the second dehydrogenation) of the β-oxidation pathway are in eukaryotic organisms performed by one and the same enzyme: the so called multifunctional enzyme. In mammals, two of such enzymes occur: multifunctional enzyme type 1 (MFE-1) is located in the mitochondria, while multifunctional enzyme type 2 (MFE-2) is found in the peroxisomes. Yeasts only possess MFE-2 and until now there is no evidence for the presence of isozymes in any yeast or fungal species. Therefore, the MFE-2 from *C. bombicola* is an attractive target for gene disruption. Furthermore, by disabling the gene, two biochemical reactions are inhibited at the same time.

The last reaction of the β-oxidation cycle is the thiolytical cleavage. In *S. cerevisiae* only one gene, *FOX3*, was found and yeasts with a mutated gene were unable to grow on fatty acids, indicating the inactivation of β-oxidation (Zhang *et al.*, 2006). However, in the genome of *C. tropicalis* four isozymes with different activities and specificities were found. Double mutants were necessary to block β-oxidation (Kanayama *et al.*, 1998). Because of the possibility that also in *C. bombicola* several isozymes are present, it was decided not to target the thiolytical cleavage for blocking the β-oxidation pathway.

## 1.2 Materials and Methods

### 1.2.1 Strains, plasmids and culture conditions

*C. bombicola* ATCC 22214 was used for the preparation of genomic DNA. *Escherichia coli* DH5α was used in all cloning experiments. *C. bombicola* was maintained on agar plates containing 10 % glucose, 1 % yeast extract and 0.1 % urea. Yeast strains presumed to lack β-oxidation activity were tested for growth on Yeast Nitrogen base with Oleic acid (YNO) plates (0.67 % yeast nitrogen base without amino acids, 0.02 % Tween 40 and 0.1 % oleic acid). Sophorolipid synthesis of yeast strains with a blocked β-oxidation activity was tested on medium as described by Lang *et al.* (2000) with either 1-dodecanol (Sigma) used as hydrophobic carbon source in a concentration of 20 or 5 g/L or rapeseed oil (Sigma) used at 37.5 g/L. The hydrophobic carbon source was added two days after inoculation and the cultures were incubated for another eight days, unless specified otherwise.

Liquid yeast cultures were incubated at 30 °C and 200 rpm. *E. coli* was grown in Luria Bertani (LB) medium (1 % trypton, 0.5 % yeast extract and 0.5 % sodium chloride) supplemented with 100 mg/l ampicillin and 40 mg/l X-gal if necessary. Liquid *E. coli* cultures were incubated at 37 °C and 200 rpm.

All PCR products were cloned into the pGEM-T® vector (Promega).

### 1.2.2 DNA isolation and sequencing

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

### 1.2.3 Transformation

*C. bombicola* cells were transformed by the method described in Section 2.2.3 from Chapter III. Transformants were selected on synthetic dextrose (SD) plates (0.67 % yeast nitrogen

base without amino acids (DIFCO) and 2 % glucose). *E. coli* cells were transformed as described by Sambrook and Russell (2001).

### 1.2.4 Degenerate and re-amplification PCR

Part of the *MFE-2* gene of *C. bombicola* was amplified by the use of the degenerate primers MFE2DegFor308, MFE2DegFor552, MFE2DegRev1413 and MFE2DegRev1422 (Table V-1) based on the conserved sequences of the known *MFE-2* genes of the *Ascomycetes*.

**Table V-1:** Primers used for isolating, cloning and knocking-out the *C. bombicola MFE-2* gene. All primers were obtained from Sigma Genosys.

| Name | Feature | Sequence |
|---|---|---|
| MFE2DegFor308 | degenerate primer | 5'-TYHTBRTYAACAAYGCBGG-3' |
| MFE2DegFor552 | degenerate primer | 5'-GARACBYTBGCNAAGGARGG-3' |
| MFE2DegRev1413 | degenerate primer | 5'-TGDCCRAARTTRCCRTARATRCC-3' |
| MFE2DegRev1422 | degenerate primer | 5'-TAGTTRGCYTGDCCRAARTTRCC-3' |
| MFE2Up1 | primary upstream GSP | 5'-ATTCTCATGAACGAGGTATCCGACAACAG-3' |
| MFE2UpN | nested upstream GSP | 5'-GAGCTCGGGCTTCAGCTTCTCTAGAATC-3' |
| MFE2Down1 | primary downstream GSP | 5'-ATCTCTGGAATGCTGTTCAACAGGTTCATC-3' |
| MFE2DownN | nested downstream GSP | 5'-GAAGCAGAAATACGGACGTGTGGTCAAC-3' |
| MFE2DownII1 | primary downstream GSP | 5'-GATTTCGATGACGGGAAGGCTACCAAC-3' |
| MFE2DownIIN | nested downstream GSP | 5'-TTGAGTCCGCCTCCGAAGCAAACATGTA-3' |
| MFE2DownIII1 | primary downstream GSP | 5'-CAACCAGGCTGCAATCTATCGACTATCA-3' |
| MFE2DownIIIN | nested downstream GSP | 5'-AACCCTCTTCACATTGATCCTGAGTTTG-3' |
| MFE2_2000For | downstream GSP | 5'-CTTTGAGTCCGCCTCCGAAGCAAAC-3' |
| MFE2DownDegRev | degenerate primer | 5'-TGBARNGGRTTKHDRTCRCC-3 |
| MFE2_3545For | cloning MFE2 | 5'-CTCCTAAGCATTTAACTGCCTTGAG-3' |
| MFE2_3545Rev | cloning MFE2 | 5'-AACCGAGATACGCCTAATCAGTC-3' |
| URA3infMFE2For | ligating Ura3 into MFE2 | 5'-TGCGTTGCCCCTACACTGACGGGCGGATAGTACA-3' |
| URA3inf MFE2Rev | ligating Ura3 into MFE2 | 5'-TGGTCTTCGCCCTCACATCATCGTCACTATACAC-3' |
| MFE2knock1000For | 4095 bp knock-out fragment | 5'-GGCAACTTTGGCCAGGCCAATTA-3' |
| MFE2knock1000Rev | 4095 bp knock-out fragment | 5'-GTTTAGATGTGGCTCAAGTA-3' |
| MFE2knock500For | 3070 bp knock-out fragment | 5'-TACTGGTGCGGGAGGCGGAATTG-3' |
| MFE2knock500Rev | 3070 bp knock-out fragment | 5'-GGTGGCTTGTTGGCTGCTGTGAT-3' |

PCR amplification was carried out with an initial denaturation of 94 °C for 4 min, a 40 fold repeat of 94 °C for 30 s, 50 °C for 1 min, 72 °C for 2 min with 5 s increasing time and a final 7 min elongation at 72 °C.

Re-amplification for low concentrated PCR products was performed with the primers of the initial PCR and the following temperature program: denaturation at 92 °C of 1 min, a 30 fold repeat of 92 °C for 10 s, a primer specific annealing temperature for 30 s, 72 °C for 4 min and a final 7 min elongation at 72 °C. The primer specific annealing temperature was 50 °C for the re-amplification of the degenerate product and 67 °C for the walking product.

### 1.2.5 Genome Walking

The unknown genomic DNA sequences upstream and downstream of the degenerate PCR fragment were identified by genome walking with the BD GenomeWalker™ Universal Kit (BD Biosciences) as described in Section 3.2.5 of Chapter III. The four gene-specific primers (GSP) are listed in Table V-1.

### 1.2.6 Nucleotide sequence accession number

The nucleotide sequence described in this paper has been deposited at the GenBank nucleotide database under the accession number EU371724.

### 1.2.7 Creation of the knock-out fragments

The 2655 bp coding fragment and 493 and 397 bp upstream and downstream of the *MFE-2* gene were amplified with the primers MFE2_3545For and MFE2_3545Rev, yielding a fragment of 3545 bp which was cloned into the pGEM-T® vector (Promega). The created vector was digested with *Eco*NI and *Bbv*CI, which both cut the coding sequence of *MFE-2* once, thus deleting 309 bp of the *MFE-2* sequence. The *URA3* gene of *C. bombicola* was inserted by means of the In-Fusion™ 2.0 Dry-Down PCR Cloning Kit (Clontech). The primers URA3infMFE2For and URA3infMFE2Rev were designed according to the guidelines of the manual and used for integration of the functional *C. bombicola URA3* sequence (2055 bp) into *MFE-2*. The primerpair MFE2knock1000For and MFE2knock1000Rev were used for the amplification of a 4095 bp fragment containing the

*URA3* sequence with approximately 1000 bp of the *MFE-2* sequence on each site. In the same way the primers MFE2knock500For and MFE2knock500Rev were used to create a construct of 3070 bp with 500 bp flanking regions. Finally, a larger fragment (5287 bp) with 1962 bp and 1272 bp flanking respectively the 5' and 3' site of the *URA3* sequence was created with the MFE2_3545For and MFE2_3545Rev primerpair. The creation of the knock-out cassettes is illustrated in Figure V-2.

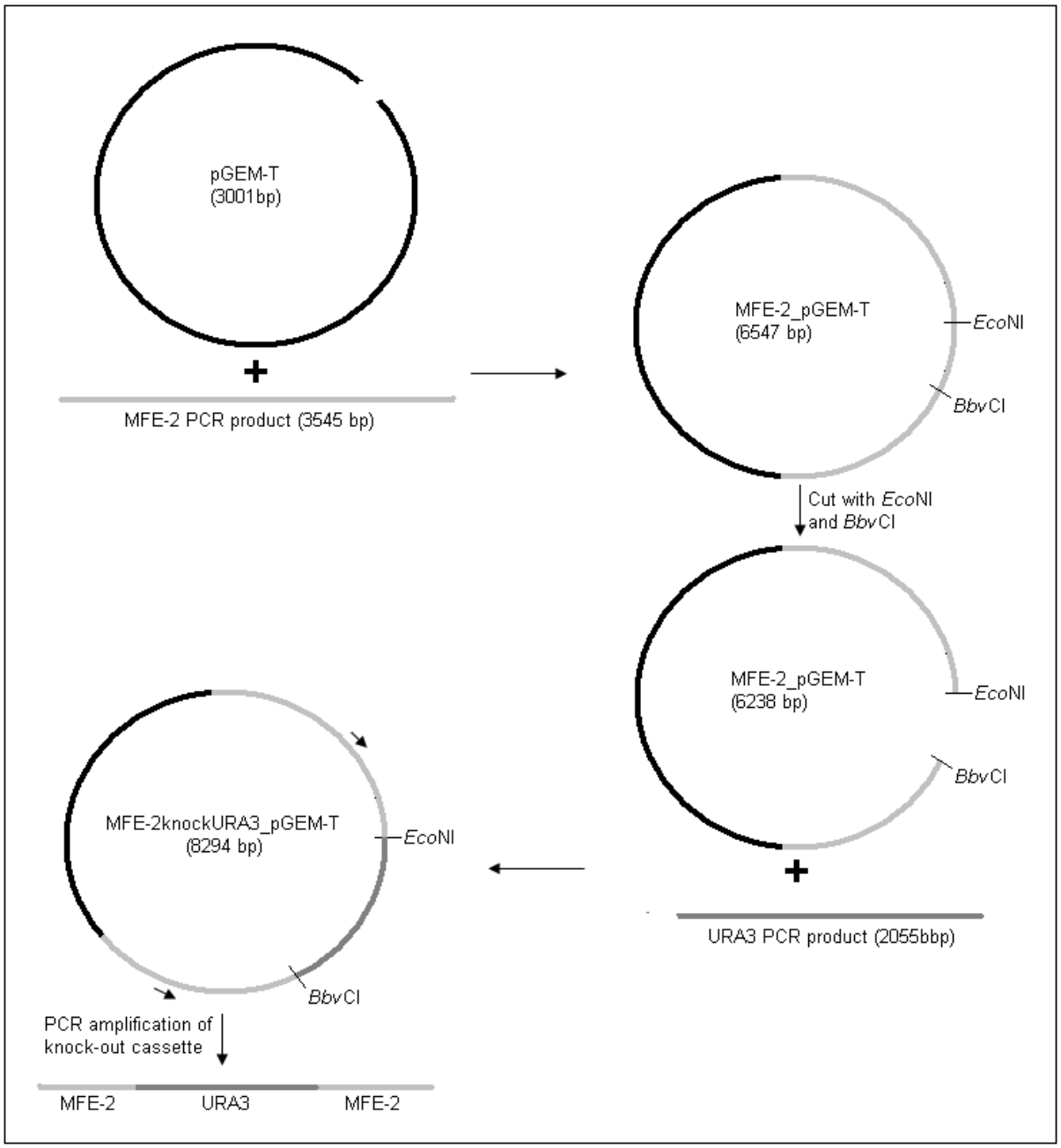


**Figure V-2:** Creation of the MFE-2 knock-out cassette with the URA3 selection marker for use in *C. bombicola*.

### 1.2.8 Sampling

CDW, CFU, sophorolipid and glucose concentration were determined as described in Chapter II, Section 2.2.

### 1.2.9 Final sophorolipid extraction from culture broth

The end extraction of sophorolipids was carried out as described in Chapter II, Section 2.3.

### 1.2.10 HPLC-analysis

HPLC analysis of the produced sophorolipids was performed as described in Chapter II, Section 2.4. In order to be able to compare and quantify the different samples, dilutions of a standard were analysed in parallel.

## 1.3 Results and Discussion

### 1.3.1 Isolation of the *C. bombicola MFE-2* gene

The multifunctional enzyme type 2 (MFE-2) catalysing the second and third step in the β-oxidation cycle was chosen as target for knocking out the β-oxidation pathway. The first step in creating the deletion strain consisted in obtaining (part of) the *C. bombicola MFE-2* sequence. Therefore, all known *Ascomycetous MFE-2* sequences were put into an alignment and degenerate primers were designed on the found conserved regions (Table V-1 and Figure V-3). At first, the primers with the lowest degeneracy, MFE2DegFor552 and MFE2DegRev1422, were used, but amplification of multiple fragments was observed, also at higher annealing temperatures and it was not clear whether the right band of about 890 bp was present. All other forward and reverse primer combinations were tested at several annealing temperatures, but again yielded multiple fragments. Therefore, a nested PCR on the MFE2DegFor308 and MFE2DegRev1422 PCR mixture was conducted with the primers MFE2DegFor552 and MFE2DegRev1413 resulting in a single very weak fragment of the expected length of about 850 bp. To increase the amplicon concentration, a re-amplification PCR was performed before isolating of the PCR-product. The nucleic acid sequence was determined and the fragment of 851 bp showed a 62 % amino acid identity with the MFE-2 from *Yarrowia lipolytica* and was therefore believed to be indeed part of the *MFE-2* gene of *C. bombicola*. The rest of the sequence was elucidated by genome walking with the BD GenomeWalker™ Universal Kit and the primers listed in Table V-1.

For obtaining the downstream sequence, a second round of genome walking was required as the stop codon was not included in the short obtained fragment of 500 bp. The second pair of primers designed on the newly acquired sequence did however not resulted in the amplification of any fragment for all four available walking banks. Possibly, the fragments were either too small or too lengthy for efficient amplification. For this reason, a semi-degenerate strategy was used: a forward primer MFE2_2000For was designed on the 3' end of the already known sequence, while a (very) degenerate primer MFE2DownDegRev was designed on the unknown region, based on the sequence alignment of the known *Ascomycetous MFE-2* sequences. Both primers are listed in Table V-1. This approach indeed

yielded a fragment of the expected length of 700 bp harbouring part of the *C. bombicola* MFE-2 gene. This sequence was then used for a third round of downstream walking. Only for one bank an amplification fragment was obtained, visible as a weak band. In order to get a higher DNA concentration, a re-amplification PCR was conducted prior to isolation of the PCR product. A schematic representation of the followed strategy can be seen in Figure V-3. The genome walking yielded a total genome fragment of 6033 bp, containing the total 2655 bp coding sequence of the *C. bombicola MFE-2* and 2909 and 469 bp respectively of the 5' and 3' regions.

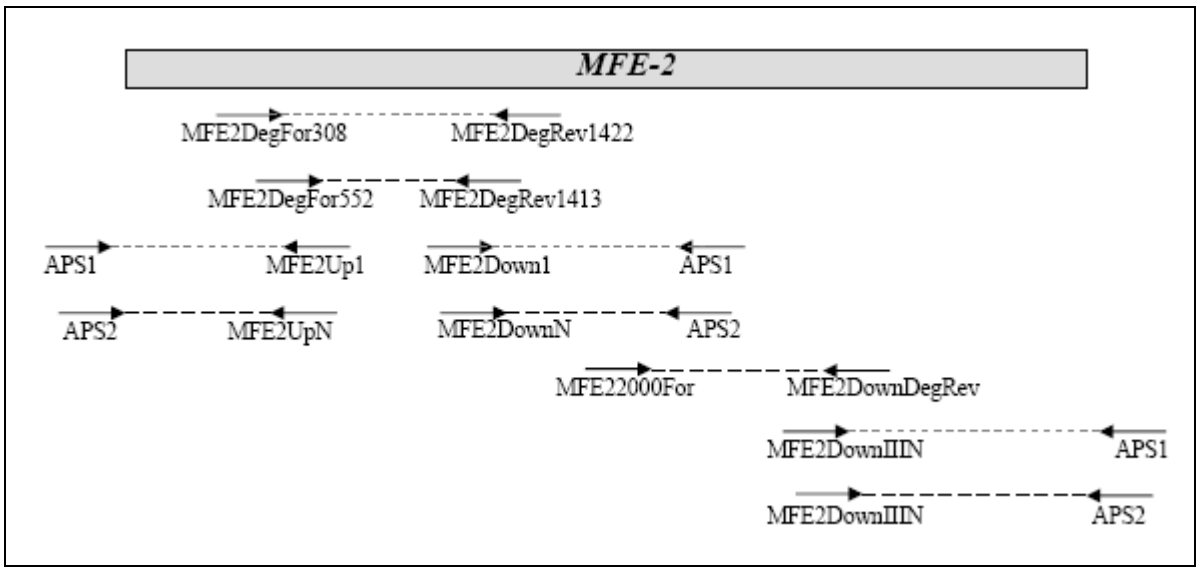


**Figure V-3:** Schematic representation of the followed degenerate PCR and walking strategy to obtain the *C. bombicola MFE-2* nucleic acid sequence.

### 1.3.2 Characterisation of the *MFE-2* sequence

In the promoter region of the MFE-2 gene, a possible TATA-box was identified either starting at position -74 or -72 (Figure V-4). A CAAT-box was also present 217 bp upstream the ATG start codon. Two putative polyadenylation signals were found in the 3' region: the conventional eukaryotic AATAAA sequence 20 bp downstream the TGA stop codon and the more yeast-specific ATAATA 216 bp away from the stop codon (Graber *et al.*, 1999).

```
   1  agaccataaa aaagcaggcg atttaatttc gccaatgccg aggctgcatg gtggagatat gtctgaattt cggcggatag gtacccttac
  91  ctcaacccaa aagtgcgggg ggaagcccga ctgtctgccg atactttatg gcgcagttcc gcaatagccg agtggggtcg tgcatatata
 181  tataggggtc actcatcatt cgattgtgag ataaactaca ttgaaaatac aagaagggaa agtgaaacat ggcggagaat cttaggtacg
   1                                                                MFE-2 →       M  A  E  N   L  R  Y

 271  acggcaaagt tgtcgttgtc actggcgcag gaggcggact tggaaaagcc tacgcgctgt tctttggtgc tcgtggtgca tcagttgttg
   8   D  G  K   V  V  V  V   T  G  A   G  G  G   L  G  K  A   Y  A  L   F  F  G   A  R  G  A   S  V  V

 361  tgaatgatct cggaggcaca ttgaatggtg gtgacggtaa ctctagagtt gctgatggag ttgtaaaaga aattgaggct cttggaggca
  38   V  N  D   L  G  G  T   L  N  G   G  D  G   N  S  R  V   A  D  G   V  V  K   E  I  E  A   L  G  G

 451  aagcagcggc aaattacgat agcgtcgaga atggtgacaa gatcgttgaa accgctatca aagcattcgg cacggtacac attatcatca
  68   K  A  A   A  N  Y  D   S  V  E   N  G  D   K  I  V  E   T  A  I   K  A  F   G  T  V  H   I  I  I

 541  ataacgcagg aattcttcgt gacgtcagtc taaaaaagat gacggataag gatttcaact ttgtccagtc tgtccacgtc tttggctcgt
  98   N  N  A   G  I  L  R   D  V  S   L  K  K   M  T  D  K   D  F  N   F  V  Q   S  V  H  V   F  G  S

 631  acgcggttac gagggctgct tggccttatt tcaaacaaca gaagttcggt cgtgtgatca acaccgcaag cgcagctggt ctatatggca
 128   Y  A  V   T  R  A  A   W  P  Y   F  K  Q   Q  K  F  G   R  V  I   N  T  A   S  A  A  G   L  Y  G

 721  actttggcca ggccaattat tctgcggcta aatccgcttt ggtgggcttt actgaaactt tggctaaaga gggcgccaaa tacaatatca
 158   N  F  G   Q  A  N  Y   S  A  A   K  S  A   L  V  G  F   T  E  T   L  A  K   E  G  A  K   Y  N  I

 811  ccgctaacgt tattgttcca ctggcggcgt cgcgcatgac tgagaccatt cttcctcctg acattctaga gaagctgaag cccgagctca
 188   T  A  N   V  I  V  P   L  A  A   S  R  M   T  E  T  I   L  P  P   D  I  L   E  K  L  K   P  E  L

 901  tcgttcctgt tgtcggatac ctcgttcatg agaatacagc agagagcaat ggaatctacg aaagtgctgc tggcgttgta accaaggtga
 218   I  V  P   V  V  G  Y   L  V  H   E  N  T   A  E  S  N   G  I  Y   E  S  A   A  G  V  V   T  K  V

 991  gatggcagcg tggagctggt gtacagttca gggctgatga ctcgttcact cccgctgcag tgttgaacaa attcgaagaa atcaacgaca
 248   R  W  Q   R  G  A  G   V  Q  F   R  A  D   D  S  F  T   P  A  A   V  L  N   K  F  E  E   I  N  D

1081  actttgagcc agcagagtac cccagtgggc ccaaggatct tctagctgcc tttgaaaatg gcaagaatct gccttcgaat gagcagggaa
 278   N  F  E   P  A  E  Y   P  S  G   P  K  D   L  L  A  A   F  E  N   G  K  N   L  P  S  N   E  Q  G

1171  gcactccagt aagtttcgag aaccaggtcg ttatcgttac tggtgcggga ggcggaattg gacagcaata tgctctcatg ctcggtaagt
 308   S  T  P   V  S  F  E   N  Q   V  V  I  V   T  G  A  G   G  G  I   G  Q  Q   Y  A  L  M   L  G  K

1261  tgggagcgaa ggttgttgtg aatgaccttg gcaacgctga tgctaccgtg gaattgatca agaaggctgg aggaacggct gtggcggata
 338   L  G  A   K  V  V  V   N  D   L  G  N  A   D  A  T  V   E  L  I   K  K  A   G  G  T  A   V  A  D

1351  agcacaatgt caccgatggt gaggcagttg tgaagactgc tctagacaat ttcggtgcta tccatgcggt tatcaacaat gccggtatca
 368   K  H  N   V  T  D  G   E  A  V   V  K  T   A  L  D  N   F  G  A   I  H  A   V  I  N  N   A  G  I

1441  ttcgtgatcg tggcattctc aagatgacgc ccgatctctg gaatgctgtt caacaggttc atctattcgg ttccttttca gtcaccaaag
 398   I  R  D   R  G  I  L   K  M  T   P  D  L   W  N  A  V   Q  Q  V   H  L  F   G  S  F  S   V  T  K

1531  ctgcatggcc tcatttccag aagcagaaat acggacgtgt ggtcaacacg acttcaacct ctggaatcta cggaaatttc ggacagacaa
 428   A  A  W   P  H  F  Q   K  Q  K   Y  G  R   V  V  N  T   T  S  T   S  G  I   Y  G  N  F   G  Q  T

1621  actactcagc ggcaaaggcc ggtctcattg gcttcaccaa gactgtggca ctagaaggtg ccaagtacaa cattctttgc aactgcgttg
 458   N  Y  S   A  A  K  A   G  L  I   G  F  T   K  T  V  A   L  E  G   A  K  Y   N  I  L  C   N  C  V

1711  cccctacagc aggaactgct atgactgctg atgtgttccc tcaagatatg ctggagacgt tgaagccaag gtacattgcg ccaatcactg
 488   A  P  T   A  G  T  A   M  T  A   D  V  F   P  Q  D  M   L  E  T   L  K  P   R  Y  I  A   P  I  T

1801  tcctgcttgc tagtgagcac tcgcccgaca ccggtaaggt ctacgaagca ggtgctggct ggattggccg cacgcgttgg cagagaactt
 518   V  L  L   A  S  E  H   S  P  D   T  G  K   V  Y  E  A   G  A  G   W  I  G   R  T  R  W   Q  R  T

1891  cgggtgtcat gattcctggt atcacagtgg aaaaggttaa gcaaaattgg cagaaaatca ccgatttcga tgacgggaag gctaccaact
 548   S  G  V   M  I  P  G   I  T  V   E  K  V   K  Q  N  W   Q  K  I   T  D  F   D  D  G  K   A  T  N

1981  ttgagtccgc ctccgaagca aacatgtaca tcttcaacat ggcagctgag ggcgaagacc aaggctctga aggtggcgaa tctgaggctt
 578   F  E  S   A  S  E  A   N  M  Y   I  F  N   M  A  A  E   G  E  D   Q  G  S   E  G  G  E   S  E  A

2071  cagcaagcgg cgaatattct tacgacgaca agacgatcat tttgtacaac ttgggagttg gtgcgagcga gaagcagctc aattatactt
 608   S  A  S   G  E  Y  S   Y  D  D   K  T  I   I  L  Y  N   L  G  V   G  A  S   E  K  Q  L   N  Y  T

2161  ttgaaaacaa tcaggatttc cagccagtgc cgagtttcgg caccatcccg ctcttcagcg ctccattccc atttgatgaa gttgtgccca
 638   F  E  N   N  Q  D  F   Q  P  V   P  S  F   G  T  I  P   L  F  S   A  P  F   P  F  D  E   V  V  P

2251  atttcaatcc aatgaagctc cttcatggag agcaatattt ggagttgaag aagtggccca ttgccccaga ggcaacgttg aagaccacgg
 668   N  F  N   P  M  K  L   L  H  G   E  Q  Y   L  E  L  K   K  W  P   I  A  P   E  A  T  L   K  T  T

2341  gcaagcttct cgatcttgca gacaagggca aagctgctgt agcgatggtg gaatatatct ctgtcgataa gaattctggt gagcctgtgt
 698   G  K  L   L  D  L  A   D  K  G   K  A  A   V  A  M  V   E  Y  I   S  V  D   K  N  S  G   E  P  V

2431  tcctcaacgt catgtcaaca ttcttgagag gctccggagg tttcgggggt gagaagaatt tcaaggacca tggccccatc acagcagcca
 728   F  L  N   V  M  S  T   F  L  R   G  S  G   G  F  G  G   E  K  N   F  K  D   H  G  P  I   T  A  A

2521  acaagccacc ggctcgcgag cccgactata tcgccaagta caagaccacg gacaaccagg ctgcaatcta tcgactatca ggagactaca
 758   N  K  P   P  A  R  E   P  D  Y   I  A  K   Y  K  T  T   D  N  Q   A  A  I   Y  R  L  S   G  D  Y

2611  accctcttca cattgatcct gagtttgctg ccgttggcgg attcgatcgt ccgattcttc acggccttgc gtctttcgga atctcatcaa
 788   N  P  L   H  I  D  P   E  F  A   A  V  G   G  F  D  R   P  I  L   H  G  L   A  S  F  G   I  S  S

2701  gattgttggt tgaaaagtat ggcgttttca agaacatcaa ggtaagattc tcgggccatg tgttccctgg tgagactctg caagtttccg
 818   R  L  L   V  E  K  Y   G  V  F   K  N  I   K  V  R  F   S  G  H   V  F  P   G  E  T  L   Q  V  S

2791  cttggaagga aggtcccaag gtgatttttg agacgacggt gctggagcgt aacaccaaag ccattactgc agcagcaatt gagctggctg
 848   A  W  K   E  G  P  K   V  I  F   E  T  T   V  L  E  R   N  T  K   A  I  T   A  A  A  I   E  L  A

2881  atgatggtaa gtctaagctg tgacggagaa gctatatagt taataaata agattacaag tcgacttctg tcaggccaga ggtgtaggtt
 878   D  D  G   K  S  K  L   *
```

**Figure V-4:** Nucleotide and deduced amino acid sequence of the *MFE-2* gene of *C. bombicola*. The deduced amino acid sequence of the putative protein upstream *MFE-2* is also indicated. Possible promoter elements are underlined and sequences presumed to be involved in polyadenylation are marked in bold. The conserved amino acid sequences of the nucleotide binding domains A and B are underlined and in bold. The PTS1is boxed.

The protein of 884 amino acids has a theoretical molecular weight of 94.45 kDa and pI of 5.93. After blasting MFE-2 of *C. bombicola* against the protein database, it became clear that it was most similar to MFE-2 of *Yarrowia lypolytica* (61 % amino acid based). Surprisingly, the sequence of *C. bombicola* then shows a better match with the MFE-2's of filamentous fungal *Ascomycota* (*Pezizomycotina*) instead of MFE-2's from the yeast-like *Ascomycota* (*Saccharomycotina*). The same trend is observed when looking at the phylogenetic tree of known and putative MFE-2's from the *Ascomycota*; the *C. bombicola* MFE-2 clusters together with the *Y. lipolytica* protein and this branch in turn clusters together with the one of the filamentous *Ascomycota* (Figure V-5). The Ln Likelihood of the tree is -20116 and all knobs and branches were statistically significant ($p < 0.01$).

There are two types of peroxisomal targeting signals (PTS): the C-terminal tripeptide signal PTS1 with the consensus sequence (S/A/C)-(K/R/H)-(L/A) occurs in most of the peroxisomal matrix proteins. The most common PTS1 is SKL and this sequence is also found in the *C. bombicola* MFE-2 (Neuberger *et al.*, 2003).

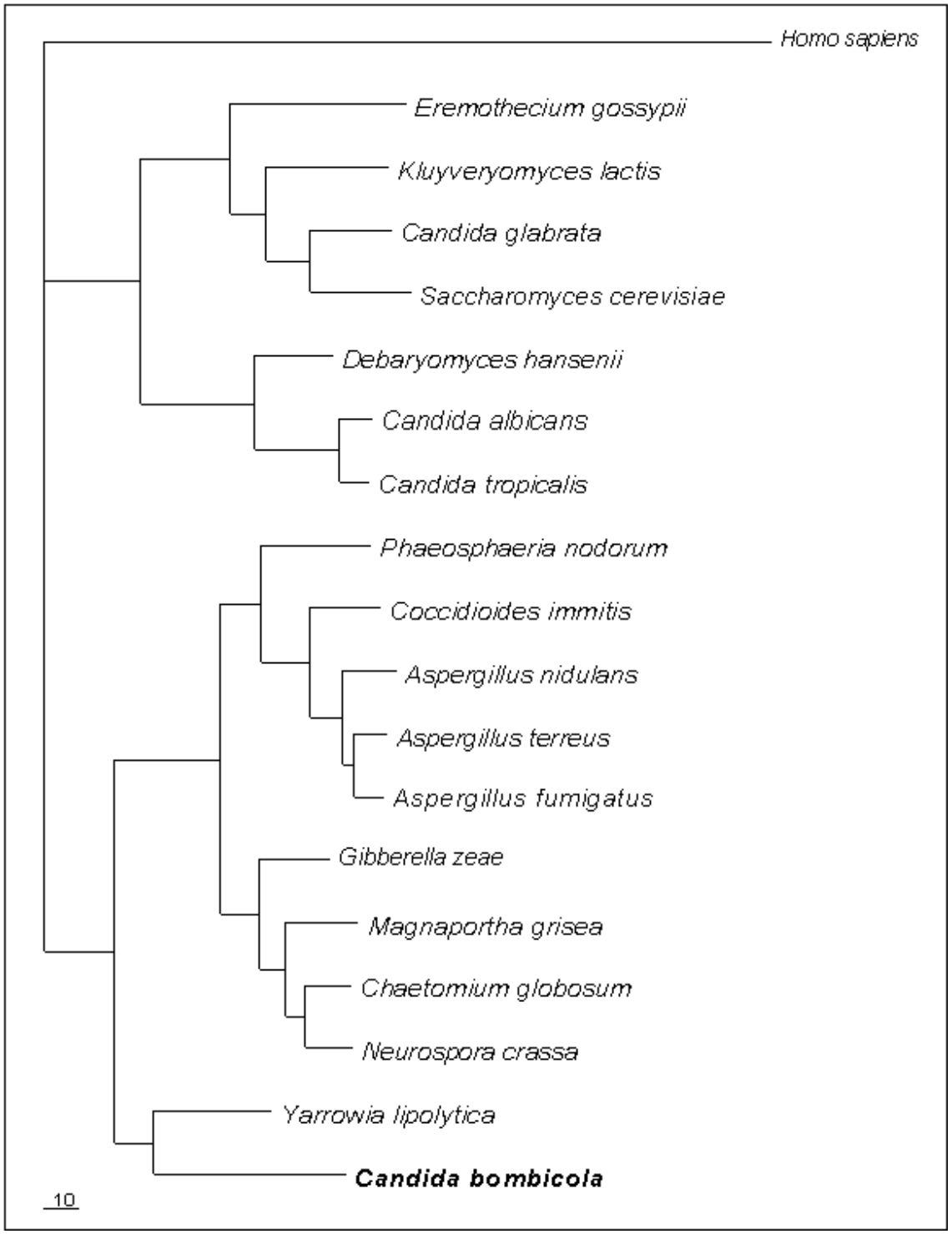


**Figure V-5:** Phylogenetic tree of known and putative MFE-2's from *Ascomycota* routed against the human MFE-2. The marker bar below denotes the integer branch length.

The 3-hydroxyacyl-CoA dehydrogenase activity is found in the first two thirds of the enzyme, while the enoyl-CoA hydratase activity is located in the last third. The dehydrogenase area has two domains belonging to the short chain alcohol dehydrogenase/reductase superfamily which are called type A and B. In those domains, the nucleotide binding sites (binding of $NAD^+$) are highly conserved, also between type A and B (Figures V-4 and V-6). The glycine at position 16 is one of the key residues; a mutation of this amino acid leads to MFE-2 deficiency (Qin *et al.*, 1999). In the human MFE-2, only one such domain is seen. The domains A and B from yeasts have different enzymatic properties: domain A shows the highest activity with medium- and long-chain (3R)-hydroxyacyl-CoA, whereas domain B prefers short-chain substrates. Nevertheless, both domains are required for optimal growth on fatty acids as the sole carbon source (Qin *et al.*, 1999).

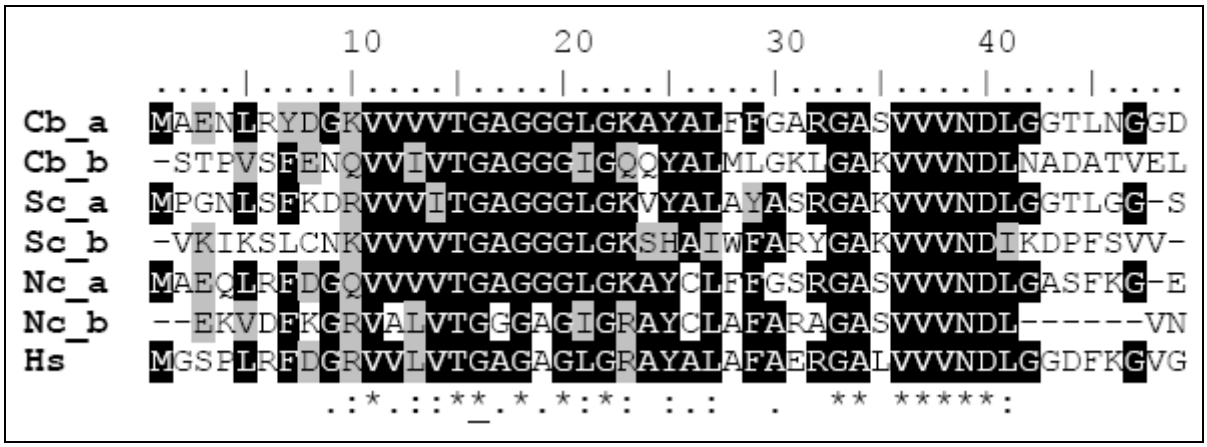


**Figure V-6:** Alignment of the nucleotide binding sites from *C. bombicola* (Cb_a and Cb_b for domain A and B respectively), *S. cerevisiae* (Sc_a and Sc_b for domain A and B respectively), *Neurospora crassa* (Nc_a and Nc_b for domain A and B respectively) and the human sequence (Hs). The nucleotide binding site is located between position 10 and 41. The glycine key residue is marked with a small bar.

### 1.3.3 Other putative genes flanking *MFE-2*

In the region upstream the MFE-2 gene sequence, a small open reading frame of 602 bp was found. The corresponding gene product shows a 68 % amino acid identity with the vacuolar ATP-synthase V0 subunit c" from the yeast *Lodderomyces elongisporus* and also has high homologies with vacuolar ATP synthase V0 subunit c" proteins from other yeasts. The vacuolar ATP synthase is an heteromultimeric enzyme composed of a peripheral catalytic V1

complex (components A to H) attached to an integral membrane V0 proton pore complex (components: a, c, c', c", d and e). The proteolipid components c, c' and c" are present as a hexameric ring that forms the proton-conducting pore.

### 1.3.4 Creating a *MFE-2* negative *C. bombicola* strain

Knocking out genes in *S. cerevisiae* is relatively simple: by means of PCR one can construct a linear fragment containing a marker flanked on each site by only 40 bp of the target gene and transform the yeast cells with this construct. Since *S. cerevisiae* laboratory strains are very efficient in homologous recombination, events of double cross-over or gene replacement occur frequently enough to yield the right transformants (Brachmann *et al.*, 1998). This strategy was also tested for *C. bombicola*; the functional *URA3* sequence was flanked with 60 bp of the *C. bombicola MFE-2* gene and used to transform the ura3-auxotrophic G9 strain described in Chapter III, Section 4. As expected, this did not yield proper transformants. 60 bp homology is probably too short for efficient recombination with the *C. bombicola* genome. Also for other non-conventional yeasts longer fragments of several hundreds or even more than thousand bp must be used (Wesolowski-Louvel *et al.*, 1998). Therefore, a disruption vector was created as described in the Materials and Methods, Section 1.2.7. It contains the 2055 bp of the *URA3* coding sequence with promoter inserted into the MFE2 sequence. The vector was then used as a template in PCR reactions with the primers listed in Table V-1 to create a linear marker sequence flanked with MFE-2 sequences of a variable length. In this way fragments with about 500, 1000 or more bp of the *MFE-2* gene on both ends were created with a total length of respectively 3070, 4095 and 5287 bp. 1.5 µg of each of the three fragments were used to transform *C. bombicola*. The fragment with 1000 bp flanking regions yielded about threefold more transformants compared to the other two.

The results of Chapter III, Section 4.3.2 demonstrated that, when transforming the ura3-auxotrophic G9 *C. bombicola* strain with a functional URA3 gene, double cross-over with a consequent gene replacement takes place at a rate of 19 %. All other transformants arise owing to events of single cross over or illegitimate recombination. For this reason, it is necessary to further test the obtained MFE2-negative candidates. The correct phenotype was tested by plating the mutants on medium with oleic acid as the sole carbon source (YNO);

mutants with a deficient β-oxidation will not succeed in growing on this medium (Figure V-7). 9 transformants were not able to grow on this medium, while the growth on general YPD medium was good: M9, M15, M16, M18, M24, M27, M30, M32 and M33.

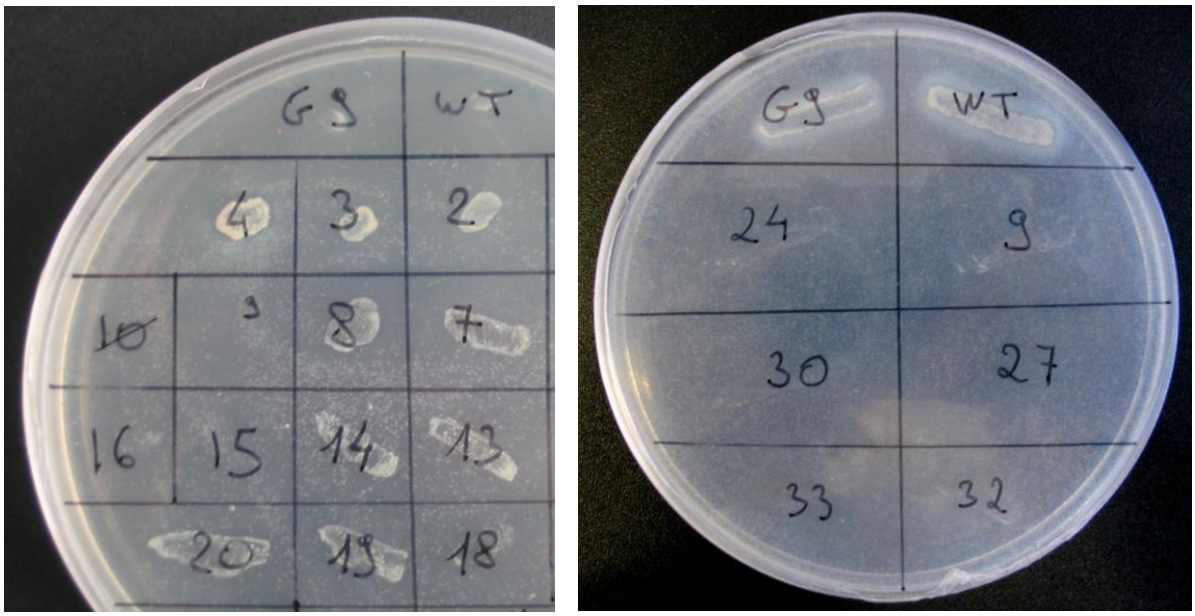


**Figure V-7:** Growth of *C. bombicola* wild type (wt), the ura3-auxotrophic G9 strain (G9) and several mutants on YNO medium. Yeast cells unable to metabolize oleic acid will not be able to grow on this medium.

The genotype of the obtained mutants was confirmed by a yeast colony PCR targeting the *MFE-2* sequence with the primers MFE2_3545For and MFE2_3545Rev (Table V-1). Presence of the intact wild type *MFE-2* gene should yield a fragment of 3545 bp, whereas for yeasts with the knocked-out *MFE-2*, this fragment must be either absent or appear as a larger band of about 5000 bp. All 9 strains selected by their non-growth on YNO-plates displayed the right genotype.

### 1.3.5 Evaluation of sophorolipid synthesis by the *MFE-2* negative strains

A phenomenon often seen regarding the synthesis of sophorolipids on terminal hydroxylated medium-chain substrates such as alcohols, diols and hydroxy-fatty acids, is the metabolization of the lipophilic substrate in the β-oxidation pathway. This observation is also described in Chapter II for the production of sophorolipids based on 12-hydroxydodecanoic acid and 1,12-

dodecanediol. Furthermore, Brakemeier *et al.* (1998a and b) prefer 2-alkanols and alkanones over 1-alkanols because the latter are is to a larger extent metabolized by the yeast and consequently not converted to sophorolipid molecules. On the other hand, the use of those terminal hydroxylated substrates is a convenient way to have *C. bombicola* synthesize medium-chain sophorolipids. Therefore, it was tested if knocking out the β-oxidation could redirect the hydroxylated medium-chain substrates towards sophorolipid biosynthesis.

It was decided to test this hypothesis with 1-dodecanol because this substrate shows a better solubility at 30 °C than 12-hydroxydodecanoic acid and 1,12-dodecanediol which facilitates sophorolipid synthesis and sampling. In addition, the composition of the sophorolipid mixture is quite simple as compared to the other substrates which makes evaluation between the wild type and mutants more straightforward. The results of the fermentations run on 1-dodecanol are given in Table V-2.

**Table V-2:** Sophorolipid production on 1-dodecanol of mutant strains relative to the wild type.

| Strains | Relative amount of sophorolipids (%) | CDW (g/L) | CDW corrected relative amount of sophorolipids (%) |
|---|---|---|---|
| Wild type | 100.0 | 7.0 | 100.0 |
| M9 | 285.7 | 9.0 | 222.2 |
| M15 | 262.9 | 9.5 | 193.7 |
| M16 | 237.1 | 9.0 | 184.4 |
| M18 | 274.3 | 8.0 | 240.0 |
| M24 | 217.1 | 7.5 | 202.7 |
| M27 | 254.3 | 9.0 | 197.8 |
| M30 | 308.6 | 7.5 | 288.0 |
| M32 | 262.9 | 11.0 | 167.3 |
| M33 | 278.0 | 8.5 | 228.9 |

A total mass of 9.5 g/L sophorolipids was synthesized by the wild type, while for all mutant strains clearly better results were obtained; the overall yield was 2.2 till 3.1 times higher.

Since biomass formation differs among the strains and is slightly lower for the wild type, this factor was taken into account for correct comparison of the obtained sophorolipid values. Still, sophorolipid production of the mutant strains stayed 1.7 to 2.9 times higher as compared to the wild type; it therefore can be concluded that strains with a blocked β-oxidation route perform better when synthesizing sophorolipids on 1-dodecanol.

Because of the good results with the mutant strains on 1-dodecanol, it was decided to further test their sophorolipid producing capacities on rapeseed oil. Since this substrate is composed of fatty acids with the optimal chain length of 16 to 18 carbon atoms, it is readily integrated into sophorolipid molecules. As illustrated in Section 1 of Chapter II, the efficiency of incorporation is high, and it is therefore questionable if the use of mutants with a blocked β-oxidation pathway can still increase the overall sophorolipid yield. Table V-3 shows however an unexpected result: the sophorolipid yield of the mutants is far less as compared to the wild type, even after correction for the differences in CDW. There is no straightforward explanation for this event, but it can be seen that the CDW value of the wild type is much higher than the values of the mutants and when looking at the graphs of the CFU, the same trend is observed (Figure V-8). Nevertheless, a certain recovery of the CFU values of all mutants except M9 is detected. It is therefore thought that the cells are somehow inhibited by rapeseed fatty acid derived products which accumulate due to the blocked β-oxidation, such as increased intracellular amounts of acyl-CoA derivatives produced via the upper part of the β-oxidation. The recovery of the mutant cells can be explained by the fact that sophorolipid synthesis becomes more important once the stationary phase is established and that this pathway consumes the rapeseed fatty acids and/or their derivates.

**Table V-3:** Sophorolipid production on rapeseed oil of mutant strains relative to the wild type.

| Strains | Relative amount of sophorolipids (%) | CDW (g/L) | CDW corrected relative amount of sophorolipids (%) |
|---|---|---|---|
| Wild type | 100.0 | 15.0 | 100.0 |
| M9 | 26.1 | 9.5 | 41.2 |
| M18 | 37.0 | 10.0 | 55.4 |
| M24 | 14.3 | 8.0 | 26.9 |
| M27 | 18.5 | 9.5 | 29.2 |
| M30 | 14.6 | 11.5 | 19.0 |

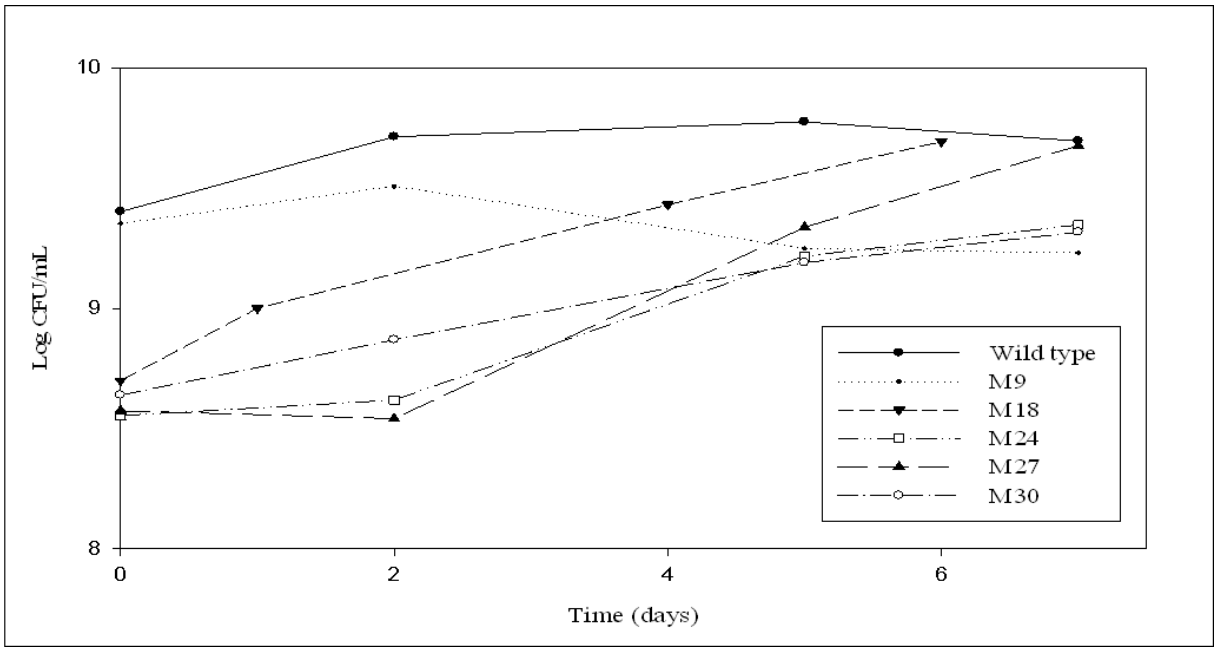


**Figure V-8:** CFU values of the different strains during the sophorolipid synthesis on rapeseed oil. Sampling was started after the addition of rapeseed oil to a 48 hours old culture (day 0).

# 2 Inhibition of *de novo* synthesis of sophorolipids

## 2.1 Introduction

The yeast *C. bombicola* can still produce sophorolipids even when no hydrophobic carbon source is available. These sophorolipids are formed *de novo*; the required fatty acids are derived by *de novo* synthesis by the fatty acid synthase system, hydroxylated and subsequently incorporated into sophorolipids. The *de novo* sophorolipids all posses a C16 or C18 fatty acid tail (Table II-1) and even in the presence of a suitable hydrophobic carbon source for sophorolipid production, *de novo* formation of sophorolipids is still observed. The relative amount of the *de novo* sophorolipids depends on the ease of integration of the presented hydrophobic carbon source; it is quite obvious that for fermentations run on substrates which are readily incorporated into the sophorolipid molecule, the *de novo* formation becomes less significant. This phenomenon has also been observed for the fermentations described in Chapter II.

Hence, when performing a fermentation on unconventional hydrophobic carbon sources to obtain medium-chain or new-to-nature sophorolipids, the achieved sophorolipid mixture will always be contaminated by the *de novo* compounds. This makes it rather difficult to test the physico-chemical characteristics of the special sophorolipids and consequently requires further purifications steps. To circumvent this inconvenience and to redirect the precursors and co-factors for sophorolipid synthesis towards the medium-chain sophorolipids, it would be opportune if the *de novo* formation of sophorolipids could in some sort of way be blocked.

Synthesis of fatty acids is carried out by fatty acid synthase (FAS). For fungi, animals and some bacteria, FAS is a multimeric polypeptide chain consisting of an acyl carrier protein (which carries the growing fatty acid chain) linked to the other enzymatic domains. The enzymatic reactions are in fact reversed to those of the β-oxidation pathway and start with the precursors acetyl-CoA and malonyl-CoA (Wakil, 1989).

Some antibiotics are known to act on the synthesis of fatty acids, in this way inhibiting growth. Cerulenin for example, is an antifungal antibiotic having strong inhibitory activity against *C. albicans*, *S. cerevisiae* and other fungi (Figure V-9). The antibiotic inhibits the

condensation step accomplished by the β-ketoacyl thioester synthetase and has a minimum inhibitory concentration (MIC) of approximately 0.8 to 1.5 mg/L (Omura, 1976). However, when applying the antibiotic on cells in the stationary phase (when their viability does not depend that strongly on the formation of fatty acids) it can be a useful biochemical tool for investigation of the lipid metabolism. However, the concentration to be applied must be higher. For example, Kitamoto *et al.* (1995) used 20 mg/L for inhibiting the fatty acid synthesis of *C. antarctica*. In this section, the effect of cerulenin on the fatty acid synthesis of *C. bombicola* and the hereto related *de novo* synthesis of sophorolipids is evaluated.

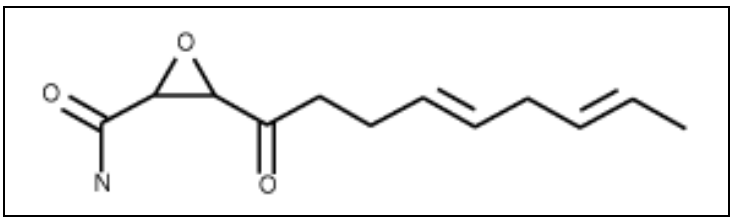


**Figure V-9:** Structural formula of cerulenin.

## 2.2 Materials and Methods

### 2.2.1 Strains and culture conditions

*C. bombicola* ATCC 22214 was used in all experiments and was grown in liquid medium as described by Lang *et al.* (2000). Cerulenin was added 48 hours after inoculation and was dissolved in ethanol. A yeast culture supplemented with the same volume of ethanol was used as a reference. Cerulenin was allowed to be taken up be the cells for a few hours before 1,12-dodecanediol (Sigma) was added (18.75 g/L). Four days after the first addition of this substrate, a second amount of 18.5 g/L was added). Yeast cultures were incubated at 30 °C and 200 rpm.

### 2.2.2 Toxic effect of cerulenin

*C. bombicola* cells were grown for 48 hours when different concentrations of cerulenin were added. Because the antibiotic has to stay active during the whole sophorolipid synthesis period, samples were taken 2 and 4 days after the addition and the CFU were determined as described in Chapter II, Section 2.2.1.

### 2.2.3 Quantification of sophorolipid synthesis

Sophorolipids were extracted from the media as described in Chapter II, Section 2.2.3 (daily sampling) and Section 2.3 (end extraction).

HPLC analysis of the produced sophorolipids was performed as described in Chapter II, Section 2.4. In order to be able to compare and quantify the different samples, dilutions of a standard were analysed in parallel.

## 2.3 Results and Discussion

### 2.3.1 Evaluating the toxic effect of cerulenin

In order to test the effect of cerulenin on stationary *C. bombicola* cells, several concentrations of the antibiotic were tested on liquid cultures as described in the Materials and Methods Section 2.2.2. As sophorolipid synthesis occurs in the stationary phase and takes several days, the toxic effect of cerulenin was evaluated 2 and 4 days after addition (Figure V-10).

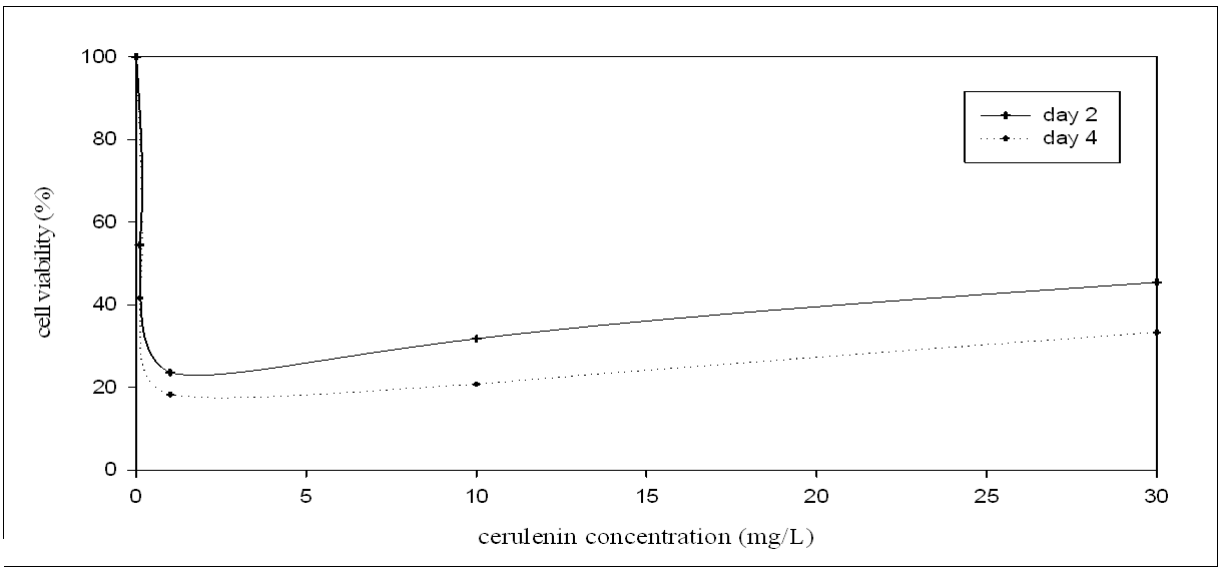


**Figure V-10:** Effect of cerulenin concentration on *C. bombicola* cell viability. Samples were taken 2 and 4 days after addition of cerulenin.

The strongest effect of cerulenin is observed at 1 mg/L. At this concentration, the CFU number is reduced to 24 and 18 %, respectively 2 and 4 days after the supplementation (these values correspond with reductions of $5.5x10^9$ to $2.5x10^9$ CFU/mL and $6x10^9$ to $2x10^9$ CFU/mL respectively). Further augmentation of the cerulenin concentration did not result in a stronger toxic effect; even at a concentration of 30 mg/L no additional lethal effect was observed. However, one can observe that cell viability declines upon longer incubation times (Figure V-10; compare the samples taken after 2 and 4 days of incubation). This decline in cell viability should not affect sophorolipid production as it has been demonstrated in Chapter

II that sophorolipid synthesis still is possible despite similar or even higher reductions of the CFU number.

Although a concentration of 1 mg/L seems sufficient for inhibiting fatty acid synthesis, it was decided to apply 20 mg/L because overdosing did not harm to the cells and because it was possible that part of the hydrophobic cerulenin would be absorbed by the sophorolipid molecules and/or the hydrophobic carbon source, thus reducing the availability for the yeast cells.

### 2.3.2 Effect of cerulenin on sophorolipid synthesis

When 20 mg/L cerulenin was added to a *C. bombicola* culture, hardly any sophorolipid could be detected except at the last day of the incubation period, *i.e.* 7 days after the addition of cerulenin. However, they only represented less then 1 % of the sophorolipids found in the control culture without addition of cerulenin. These finding clearly illustrate the inhibitory effect of cerulenin on the *de novo* synthesis of sophorolipids; when sophorolipids are produced in a medium containing merely glucose and no lypophilic carbon source, the necessary fatty acids for *de novo* sophorolipid production are 100 % contributed by the fatty acid synthase protein complex.

In a further experiment, cerulenin was used in combination with the hydrophobic carbon source 1,12-dodecanediol. Fermentations run on this hydroxylated substrate give rise to medium-chain sophorolipids, but at the same time a noticeable level of *de novo* sophorolipids can be detected. This event is discussed in Chapter II, Section 3.2.2.1. Furthermore, 1,12-dodecanediol exposes only a minor toxicity towards *C. bombicola* cells (which are already challenged by the cerulenin antibiotic). It also remains in the solid state in the culture medium due to its melting temperature of about 80 °C. This latter fact makes absorption of the hydrophobic cerulenin less plausible.

No *de novo* sophorolipids were detected in the culture with addition of cerulenin. The HPLC-profile differs somewhat from those previously obtained in Chapter II, Figures II-7 and II-8. Instead of several hydrophilic peaks, only one major hydrophilic peak is detected in the sample of the end extraction, making it quite similar to the HPLC profile of the ethylacetate

extracted samples (Figure II-8). Unfortunately, something went wrong with the control without addition of cerulenin: no sophorolipid production at all was observed and this while the CFU numbers always stayed higher than those of the cerulenin culture. It therefore is difficult to verify if the somewhat different HPLC chromatogram is caused by the cerulenin or is just a coincidence. For fermentations on rapeseed oil for example one sometimes can observe differences in the relative importance of the HPLC peaks amongst fermentations despite the application of the same culture conditions.

# 3 Conclusion

Synthesis of medium-chain sophorolipids by *C. bombicola* is a challenging objective. First of all, these sophorolipids can only be obtained by fermentations on unconventional carbon sources which often have a toxic effect on the cells. Furthermore, medium-chain hydroxylated substrates and substrates with an internal ester bond and alkyl endings are partially metabolized in the β-oxidation pathway. Finally, the obtained sophorolipids always represent a mixture of medium-chain and native *de novo* formed sophorolipids.

In order to redirect unconventional substrates towards sophorolipid synthesis, the β-oxidation pathway was blocked. Since previous experiments in the lab with chemical inhibitors of the β-oxidation route did not result in improved use of the substrate, β-oxidation was tackled at the genome level. The multifunctional enzyme 2 was chosen as the target enzyme. In contrast to other enzymes contributing to the metabolization of fatty acids, *MFE2* is believed to occur as a single copy in the genome thus making the creation of *C. bombicola* strain with a blocked β-oxidation quite feasible.

The total gene sequence of the *C. bombicola MFE-2* (6033 bp) was cloned (GenBank accession number EU371724). The large protein of 884 amino acids shows a 61 % amino acid homology with the MFE-2 of *Y. lipolytica*. Surprisingly, the protein shows a better match with MFE-2's of filamentous fungi compared to yeast-like fungi. The obtained nucleotide sequence was used to construct a knock-out cassette. Homologous recombination turned out to be most efficient with large fragments of 1000 bp of the target gene flanking the selection marker. Those fragments are quite large as compared to the 40 bp regions which are sufficient for homologues recombination in *S. cerevisiae*. However, for unconventional yeasts larger fragments are indeed necessary (Wesolowski-Louvel *et al.*, 1988).

Several knock-out mutants with the correct geno- and phenotype were tested in a fermentation on 1-dodecanol. The wild-type strain incorporates this substrate into sophorolipids, but a large extent of it is metabolized in the β-oxidation route. All mutants with a blocked β-oxidation showed a 1.7 to 2.9 times higher production of sophorolipids, indicating that in those strains the substrate is redirected towards the sophorolipid synthesis. This experiment illustrates that the mutants can be used to improve the production and yield of sophorolipids on medium-

chain substrates. In future experiments, other substrates which are subjected to metabolization by the yeast will be evaluated. The best mutants will be selected and with those strains, larger scale fermentations will be performed in order to optimize the culture conditions and feeding of the substrate.

Somewhat unexpected results were obtained when running fermentations with the mutants on rapeseed oil. Since rapeseed oil is already very efficiently incorporated into sophorolipid molecules, it remains to be tested if the sophorolipid yield could be further increased by using mutants with a blocked β-oxidation pathway. Nevertheless, instead of equal or slightly better sophorolipid yields, inferior results were obtained for the mutants. The lower CDW and CFU values could indicate that the cells are somehow inhibited by rapeseed fatty acid derived products which accumulate due to the blocked β-oxidation. Thus, for production of sophorolipid on conventional substrates, the wild type strain should be preferred above the mutant strains.

Another pathway interfering with the formation of medium-chain sophorolipids is the synthesis of native *de novo* sophorolipids. In this way, the obtained sophorolipids are in fact always a mixture of medium- and long-chain ones. The fatty acid moiety of *de novo* sophorolipids is derived from the *de novo* synthesis of fatty acids by the fatty acid synthase complex. In contrast to the β-oxidation, this pathway is essential for cell viability and consequently can not be blocked by knocking out one of the involved genes. Therefore, it was decided to apply an inhibitor of fatty acid biosynthesis: the antibiotic cerulenin. In this way the cells could first go through their exponential growth phase without presence of the inhibitor and synthesise the fatty acids needed for assembling the cell membrane. Cerulenin was added in the stationary phase: the antibiotic effect on fully developed yeast cells is not that strong and sophorolipids are only produced in the stationary phase; there is no need for early addition of cerulenin. Cerulenin could indeed block synthesis of *de novo* sophorolipids in glucose grown cultures. Also for fermentations run on 1,12-dodecanediol, all detected sophorolipids originated from the hydrophobic substrate. These results clearly indicate that when blocking the *de novo* fatty acid synthesis, a more uniform mixture of medium-chain sophorolipids can be obtained.

There are however a few drawbacks on the use of cerulenin. As the antibiotic is very hydrophobic, it tends to absorb at oily substrates and as such the cells are only to a minor extent exposed to it. Consequently, not every hydrophobic carbon source can be combined with the supplementation of cerulenin. Furthermore, cerulenin is a very expensive antibiotic, making it only applicable to small laboratory-scale experiments. Finally, cerulenin also inhibits the human fatty acid synthase. This latter fact combined with the high cost renders the use of cerulenin not suitable for industrial applications.

One way to circumvent the use of cerulenin could be blocking the *de novo* fatty acid synthesis on the genome level, but as mentioned before, the fatty acid synthase cannot be targeted as it is essential for cell growth. Another way of eliminating the *de novo* sophorolipid synthesis is knocking out the cytochrome P450 monooxygenase(s) which hydroxylate(s) the fatty acids. However, those mutants must be provided with already hydroxylated substrates if sophorolipid production is wanted.

Blocking the *de novo* sophorolipid synthesis gives rise to a sophorolipid mixture with a reduced complexity. This simplifies further evaluation of the physico-chemical characteristics of medium-chain sophorolipids and makes additional purification steps unnecessary or less complex. Furthermore, a disabled *de novo* synthesis will result in the redirection of precursors and co-factors towards the medium-chain sophorolipids, thereby giving rise to better yields.

# Chapter VI: Genetic characterisation of the cytochrome P450 monooxygenase system of *Candida bombicola*

Part of this chapter has been published as:

Van Bogaert INA, Develter D, Soetaert W, Vandamme EJ (2007) Cloning and characterisation of the NADPH cytochrome P450 reductase gene (*CPR*) from *Candida bombicola*. FEMS Yeast Res 7, 922-928

Van Bogaert INA, Develter D, Soetaert W, Vandamme EJ (2008) New members of the cytochrome P450 monooxygenase CYP52 family obtained from the yeast *C. bombicola*. FEMS Yeast Res, in preparation

# 1 Importance of the cytochrome P450 monooxygenase family CYP52 for the sophorolipid producing yeast *C. bombicola*.

## 1.1 Introduction

Cytochrome P450 monooxygenases (P450) are heme-thiolate proteins involved in the hydroxylation of a wide range of endogenous and xenobiotic compounds. They are present in all eukaryotic organisms and in a limited number of prokaryotic species. P450's are mainly studied for their role in steroid biosynthesis and metabolization of drugs and xenobiotic compounds in the liver. Until now, over 7700 P450 genes are known and this number is still increasing thanks to the recent genome sequencing projects. Nelson (1998) developed a classification system based on structural homology of the amino acid sequences of the P450 genes.

Alkane assimilating yeasts such as *Yarrowia lipolytica*, *Candida tropicalis* and *C. maltosa* possess several specific P450's. These enzymes are classified in the CYP52 family and participate in terminal hydroxylation of alkanes. The derived alcohols are further converted to fatty acids and then enter β-oxidation. The CYP52 family also includes several P450's able to hydroxylate fatty acids such as CYP52A13 and CYP52A17 of *C. tropicalis* (Eirich *et al.*, 2004). CYP52's are microsomal class II P450's, meaning that they form a small electron transfer chain together with the NADPH cytochrome P450 reductase (CPR). Both enzymes are N-terminally anchored to the endoplasmatic reticulum and derived structures (Edwards *et al.*, 1991).

*Candida bombicola* is a non pathogenic yeast capable of growing on alkanes when supplied as the sole carbon source, indicating that in this organism CYP52 activity must be involved. Furthermore, *C. bombicola* is known to synthesize sophorolipids, a biosurfactant group of commercial interest consisting of the disaccharide sophorose to which one hydroxy fatty acid is linked (Asmer *et al.*, 1988). This hydroxy fatty acid is synthesized by the yeast itself via hydroxylation of a common fatty acid by cytochrome P450 mediated monooxygenase activity (Lottermoser *et al.*, 1996). Hence, the cytochrome P450 monooxygenase system plays a crucial role in both alkane degradation and sophorolipid synthesis of *C. bombicola*.

Since the evolutionary history of cytochrome P450 monooxygenase genes is characterized by several events of gene duplication and conversion, a broad diversity among these genes is also found within the genome of a single organism (Nerbert and Gonzalez, 1987; Nelson, 1999). For example, in *C. maltosa* eight *CYP52* genes have been reported, while for *Yarrowia lipolytica* at least 12 genes were found (Fickers *et al.*, 2005: Zimmer *et al.*, 1998). For this reason, it is suggested that in the genome of *C. bombicola* several CYP52 genes are also present.

This section describes the cloning and characterisation of three *CYP52* genes of *C. bombicola*. The *CYP52* genes are believed to play a key role in sophorolipid synthesis: their selectivity towards a fatty acid with an optimal chain length of 16 to 18 carbon atoms determines the structure of the sophorolipid molecules. In order to find out whether the *CYP52* genes of *C. bombicola* intervene in sophorolipid synthesis or in alkane metabolization, an expression analysis was performed.

## 1.2 Materials and Methods

### 1.2.1 Strains, plasmids and culture conditions

*Candida bombicola* ATCC 22214 was used in all experiments. When sophorolipid production was intended, the medium described by Lang *et al.* (2000) was used. To study how the yeast cells metabolize alkanes, media containing 1 % hexadecane, 1 % yeast extract and 0.1 % urea were used. Yeast cultures were incubated at 30 °C and 200 rpm.

All PCR products intended for sequence analysis were cloned into the pGEM-T® vector (Promega). *Escherichia coli* DH5α was used in all cloning experiments and was transformed as described by Sambrook and Russell (2001). *E. coli* cells were grown in Luria-Bertani (LB) medium (1 % trypton, 0.5 % yeast extract and 0.5 % sodium chloride) supplemented with 100 mg/L ampicillin and 40 mg/L X-gal if necessary. Liquid *E. coli* cultures were incubated at 37 °C and 200 rpm.

### 1.2.2 DNA and RNA isolation and sequencing

Yeast genomic DNA was isolated with the GenElute™ Bacterial Genomic DNA Kit (Sigma), while RNA isolation was conducted with the RNeasy Mini Kit from Qiagen. In both cases cell lysis was performed by incubation at 30 °C for 90 minutes with zymolyase (Sigma). The RNA was treated with the RNase-Free DNase set from Qiagen to remove contaminating genomic DNA.

Bacterial plasmid DNA was isolated with the QIAprep Spin Miniprep Kit (Qiagen). All DNA sequences were determined at the VIB Genetic Service Facility (Belgium).

### 1.2.3 Degenerate PCR

Part of the P450 genes of *C. bombicola* were amplified using degenerate primers (Table VI-1). PCR amplification was carried out with an initial denaturation of 94 °C for 4 min, 40 cycles of 94 °C for 30 s, 40 °C for 1 min, 72 °C for 2 min with 5 s increasing time and a final 7 min elongation at 72 °C.

### 1.2.4 Genome Walking

The unknown genomic DNA sequences upstream and downstream of the degenerate PCR fragments were identified by genome walking with the BD GenomeWalker™ Universal Kit (BD Biosciences) as described in Chapter III, Section 3.2.5. Four gene-specific primers (GSP) were designed for each gene and are listed in Table VI-1.

### 1.2.5 Sequence analysis

Sequences were analysed with the Clone Manager Professional Suite Software (Version 6.0). The BLAST program (Atschul *et al.*, 1997) was used for similarity searches in databases available on the NCBI website (http://www.ncbi.nlm.nih.gov). Multiple sequence alignments were made with the CLUSTAL W program (Higgins *et al.*, 1992). Subsequently, a phylogenetic tree was constructed in BioEdit using the Protein Maximum Likelihood (ProML) algorithm.

The hydropathy of P450 proteins was analysed using the method of Kyte and Doolittle (1982). A window size of 21 amino acids was used. Peaks with scores greater than 1.6 indicate possible transmembrane regions.

### 1.2.6 Reverse transcriptase real-time PCR

cDNA was synthesized with the QuantiTect Reverse Transcription Kit from Qiagen which is suitable for sensitive real-time two-step RT-PCR and eliminates contaminating genomic DNA. Real-time PCR was conducted with a Rotor-Gene™ 6000 from Corbett Life Science's and SYBR® Green detection. The reaction mixture consisted of 2 µL sample, 12.5 µL ABsolute™ QPCR SYBR® Green Mix (ABgene) and 200 nM forward and reverse primer in a total volume of 25 µL. PCR was performed by 15 min of enzyme activation at 95 °C and a 40 fold cycling of 15 s at 94 °C, 30 s at 60 °C and 30 s at 72 °C. The primers used for real-time detection can be found in Table VI-1. Samples and standards were run in duplicate and the sample signals for the P450 genes were normalized against the housekeeping gene coding for glyceraldehyde 3-phosphate dehydrogenase (GAPD, Chapter IV).

**Table VI-1:** Primers used for obtaining *Candida bombicola* P450 genes and reverse transcriptase real time PCR. All primers were obtained from Sigma Genosys.

| Name | Feature | Sequence |
|---|---|---|
| P450deg470For | Degenerate primer | 5'-ATHTTYACHYTBGAYGG-3' |
| P450Deg680For | Degenerate primer | 5'-GCYACBGARTTYYTRTTYGG-3' |
| P450DegHelixFor | Degenerate primer | 5´-GCHGGWMGAGAYACCACHGC-3' |
| P450DegHelixRev | Degenerate primer | 5'-GCDGTGGTRTCTCKWCCDGC-3' |
| P450DegHemeRev | Degenerate primer | 5'-TGDCCHARRCARATTCKWGGWCC-3' |
| P450f1up1 | 5' primary GSP for fragment 1 | 5'-TTCGGCAGTAATCTCCGAACCATCAC-3' |
| P450f1UpN | 5' nested GSP for fragment 1 | 5'-CATGACTGATTCTCAGGGTTCAAGCTCTC-3' |
| P450f1Down1 | 3' primary GSP for fragment 1 | 5'-TCAATTTACATGGTCCACAGGGACATC-5' |
| P450f1DownN | 3' nested GSP for fragment 1 | 5'-CCAATCGATGGGAGAACTTGAAGCTAGA-3' |
| P450f2Up1 | 5' primary GSP for fragment 2 | 5'-TCAGCAAACTGCTCACCGGTTACAAGA-3' |
| P450f2UpN | 5' nested GSP for fragment 2 | 5'-TCCATGCTGAGCAAGAGCAGGGTCAAG-3' |
| P450f2Down1 | 3' primary GSP for fragment 2 | 5'-CCTGACCGCTGGAACGAGTCAAGAGA-3' |
| P450f2DownN | 3' nested GSP for fragment 2 | 5'-TCCGGATGGGACTACATTCCTTTCAAC-3' |
| P450f3up1 | 5' primary GSP for fragment 3 | 5'-ATGATGCGTCTCCATTAGGGAAGACC-3' |
| P450f3UpN | 5' nested GSP for fragment 3 | 5'-CCTTGGCGTAGATAGCGGGATTCTTCG-3' |
| P450f3Down1 | 3' primary GSP for fragment 3 | 5'-TTCCCAAGGGCACTGCTGTCAACTAC-3' |
| P450f3DownN | 3' nested GSP for fragment 3 | 5'-TCTATGGTGCCGACGCGTACGAGTTC-3' |
| GAPDqFor | Real-time PCR primer | 5'-TCATCGCTCCGGACTATGCT-3' |
| GAPDqRev | Real-time PCR primer | 5'-TGAAGCGGCCGTGAGTAGA-3' |
| M1qFor | Real-time PCR primer | 5'-GCAGCACGTAGGCACAAGTG-3' |
| M1qRev | Real-time PCR primer | 5'-CCCAGTGGGAACGATTTCTC-3' |
| E3qFor | Real-time PCR primer | 5'-TCACCTTCAAATTCGAGGAGTTC-3' |
| E3qRev | Real-time PCR primer | 5'-AGTTGCCCAGCGACTGATTG-3' |
| N1qFor | Real-time PCR primer | 5'-TTCCACTCACCCTGGTAGTTCTC-3' |
| N1qRev | Real-time PCR primer | 5'-GCGCTGCCATGAATTGC-3' |

### 1.2.7 Nucleotide sequence accession numbers

The nucleotide sequences described in this paper have been deposited at the GenBank nucleotide database under the accession numbers EU315245 (GAPD), EU552419 (CYP52-M1), EU552420 (CYP52-E3) and EU552421 (CYP52-N1).

## 1.3 Results and Discussion

### 1.3.1 Isolation of the *C. bombicola CYP52* genes

As all yeast P450's able to hydroxylate alkanes or fatty acids can be found in the CYP52 family, it is more than likely that the P450's from *C. bombicola* involved in alkane metabolization and sophorolipid synthesis also belong to this group. Since P450 enzymes are highly conserved in specific regions such as the helix-I and the heme binding region, all CYP52 genes were aligned and degenerate primers were designed on the two regions mentioned above and on two additional regions where a strong conservation was observed. All possible degenerate primer combinations were tested, but the only ones resulting in the amplification of *CYP52* genes were P450Deg680For and P450DegHemeRev, and P450DegHelixFor and P450DegHemeRev (Table VI-1), giving rise to PCR-products of about 820 and 460 bp respectively. The amplicons were cloned into the pGEM-T® vector and after transformation of the construct in *E. coli*, several colonies were picked up for plasmid isolation, as it was quite possible that multiple P450 genes were amplified due to their presumed similarity and abundance in the genome combined with the degenerate characteristics of the primers.

Subsequent sequencing of clones obtained with primer combination P450Deg680For and P450DegHemeRev, yielded two different fragments of 848 and 818 bp, showing respectively a 50 % amino acid identity with CYP52A3-A from *C. maltosa* and a very high amino acid identity of 95 % with CYP52E2 from *C. apicola*. Also the primer combination P450DegHelixFor and P450DegHemeRev yielded two different fragments. Both were 467 bp long and the first one was closest related to CYP52A3-A of *C. maltosa* (51 % amino acid identity), while the other one showed highest homology with the CYP52F7 or ALK7 of *Yarrowia lipolytica* (53 % amino acid identity).

For further cloning of the genes, genome walking was performed with the BD GenomeWalker™ Universal Kit and the primers listed in Table VI-1. Total coding regions of 1617, 1560 and 1572 bp and part of the 5' and 3' regions of the first three genes were clarified. Nevertheless, the last gene could not be cloned completely following this strategy despite the use of several primers or primer combinations. By performing several attempts of

genome walking on this gene, fragments from five different, but very similar genes were obtained. Due to the very high similarity among those genes, it was impossible to complete the upstream sequence from the last gene with a primer based strategy.

### 1.3.2 Characterisation of the *CYP52* sequences

For the first gene, 1409 and 1173 bp of respectively the 5' and 3' region were determined as well. According to Hamilton (1987), highly expressed genes of *S. cerevisiae* show no G-residues in the 7 bases preceding the start codon. This phenomenon is also observed for this *C. bombicola P450* gene, just like for its *URA3* and *GAPD* gene. 75 bp upstream the ATG start codon, a putative TATA signal TATATATT was found (Figure VI-1), corresponding with the TATA(A/T)A(A/T)(A/G) consensus sequence for yeast as described by Basehoar *et al*. (2004). The consensus sequence of the polyadenylation signal found in most eukaryotic genes is AATAAA, although for yeasts several variations are possible (Guo and Sherman, 1996). In the 3' region of this gene of *C. bombicola,* a putative polyadenylation signal sequence (AAGAAA) was observed 20 bp downstream of the TAA stop codon.

The protein of 538 amino acids is 48 % identical to the alkane inducible CYP52-A2 from *C. tropicalis*, which is able to hydroxylate hexadecane, but not lauric acid (Seghezzi *et al.*, 1991). For other members of the CYP52-A subfamily, the similarity ranges between 40 and 47 %. Since the identity is higher than 40 %, this gene can be classified into the CYP52 family. However, this is lower than 55 %, so this protein was classified into a new subfamily and will from now on be referred to as CYP52-M1, according to the internationally used nomenclature system (Nelson, 1998).

```
   1  tttctcaaca gacttagatg attcaatcta aattctgagt gataggttga gcattagcac agtttggttt
  71  agatagtaac tgagtttccc agtgacgaaa ctctacctat ggtgtaagag tcttgaggct gaggttacag
 141  cgactcaaac tacagcgtgc atgagaaact cgctaatata tattagggtt gatgcagctt gcatatagag
 211  tcgggcgtta tttctccgtt tatcatattg aaaagtacat atatgttaat caaagacatt attctaactc
   1                                  CYP52-M1 → M  L   I  K  D  I   I  L  T
 281  caatgagttt atccgctgtt gctggcttgt tgccactgct cttcgtagct ttcttagttc tacacgagcc
  10   P  M  S   L  S  A  V   A  G  L   L  P  L   L  F  V  A   F  L  V   L  H  E
 351  tatctggctc ctatggtacc gctatgcagc acgtaggcac aagtgtagta tgcctcgctt cattgagaaa
  33  P  I  W  L   L  W  Y   R  Y  A   A  R  R  H   K  C  S   M  P  R   F  I  E  K
 421  tcgttcccac tgggaataca aagaaccatg gacatgatca agacggccaa gtcatacacc ttactggaag
  57    S  F  P   L  G  I   Q  R  T  M   D  M  I   K  T  A   K  S  Y  T   L  L  E
 491  ttcaatacga cagagtcttc aataagttca aagcacggac gtatcttcga caagctcccc ttcaatacca
  80   V  Q  Y   D  R  V  F   N  K  F   K  A  R   T  Y  L  R   Q  A  P   L  Q  Y
 561  aatcttcaca atcgagccag aaaacattaa gacaatcctg gcaaccaaat tcaatgattt tggtcttgga
 103  Q  I  F  T   I  E  P   E  N  I   K  T  I  L   A  T  K   F  N  D   F  G  L  G
 631  gcacgtttcc acacagtggg aaaagtgttt ggccaaggga tatttacact cagcggaaat ggatggaaac
 127    A  R  F   H  T  V   G  K  V  F   G  Q  G   I  F  T   L  S  G  N   G  W  K
 701  agtctcgatc gatgttgaga cctcagttca ctaaagatca ggtttgcaga attgatcaga tttccagtca
 150   Q  S  R   S  M  L  R   P  Q  F   T  K  D   Q  V  C  R   I  D  Q   I  S  S
 771  tgctgcggag ttaataaagg agatgaaccg tgcaatgaaa gtggaccaat ttattgatgt tcaacattat
 173  H  A  A  E   L  I  K   E  M  N   R  A  M  K   V  D  Q   F  I  D   V  Q  H  Y
 841  ttccacaaac ttacgctgga tacagcgact gaattcctat ttggggagtc ctgcgagagc ttgaaccctg
 197    F  H  K   L  T  L   D  T  A  T   E  F  L   F  G  E   S  C  E  S   L  N  P
 911  agaatcagtc atgtattgta gcccgtgatg gttcggagat tactgccgaa caattcgtgg agtcctacaa
 220   E  N  Q   S  C  I  V   A  R  D   G  S  E   I  T  A  E   Q  F  V   E  S  Y
 981  ctttctactg aattacgctt tcaaacggac cctatcaagc aaagtctact ggttgttcaa ctctaaggaa
 243  N  F  L  L   N  Y  A   F  K  R   T  L  S  S   K  V  Y   W  L  F   N  S  K  E
1051  ttccgagatc acaagaaacg tgctcagtcc tatattgact actacgttga taaggctctt tacgccacat
 267    F  R  D   H  K  K   R  A  Q  S   Y  I  D   Y  Y  V   D  K  A  L   Y  A  T
1121  ctttcgctgc tgagaactct attgcagaga aggatgctgc tgcagagtct agtggcatct atgtgttctc
 290   S  F  A   A  E  N  S   I  A  E   K  D  A   A  A  E  S   S  G  I   Y  V  F
1191  gcttgagatg gctaaagtta cccgagaccc agtgacgata cgtgatcaaa ttttcaacat tctcattgct
 313  S  L  E  M   A  K  V   T  R  D   P  V  T  I   R  D  Q   I  F  N   I  L  I  A
1261  ggtagagata caacagctgc tacgttgagc ttcgctattc atttccttgc cagaaatcct gacgtattca
 337    G  R  D   T  T  A   A  T  L  S   F  A  I   H  F  L   A  R  N  P   D  V  F
1331  acaaactacg tgaggaggtc ctcgatcatt ttggaaccaa ggaggagcaa aggcctttat cattcgaact
 360   N  K  L   R  E  E  V   L  D  H   F  G  T   K  E  E  Q   R  P  L   S  F  E
1401  tctgaagcaa gcaccttatt tgaagcaagt tataaatgaa gtcttgcgtc ttgcgccggt attgccattg
 383  L  L  K  Q   A  P  Y   L  K  Q   V  I  N  E   V  L  R   L  A  P   V  L  P  L
1471  aacttccgta ctgctgtgag agatacaact ctacccatag gtggtggtcc cgagcagaag gatccgatct
 407    N  F  R   T  A  V   R  D  T  T   L  P  I   G  G  G   P  E  Q  K   D  P  I
1541  tcgttcctaa gggcaccgca gtttactatt caatttacat ggtccacagg gacatcaagt attggggtcc
 430   F  V  P   K  G  T  A   V  Y  Y   S  I  Y   M  V  H  R   D  I  K   Y  W  G
1611  tgacgcccac gaattcaatc ccaatcgatg ggagaacttg aagctagata atgtgtgggc attcttgccc
 543  P  D  A  H   E  F  N   P  N  R   W  E  N  L   K  L  D   N  V  W   A  F  L  P
1681  ttcaatggcg gtccccgaat ttgtctcggc caacaattcg ccctgacaga gctttcgcta actctggtga
 477    F  N  G   G  P  R   I  C  L  G   Q  Q  F   A  L  T   E  L  S  L   T  L  V
1751  gactcttaca ggagtattcc aagattgaga tgggtcccga cttcccagag tcccctcgtt tctcaacaac
 500   R  L  L   Q  E  Y  S   K  I  E   M  G  P   D  F  P  E   S  P  R   F  S  T
1821  gcttacagct caacacgctc ctcccggtgt ggttgtgcgg ttttcttaag tttcttagcc tcccatggaa
 523  T  L  T  A   Q  H  A   P  P  G   V  V  V  R   F  S  *
1891  gaaacgttcc ctccttaatt ggttcaatca acccatgtta actcaacctg tggcaacctt tttattattc
1961  tgagcactct actcaagaat ggctgcagtt ctttatttat gctctttatg aagaatcatg taaatcatgt
```

**Figure VI-1:** Coding sequence of the *CYP52-M1* gene and part of the5' and 3' regions. The supposed TATA-box is underlined and sequences presumed to be involved in polyadenylation are indicated in bold. The highly conserved amino acid sequences of the helix-I and the heme domain are marked in bold and underlined cases.

For the second fragment a total sequence of 3816 bp was determined. Also in this 5' region a possible TATA box was found: the sequence TATATAAA 173 bp upstream of the translation start (Figure VI-2). The specific pentanucleotide sequence CACAT and its complement ATGTG are counted five times between the positions -321 and -85. This sequence occurs in the promoter region of cytochrome P450 monooxygenase genes of yeasts (Yadav and Loper, 1999). No clear polyadenylation signal was found for this sequence.

The second gene of 519 amino acids shows a 91 % match with CYP52-E2 and a 82 % match to CYP52-E1 from *C. apicola*, consequently it was added to this subfamily and called CYP52-E3. It is worth mentioning that *C. apicola* is also a sophorolipid producing yeast strain, closely related to *C. bombicola*. Nevertheless, the function or biochemical activity of the CYP52-E2 and –E1 gene product is not clarified and southern hybridization experiments revealed the existence of further P450-related genes in *C. apicola* (Lottermoser *et al.*, 1996); so it remains unclear whether or not these genes are directly involved in sophorolipid production or not.

```
   1  ctgacattat gctacttcga aatgttttta ataatttatc taaaaatatc ctactaggtc gaaccacccg
  71  catcgtcaac attcgcggga catcggacac aatttccgtg cttcactcaa tgcgttcttc acatgcatga
 141  ttgatttcac attgcgttgc atcgaaattg agccgcacac gtgattaggt ggccaatcac atgaggcccc
 211  tgttgcgggg tatgtgtgtt atcgtcccca gaaacccaat ccagcgtgtg agaattcata tataaagctg
 281  atttgcggtt tttagactgg tcaatcgaaa tccgccggta ttgaccagat tagattggaa aattctcaga
 351  acaagacaca tcacacccag agctattgtt ggcaagcaat actattttta acccccttc caagagtgtc
 421  caaactgttc aatacggcct catgaacatt aatttctctg acgtgctcgt gctaggaggc atcagcgtga
   1              CYP52-E3 → M  N  I   N  F  S   D  V  L   V  L  G  G   I  S  V
 491  gctttttgct cgcctaccag gcgatttact tttatttcat ttactcgcca cgagccaaaa agctcggttg
  17   S  F  L   L  A  Y  Q   A  I  Y   F  Y  F   I  Y  S  P   R  A  K   K  L  G
 561  cgctcttcca ccggtcttct tctctttccc actcggaata ccggaggtca tacgtcttgt gaacgcctgg
  40  C  A  L  P   P  V  F   F  S  F   P  L  G  I   P  E  V   I  R  L   V  N  A  W
 631  ttcaacgatg atctccttga gtatttcacc ttcaaattcg aggagttcca gcgcaaaacc ggattccaat
  64    F  N  D   D  L  L   E  Y  F  T   F  K  F   E  E  F   Q  R  K  T   G  F  Q
 701  cagtcgctgg gcaactatgg attgggacta ttgagcccga gaacatcaag actatgctcg ctacttcatt
  87   S  V  A   G  Q  L  W   I  G  T   I  E  P   E  N  I  K   T  M  L   A  T  S
 771  taaagactac tccctaggct tccgttacga ggccatgtac ggccttctcg gaaatggcat tttcactctc
 110  F  K  D  Y   S  L  G   F  R  Y   E  A  M  Y   G  L  L   G  N  G   I  F  T  L
 841  agtggtgagg gctggaagca cagccgcgct ttgttgcgtc cgcaatttag tcgtgagcaa gtctctcacc
 134    S  G  E   G  W  K   H  S  R  A   L  L  R   P  Q  F   S  R  E  Q   V  S  H
 911  ttgaatcaat gcgcacacac atcaatatgt tgatcaacaa ccacttcaag ggtggcaaag tcgtcgatgc
 157   L  E  S   M  R  T  H   I  N  M   L  I  N   N  H  F  K   G  G  K   V  V  D
 981  tcaggttttg ttccacaatc taaccattga tactgctacc gaattcctat tcggagagag caccaacact
 180  A  Q  V  L   F  H  N   L  T  I   D  T  A  T   E  F  L   F  G  E   S  T  N  T
1051  cttgaccctg ctcttgctca gcatggattc cctggaccta agggtcttgt aaccggtgag cagtttgctg
 204    L  D  P   A  L  A   Q  H  G  F   P  G  P   K  G  L   V  T  G  E   Q  F  A
1121  aggcttttac ctctgctctc gaattgcttt ctgtgcgagt tatggccggc gccgcatggt tcctcgtttg
       E  A  F   T  S  A  L   E  L  L   S  V  R   V  M  A  G   A  A  W   F  L  V
1191  gacccccaaa ttctggcgct catgcaaagt ctgccacaac ttcattgatt acttcgtttt caaggctctg
      W  T  P  K   F  W  R   S  C  K   V  C  H  N   F  I  D   Y  F  V   F  K  A  L
1261  gccactccta tggagaagga ccaggaagct gatcgctacg tctttattcg agaactcaca aaggagacct
        A  T  P   M  E  K   D  Q  E  A   D  R  Y   V  F  I   R  E  L  T   K  E  T
1331  ctgacccacg ggtcatccgc gaccaggccc tcaacatcct cttggctggt cgtgatacca ctgcggcact
       S  D  P   R  V  I  R  D  Q  A  L  N  I   L  L  A  G   R  D  T   T  A  A
1401  tctcagcttc accacctact accttggtgc ctaccctgag gtctacgatg agcttcgcga ggctgttatt
      L  L  S  F   T  T  Y   Y  L  G   A  Y  P  E   V  Y  D   E  L  R   E  A  V  I
1471  gcggacttcg gcaaggaaga tgctgagccc cctacgtttg agcagcttaa gcagtgcaag gtgctacaga
        A  D  F   G  K  E   D  A  E  P   P  T  F   E  Q  L   K  Q  C  K   V  L  Q
1541  acgtcattcg ggaagttttg cgattgcacc cgaatgtgcc cctcaacttc cgcgaggcca ttaccgatac
       N  V  I   R  E  V  L   R  L  H   P  N  V   P  L  N  F   R  E  A   I  T  D
1611  taagttcccc acaggaggcg gcccgaatgg agaccagccc gttttcgttc ccaagggaca gaaagtgttt
      T  K  F  P   T  G  G   G  P  N   G  D  Q  P   V  F  V   P  K  G   Q  K  V  F
1681  tacgccacct acgtcatgca gcgaaatgag ggtctctggg gtcctgactc cacaacattc cgccctgacc
        Y  A  T   Y  V  M   Q  R  N  E   G  L  W   G  P  D   S  T  T  F   R  P  D
1751  gctggaacga gtcaagagag gccatcgcat ccggatggga ctacattcct ttcaacggcg gccctcgtat
       R  W  N   E  S  R  E   A  I  A   S  G  W   D  Y  I  P  F  N  G   G  P  R
1821  ttgcctgggt cagcagttcg ctctcacaga ggcgagctac acgctcgtgc gtatctgcca agagttctcc
      I  C  L  G   Q  Q  F   A  L  T   E  A  S  Y   T  L  V   R  I  C   Q  E  F  S
1891  aggattgagg ttctccaccc tgatgttatt acctccagga acgtgatgaa acagcgcatg cgtttgacca
        R  I  E   V  L  H   P  D  V  I   T  S  R   N  V  M   K  Q  R  M   R  L  T
1961  actcttccag cggcggcgtc atagcgaagt tcattcgcta gttctgtgct tcgagcatac atatcatgag
       N  S  S   S  G  G  V   I  A  K   F  I  R   *
```

**Figure VI-2:** Coding sequence of the *CYP52-E3* gene and part of the 5' and 3' regions. Possible promoter elements such as the TATA-box and CACAT/ATGTG sequences are underlined. The highly conserved amino acid sequences of the helix-I and the heme domain are marked in bold and underlined cases.

The third gene also possesses a conventional yeast TATA-box: TATAAAAG at position -84 and here the specific CACAT sequence is present at -96 (Figure VI-3).

The corresponding gene product of 523 amino acids is 45 % identical to CYP52-A3-A of *C. maltosa*. The similarity to other proteins of the CYP52-A subfamily is situated between 39 and 44 % and also the newly described CYP52-M1 can be found in the same range with a 43 % match. Therefore this protein of 523 amino acids is the first member, CYP52-N1, of a new subfamily.

Together with the coding region for *CYP52-N1*, 1331 bp of the region preceding the gene and 1005 bp of the downstream region were determined. In both regions open reading frames were found: in the upstream region the putative protein matched with the F-box protein HRT3 (High level expression Reduces Ty3 transposition protein 3) from *S. cerevisiae* (22 % amino acid based). This should be a component of an ubiquitin ligase complex taking part in protein ubiquitinylation and degradation. The other predicted protein is more than 50 % identical to putative and proven Vesicle transport v-SNARE proteins. Those small proteins are required for protein trafficking between eukaryotic organelles. The v-SNAREs on transport vesicles interact with t-SNAREs on target membranes in order to facilitate this.

For none of the three gene products, there were indications for the presence of introns.

```
   1  tcatactatt ttcataagct gaggtgatgg cggtttcacc tgtaccctcg taaaccttca cggcaagctt
        S  Y  Y   F  H  K   L  R  *  ← putative F-box protein
  71  gcgatgatcc cgggcccttt cattcccttt cctcagttcc cgtagaattc tcgagaccaa caatcattca
 141  atgatatcgt gccacttacc gaggtaaccc cctcacccct ttctattcac aaggcctgta cattgccgtt
 211  ttgatacttc cccaaccctt cagggcacat acatttctat aaaagcatgt agaccgtgag ctgatggtgc
 281  cgcggtaact gaaaatcaaa tattctcctt tcattcgaat ctatattcac aatgattctt tatgctgtgc
   1                                                  CYP52-N1 → M  I  L   Y  A  V
 351  tgggcgcatt cgccgccttc ttgctttaca tggatgtact ttacccttc gtgatttacc ctctgagagc
   7   L  G  A   F  A  A  F   L  L  Y   M  D  V   L  Y  P  F   V  I  Y   P  L  R
 421  gcgatggcac aaatgtggtt acatccctag agatttgagc tggccattgg ggattccact caccctggta
  30  A  R  W  H   K  C  G   Y  I  P   R  D  L  S   W  P  L   G  I  P   L  T  L  V
 491  gttctctcga agttgaggaa agatatgctg ctgcaattca tggcagcgca agaccttagt cgcccttaca
  54   V  L  S   K  L  R   K  D  M  L   L  Q  F   M  A  A   Q  D  L  S   R  P  Y
 561  agacatcctt acgtcaattt ctgggtaaat gggtaatcgc cactagagat cctgagaaca tcaaggctgt
  77   K  T  S   L  R  Q  F   L  G  K   W  V  I   A  T  R  D   P  E  N   I  K  A
 631  tctatccacc aagttcaatg acttctcgct gaaagaaaga gggaatagga tgaggcatgt aatcggtgat
 100  V  L  S  T   K  F  N   D  F  S   L  K  E  R   G  N  R   M  R  H   V  I  G  D
 701  ggaattttta cccaagatgg cgcaccatgg aagcactcgc gagatatgct caggcctcag ttcaccaagg
 124   G  I  F   T  Q  D   G  A  P  W   K  H  S   R  D  M   L  R  P  Q   F  T  K
 771  atcaaatcag ccgagtggaa ttgttgagcc accacatcga cgttttgatt cgtgaaatca ggaagtcggg
 147   D  Q  I   S  R  V  E   L  L  S   H  H  I   D  V  L  I   R  E  I   R  K  S
 841  aggtaacgtc gagttgcaac gtttattcca cctcatgact atggacaccg ccactcactt tctattcggc
 170  G  G  N  V   E  L  Q   R  L  F   H  L  M  T   M  D  T   A  T  H   F  L  F  G
 911  gagtccgttg gctcgttgga ggtcagtggc gaaagcaagg gcattgagat caccgaccca aagactggag
 194   E  S  V   G  S  L   E  V  S  G   E  S  K   G  I  E   I  T  D  P   K  T  G
 981  agattgtgaa caccgttgat tttgttgagt cttatacttt tgcaaacaag tttgctctca agaagattat
 217   E  I  V   N  T  V  D   F  V  E   S  Y  T   F  A  N  K   F  A  L   K  K  I
1051  cctcaacgac ttggagtttt tagccgactt gacggagccc tcgtataagt ggcatctgcg ccgtgtccac
 240  I  L  N  D   L  E  F   L  A  D   L  T  E  P   S  Y  K   W  H  L   R  R  V  H
1121  acagtcatgg atcactacgt tcagctggct ttgaaggcta ctgagaagta tgatcctgat gatgatagcg
 264   T  V  M   D  H  Y   V  Q  L  A   L  K  A   T  E  K   Y  D  P  D   D  D  S
1191  agaagggaga atactacttt agccatgagc tggcgaaact cacgagagac cccttgtcgt tgagagatca
 287   E  K  G   E  Y  Y  F   S  H  E   L  A  K   L  T  R  D   P  L  S   L  R  D
1261  gcttttcaat attctcattg ctggccgcga cactaccgca gcaactttgt cctatgcctt ccactatcta
 310  Q  L  F  N   I  L  I   A  G  R   D  T  T  A   A  T  L   S  Y  A   F  H  Y  L
1331  acgaagaatc ccgctatcta cgccaaggtc cgcgaagatg tgctcacggt cttccctaat ggagacgcat
 334   T  K  N   P  A  I   Y  A  K  V   R  E  D   V  L  T   V  F  P  N   G  D  A
1401  cattggcgac ttacgaggac ttgcgaaagg ctaagtatct ccaaatggtg atcaaggagg tattgcgtct
 357   S  L  A   T  Y  E  D   L  R  K   A  K  Y   L  Q  M  V   I  K  E   V  L  R
1471  tgcgcctgcg gttcccttga acacgcgtgc cgcggttcgt gacacatatc tgccacgggg cggaggccca
 380  L  A  P  A   V  P  L   N  T  R   A  A  V  R   D  T  Y   L  P  R   G  G  G  P
1541  gccggaaacc tgcccgtttt tgttcccaag ggcactgctg tcaactaccc tacatatatt ttgcaccgcg
 404   A  G  N   L  P  V   F  V  P  K   G  T  A   V  N  Y   P  T  Y  I   L  H  R
1611  atccagatat ctatggtgcc gacgcgtacg agttcaaccc cgagagatgg aggcctgaga ataagcttcc
 427   D  P  D   I  Y  G  A   D  A  Y   E  F  N   P  E  R  W   R  P  E   N  K  L
1681  gaatagccca atgtactctt ggggatacat tcccttcaat ggtggccctc gcatctgcat tggacagcag
 450  P  N  S  P   M  Y  S   W  G  Y   I  P  F  N   G  G  P   R  I  C   I  G  Q  Q
1751  ttcgccttga ctgagatcgc tttgacgatg atcaagctgg ttctggaatt tgagaggctg gagcctgccg
 474   F  A  L   T  E  I   A  L  T  M   I  K  L   V  L  E   F  E  R  L   E  P  A
1821  acgactttga gcccaatctt caagacaagt cctctttaac tgtcatggtc ggagggtcgg gcgtccgagt
 497   D  D  F   E  P  N  L   Q  D  K   S  S  L   T  V  M  V   G  G  S   G  V  R
1891  gaaactgagt taagtctctt tcgccttcca accctatggg ttcttttgtt cggcacatta tccagtctgt
 520  V  K  L  S   *
```

**Figure VI-3:** Coding sequence of the *CYP52-N1* gene and part of the 5' and 3' regions. Possible promoter elements such as the TATA-box and CACAT/ATGTG sequences are underlined. The highly conserved amino acid sequences of the helix-I and the heme domain are marked in bold and underlined fonts.

Microsomal cytochrome P450 enzymes like those of the CYP52 family are believed to be N-terminally anchored to the endoplasmatic reticulum and its related structures. This is confirmed for the three genes when conducting a Kyte-Doolittle hydropathy plot (Kyte and Doolittle, 1982). As an example, the hydropathy plot of CYP52-M1 is given in Figure VI-4, which clearly shows a N-terminal transmembrane region.

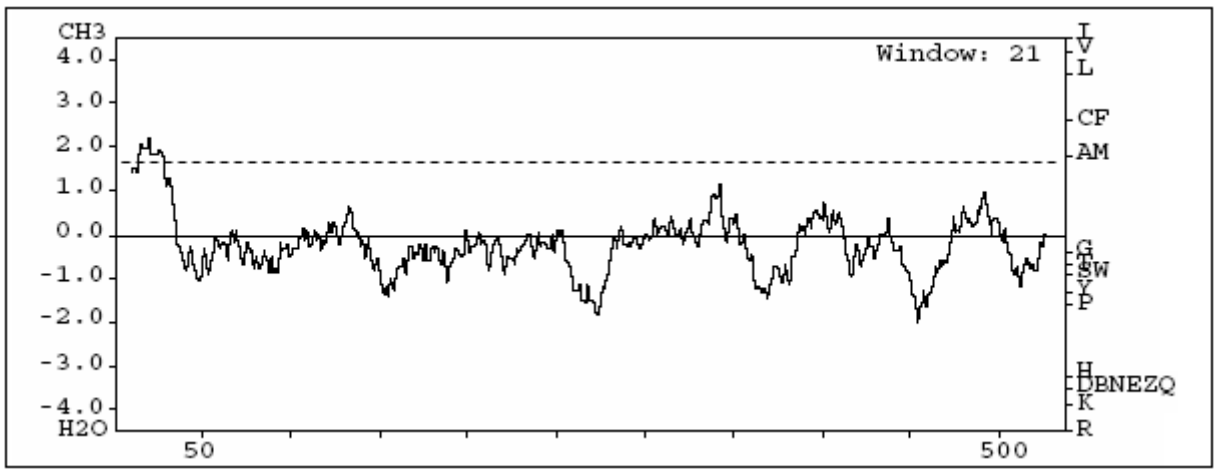


**Figure VI-4:** Kyte-Doolittle hydropathy plot with a window size of 21 amino acids for the CYP52-M1 protein. Regions with a score higher than 1.6 are believed to be transmembrane regions.

The three new amino acid sequences are aligned with other members of the CYP52 family in Figure VI-5. Cytochrome P450 monooxygenases are highly conserved in the regions involved in heme-binding. Residues 507 to 527 form the proximal heme-binding region and contain a cysteine residue as the fifth ligand to the iron molecule (position 516). Furthermore, the conserved phenylalanine residue at position 509 and the glycine at position 512, taking part in heme binding and formation of the cysteine pocket, are present. Also the distal heme-binding region or helix-I region with its conserved threonine residue (position 362) comes forward from the alignment (position 348 to 370; Hasemann *et al.*, 1995). Finally, the three arginine residues likely to interact with the heme propionates can be observed at positions 169, 425 and 514 (Nelsone and Strobel, 1989).

The alignment reveals that CYP52M1 and CYP52N1 from *C. bombicola* tend to cluster together in their differences towards the other sequences (*e.g.* at positions 152, 177, 179, 183, 188, 260, 262 and 263). This trend is confirmed when the sequences are put into a phylogenetic tree (Figure VI-6). The tree was constructed based on the protein maximum

likelihood principle and comprises one CYP52 protein from each subfamily. The sequence of CYP4A1 from *Rattus norvegicus* was used as outgroup and the tree was rooted against it. This mammalian P450 also shows activity towards fatty acids, but still differs enough from the fungal CYP52's. As expected, CYP52E3 is retrieved on a separate branch together with CYP52E1. Interestingly, also the other *C. bombicola* genes tend to cluster together although sequence identity among them is not extremely high (E3-M1: 40 %, E3-N1: 38 % and M1-N1: 43 %).

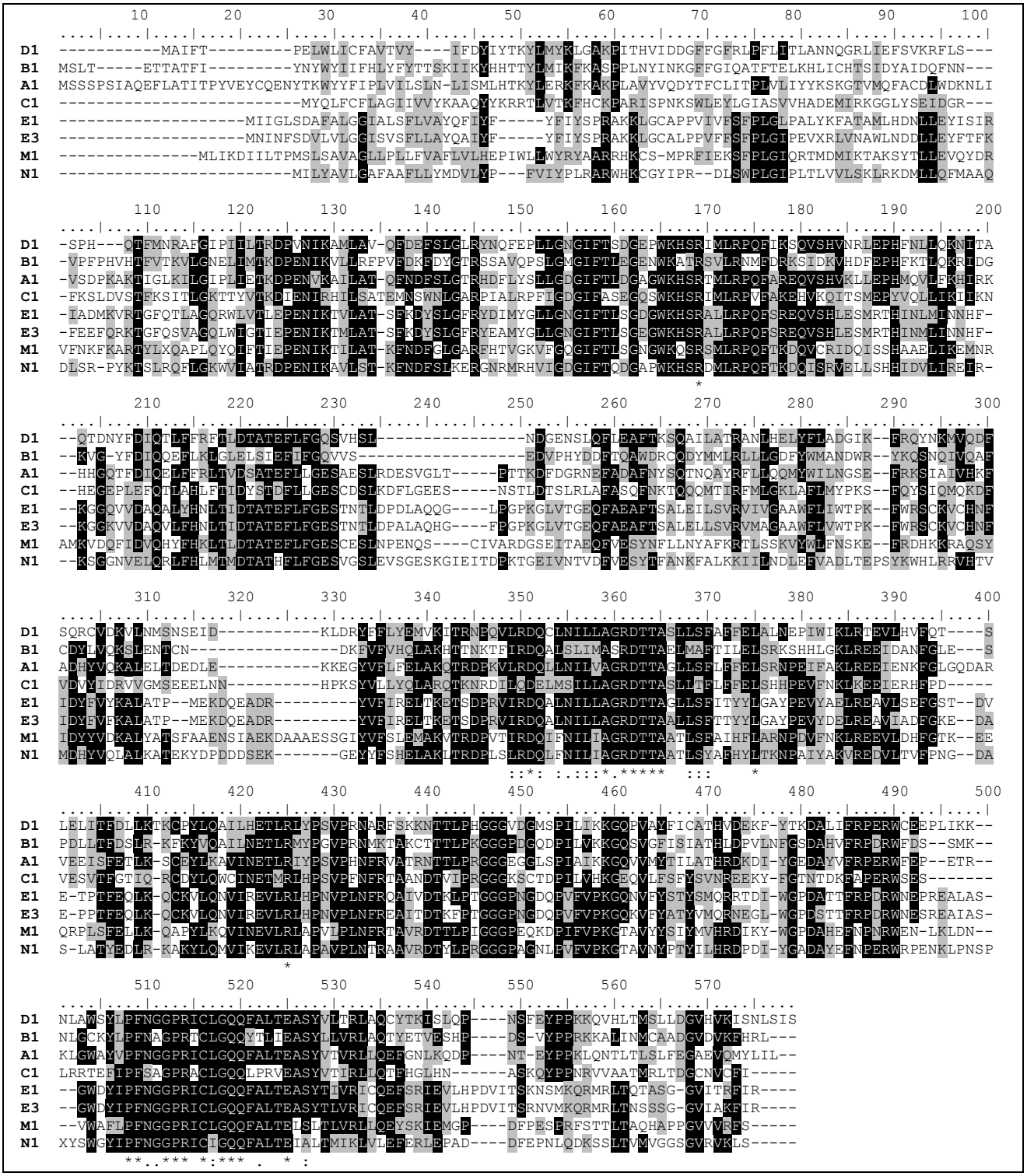


**Figure VI-5:** Alignment of the three new CYP52 proteins from *C. bombicola* (E3, M1 and N1) with CYP52A1, CYP52B1 and CYP52C1 from *C. tropicalis* (A1, B1 and C1), CYP52D1 from *C. maltosa* (D1) and CYP52E1 from *C. apicola*. The heme binding regions and other conserved amino acids as discussed in the text are marked.

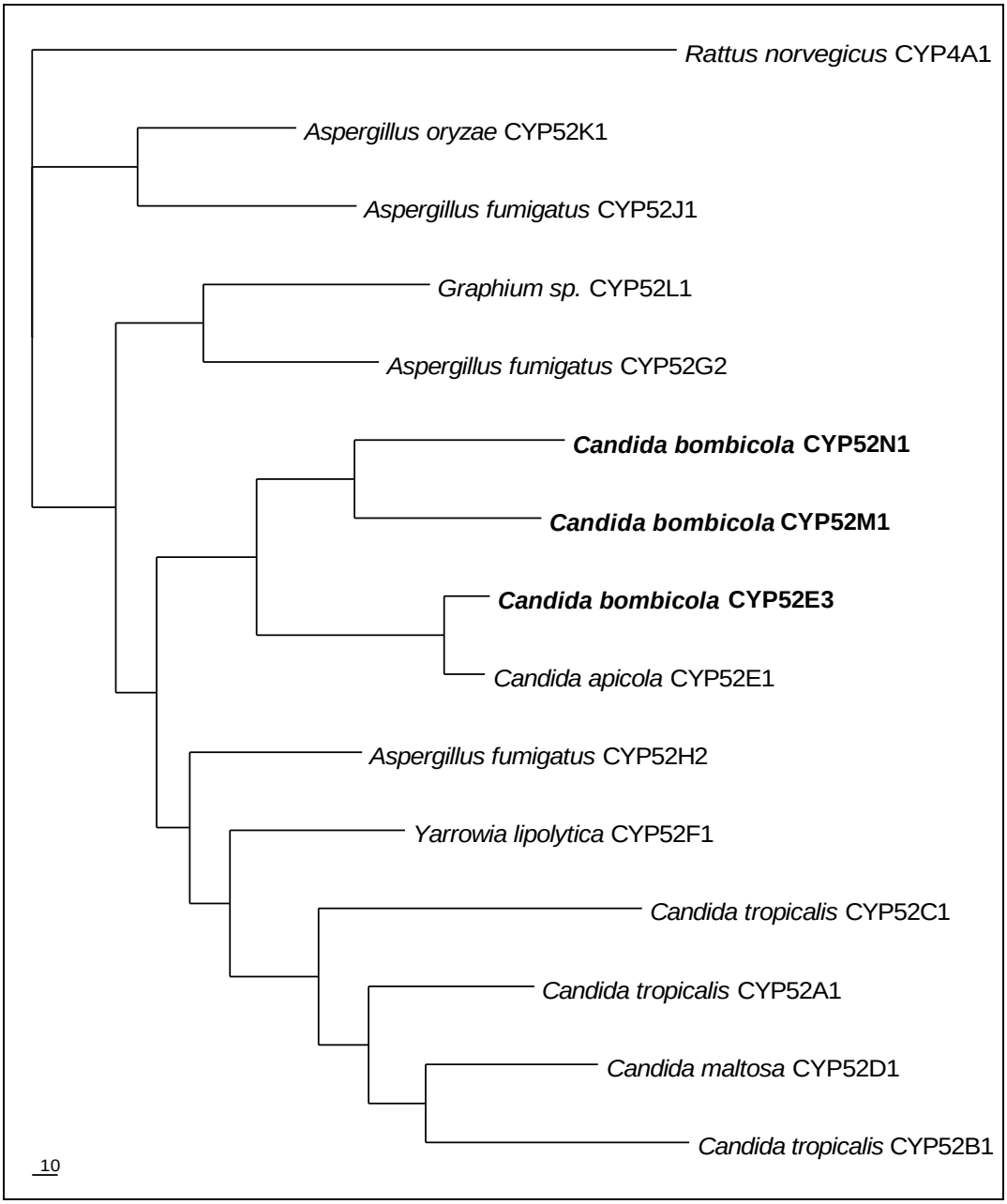


**Figure VI-6:** Phylogenetic tree for cytochrome P450 monooxygenases from the CYP52 family rooted against the *Rattus norvegicus* CYP4A1. The Ln Likelihood of the tree is -16100 and all knobs and branches were statistically significant ($p < 0.01$). The marker bar below denotes the integer branch length.

### 1.3.3 Expression profile of the isolated CYP52's from *C. bombicola*

The above experiments already demonstrated that also for *C. bombicola* multiple *CYP52* genes are present: three of them were completely isolated and there is evidence for at least five additional genes which are very similar to each other. Therefore, it can be expected that not all genes display a certain biochemical function, as can also be seen for *C. tropicalis*. For this yeast, 18 *CYP52* genes have been identified of which ten were further investigated and only five of them took actually part in biochemical reactions (Craft *et al.*, 2003). The experiment described here wants to shed light on the hypothetical function of CYP52E3, -M1 and N1, since *C. bombicola* needs CYP52 activity both for alkane assimilation as for sophorolipid synthesis and it is quite likely that different enzymes or enzyme groups are responsible for those reactions.

Therefore, a differentiating experiment for the isolation of mRNA was set up. Growth on alkanes as the sole carbon source will gather information on P450's necessary for alkane assimilation. Hexadecane was used for this purpose as biomass formation of *C. bombicola* is highest on this alkane (Inoue and Ito, 1982). When both glucose and fatty acids are added, information on the P450's involved in sophorolipid synthesis should become clear. In the latter experiment, the conditions for traditional sophorolipid fermentation are as follows: during the exponential growth phase only glucose is present, while rapeseed oil is only added after two days of cultivation, *i.e.* when the cells have reached stationary phase. A slight variant of the previous experiment is the addition hexadecane instead of rapeseed oil. For all situations, samples were taken in the exponential growth phase and stationary phase, as it is expected that the P450's involved in alkane assimilation are predominantly expressed in the exponential growth phase, while sophorolipid synthesis is only initiated in the stationary phase; therefore, it can be expected that induction of the responsible P450's also occurs during that phase.

The expression level was determined by a two-step reverse transcriptase real-time PCR with Sybr® Green detection. The primers used in this experiment are listed in Table VI-1, and were designed in this way that amplification of other know *C. bombicola CYP52* genes was impossible (the fragmentary sequences were also taken into account). The obtained signal was normalized against the expression of the housekeeping gene coding for glyceraldehyde-3-phosphate dehydrogenase (GenBank accession number EU315245, Chapter IV). The

corresponding enzyme takes part in the Embden-Meyerhof-Parnas pathway and shows a high and stable expression.

A graphical representation of the relative expressions of the genes can be seen in Figure VI-7. As the exponential phase sample for sophorolipid synthesis is the same for the rapeseed oil and alkane experiment (those are only added in the stationary phase), just one sample is shown. The most striking event is the high expression of CYP52M1 in the stationary phase after the addition of rapeseed oil or hexadecane when sophorolipids are produced. The expression level is almost the same for both substrates and is 30 fold higher as compared to the exponential phase where no sophorolipids are formed. The enzyme is of no importance for alkane assimilation: expression levels there are 258 to 584 times lower than for sophorolipid synthesis and are the lowest ones in the whole experiment. The other two *P450* genes do not seem to contribute to sophorolipid synthesis: the expression level is quite low and does not increase in the stationary phase.

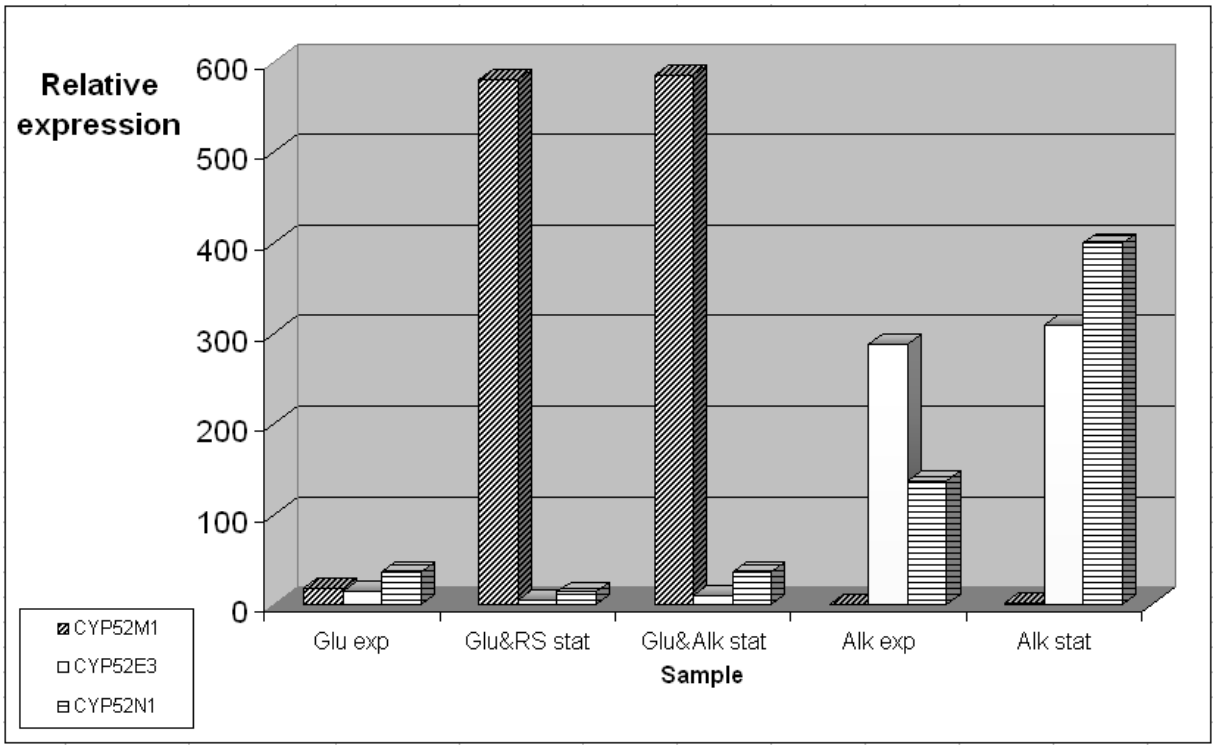


**Figure VI-7:** Relative expression of CYP52M1, -E3 and -N1 in different experimental conditions: exponential and stationary phase sample for sophorolipid production on glucose and rapeseed oil (**Glu exp** and **Glu&RS stat** respectively), stationary phase sample for sophorolipid production on glucose and hexadecane (**Glu&Alk stat**) and exponential and stationary phase sample for growth on alkanes (**Alk exp** and **Alk stat** respectively).

Furthermore, one can remark that the expression of CYP52E3 and -N1 declines when rapeseed oil is added, but stays more or less the same when alkanes are added. This is already an indication for the alkane- inducible properties of the genes which is also observed for the alkane grown samples. Both CYP52E3 and -N1 are clearly induced during the initial growth on alkanes, and are therefore believed to take part in the metabolization of alkanes. The expression level of CYP52E3 stays more or less the same in the stationary phase, while CYP52N1 is threefold higher expressed here and becomes the dominant expression product.

# 2 Cloning and characterisation of the NADPH cytochrome P450 reductase gene (CPR) of *Candida bombicola*

## 2.1 Introduction

The previous section describes the importance of the CYP52 cytochrome P450 monooxygenase family for the glycolipid producing yeast *C. bombicola*. The cytochrome P450 monooxygenase system plays a crucial role in both alkane degradation and sophorolipid synthesis.

Most cytochrome P450 monooxygenases are not self-sufficient and thus receive the necessary electrons for oxygen cleavage and substrate hydroxylation from different redox partners. There are several P450 redox systems and members of the CYP52 cytochrome P450 monooxygenases, where the *C. bombicola* P450's fit in, belong the class II redox system. They form a small electron transfer chain together with the NADPH cytochrome P450 reductase (CPR). Both enzymes are N-terminally anchored to the endoplasmatic reticulum (ER) and derived structures and the body of the protein is located in the cytoplasmatic space (Edwards *et al.*, 1991). CPR is a flavoprotein containing the flavin cofactors FAD and FMN. It transfers the hydride ion of NADPH to the lower redox potential FAD. FAD then transfers single electrons to FMN, which in turn reduces the cytochrome P450 monooxygenase heme centre as required to activate molecular oxygen (Nerbert and Gonzalez, 1987). A schematic representation of this electron transfer process is given in Figure VI-8. Other final electron acceptor proteins are cytochrome b5 (Enoch and Strittmatter, 1979), heme oxygenase (Schacter *et al.*, 1972), squalene epoxidase (Ono *et al.*, 1977) or fatty acid elongase (Ilan *et al.*, 1981). CPR is also capable of reducing cytochrome c *in vitro*.

This section describes the cloning and sequence analysis of the *CPR* gene of *Candida bombicola* ATCC 22214. The identity of the gene was demonstrated by its functional expression in *E. coli* and was supported by its homology with other eukaryotic NAPDH cytochrome P450 reductases, which was particularly high for the cofactor binding regions.

CPR plays an important role in supporting cytochrome P450 monooxygenases and can even be the limiting factor for monooxygenase activity (Pompon *et al.*, 1996). Consequently, cloning and characterisation of this enzyme is an important step in the study of the

cytochrome P450 monooxygenase systems of *Candida bombicola* involved in the initiation of alkane degradation and the biosynthesis of sophorolipids.

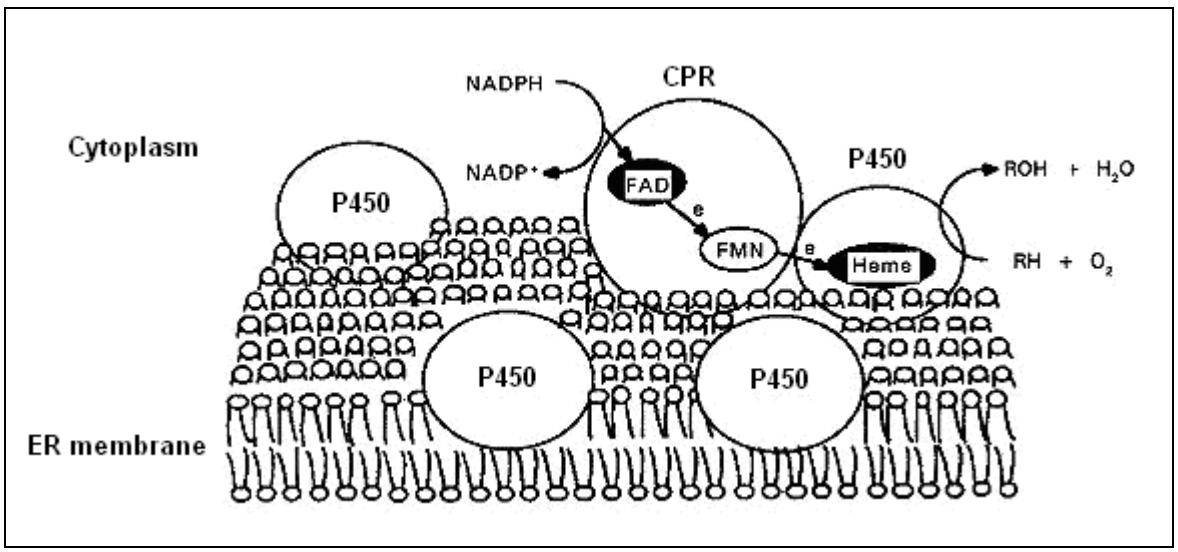


**Figure VI-8:** Presentation of the cytochrome P450 monooxygenase class II redox system between CPR and P450 (after Ohkawa *et al.*, 1998).

## 2.2 Materials and Methods

### 2.2.1 Strains, plasmids and culture conditions

*Candida bombicola* ATCC 22214 was used for the preparation of genomic DNA. All PCR products intended for sequence analysis were cloned into the pGEM-T® vector (Promega). *Escherichia coli* DH5α was used in all cloning experiments, the pTrcHis TOPO® and pTrcHis TOPO®/*lacZ* vector and *E. coli* TOP10 from Invitrogen were used for expression experiments.

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

### 2.2.2 DNA isolation and sequencing

Yeast genomic DNA and bacterial plasmid DNA was isolated as described in Section 1.2.2. of this chapter All DNA sequences were determined at the VIB Genetic Service Facility (Belgium).

### 2.2.3 Degenerate PCR

Part of the *CPR* gene of *C. bombicola* was amplified using degenerate primers DegFor and DegRev (Table VI-2). PCR amplification was carried out with an initial denaturation of 94 °C for 4 min, 40 cycles of 94 °C for 30 s, 50 °C for 1 min, 72 °C for 2 min with 5 s increasing time and a final 7 min elongation at 72 °C.

### 2.2.4 Genome Walking

Genome Walking was performed as described in Section 1.2.4 of this chapter. Four gene-specific primers (GSP) were designed and are listed in Table VI-2.

**Table VI-2:** Primers used for the isolation and cloning of the *Candida bombicola CPR* gene and reverse transcriptase real time PCR.. All primers were obtained from Sigma Genosys.

| Name | Feature | Sequence |
|---|---|---|
| DegFor | degenerate primer | 5'-ACH GGW ACB GCH GAR GAY TAY GC-3' |
| DegRev | degenerate primer | 5'-GAV GAM GAR ATV GAG TAG TAA CGW GG-3' |
| Up1 | primary upstream GSP | 5'-GGC GTT GTC AGT AGG CTC TCC ATC ACC-3' |
| UpN | nested upstream GSP | 5'-CCA TAT GAT GCC ATG AGG AAG ACA GCA AGA T-3' |
| Down1 | primary downstream GSP | 5'-CGA TAA GAC CTC GAC TGT GCG TAT ACC TTC-3' |
| DownN | nested downstream GSP | 5'-TTG TTC GCA AGC CAT GTG GCC GCG AAG A-3' |
| TotFor | primer for isolation of coding sequence | 5'-GCC GAT ATT AAT TTT ATC GCT TCG GTC GTT-3' |
| TotRev | primer for isolation of coding sequence | 5'-GCT ACC AAA CGT CCT CTT GGT ACT-3' |
| CPRqFor | real-time PCR primer | 5'-CCC GAC GTC CAG AGC AGT A-3' |
| CPRqRev | real-time PCR primer | 5'-CAT CCT CGT CAG CTT TTG AAT TG-3' |

### 2.2.5 Cloning and expression

The isolation of the complete *CPR* gene with its upstream and downstream flanking regions was carried out with the two specific primers TotFor and TotRev (Table VI-2). The fragment was amplified from genomic DNA with the High Fidelity PCR Master kit (Roche Diagnostics) using the following temperature program: initial denaturation at 94 °C for 2 min, 10 cycles of 94 °C for 10 s, 50 °C for 30 s, 72 °C for 2 min, 15 cycles of 94 °C for 15 s, 50 °C for 30 s, 72 °C for 2 min with 5 s increasing time and a final 7 min elongation at 72 °C. The obtained PCR-fragment was cloned in pTrcHis TOPO® and a vector with the fragment in the proper orientation for expression was selected after restriction analysis. The resulting vector was called pTopoCPR and was used for CPR expression in *E. coli* TOP10 cells following the guidelines of the pTrcHIS TOPO®TA Expression Kit user's manual. The pTrcHis TOPO®/*lacZ* control vector, constructed for the expression of a fragment of the *lacZ* gene, was used as negative control in the CPR activity assay.

### 2.2.6 CPR activity assay

Expression in *E. coli* TOP10 was induced with 1mM IPTG and from this stage cells were harvested every hour over a time period of 10 hours. Cell free extracts were obtained by enzymatic lysis with EasyLyse™ Bacterial Protein Extraction Solution (Epicentre). Protein

concentrations were determined as described by Bradford *et al.* (1976). The NADPH cytochrome c reductase activity was measured spectrophotometrically at room temperature with a Uvikom 922 spectrophotometer (BRS). The 1 mL reaction mixture contained an appropriate volume of cell free extract containing 0.5 mg of protein, 1.3 mM $NADP^+$, 3.3 mM glucose-6-phosphate, 0.4 U/mL glucose-6-phosphate dehydrogenase, 3.3 mM $MgCl_2$ and 0.95 mg cytochrome c in 250 mM potassium phosphate buffer. The extinction coefficient for reduced cytochrome c at 550 nm under the described conditions is 21 $mM^{-1}$ $cm^{-1}$.

### 2.2.7 Transformation

*E. coli* DH5α cells were transformed as described by Sambrook and Russell (2001). *E. coli* TOP10 was transformed following the guidelines of the pTrcHIS TOPO®TA Expression Kit user's manual.

### 2.2.8 Sequence analysis

Sequence analysis, similarity searches, construction of the phylogenetic tree and creation of the hydropathy plot were conducted as described in Section 1.2.5 of this chapter.

### 2.2.9 Reverse transcriptase real-time PCR

Reverse transcriptase real-time PCR for the *CPR* gene was performed as described in Section 1.2.6. of this chapter. The real-time PCR primers are listed in Table VI-2.

### 2.2.10 Nucleotide sequence accession number

The nucleotide sequence described in this paper has been deposited at the GenBank nucleotide database under the accession number EF050789.

## 2.3 Results and Discussion

### 2.3.1 Isolation of the *C. bombicola CPR* gene

The *CPR* genes of several yeast species of the subphylum *Saccharomycotina* were compared in a multiple alignment, revealing several conserved regions corresponding to conserved amino acid sequences previously described by Van den Brink *et al.* (1995). Degenerate oligonucleotides DegFor and DegRev were designed based on the highly conserved FMN-1 and FAD-2 binding regions, respectively (Figure VI-8, amino acids 69 - 92 and 447 - 464 for *Candida bombicola*).

Degenerate PCR conducted with the primers DegFor and DegRev resulted in the appearance of several amplified products, probably due to the ability of the primers to anneal with various genome fragments encoding proteins with similar co-factor binding properties. However, based on the *CPR* sequences of the *Saccharomycotina*, a fragment ranging from 1133 till 1199 bp was expected for the *CPR* gene of *Candida bombicola*. An 1146 bp fragment was therefore isolated and subsequent sequence analysis revealed a 49 % amino acid sequence identity with the NADPH cytochrome P450 reductase of *Yarrowia lipolytica.*

The regions adjacent to the obtained fragment were cloned using the BD GenomeWalker™ Universal Kit. As a result, a DNA sequence of 4362 bp was obtained and showed to contain the total coding sequence of 2064 bp and 5’ and 3’ regions of 461 and 1837 bp, respectively.

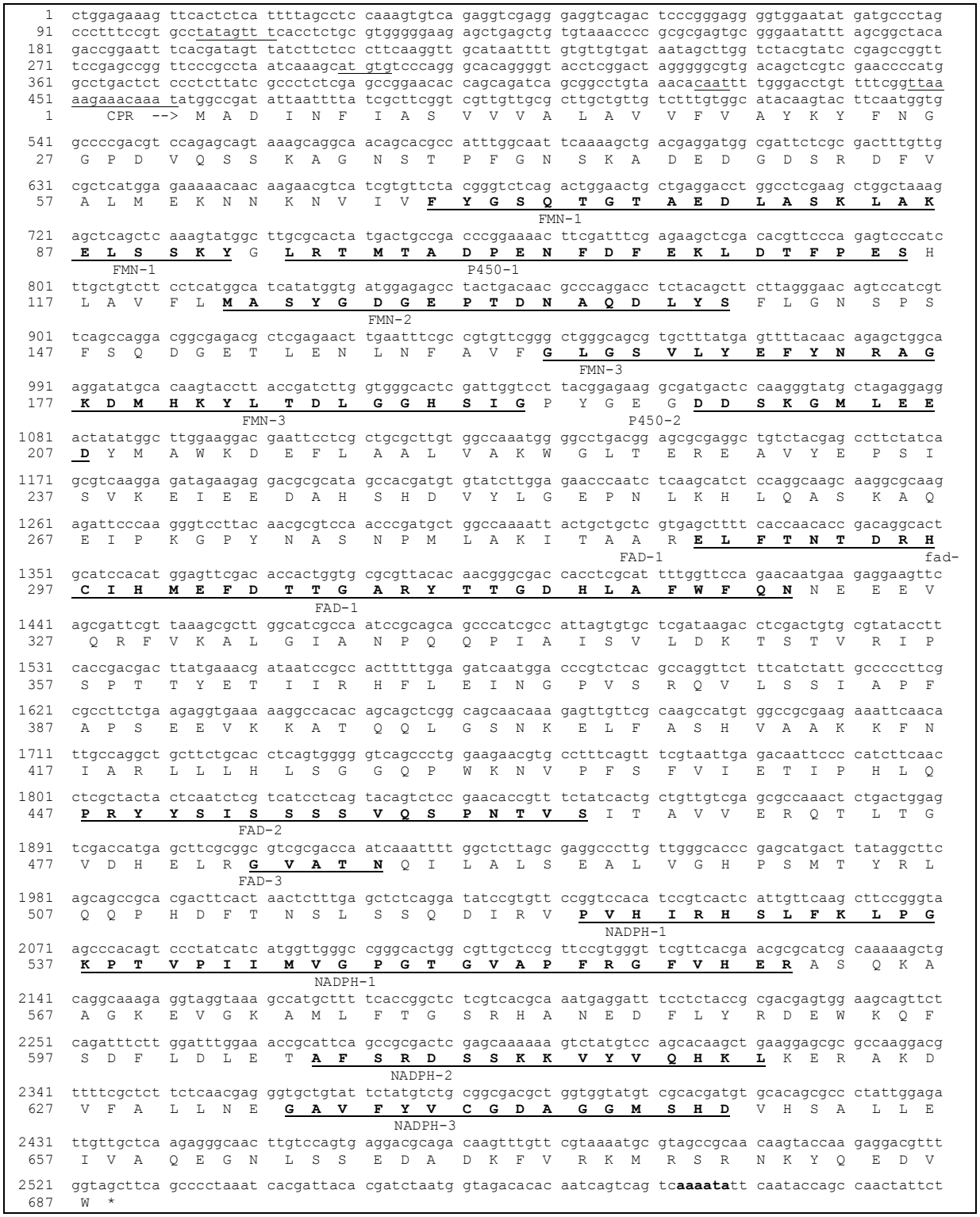


**Figure VI-8:** Nucleotide and deduced amino acid sequence of the *CPR* gene of *C. bombicola*. Possible promoter elements are underlined and sequences presumed to be involved in polyadenylation are indicated in bold. The conserved amino acid sequences of the CPR as described by Van den Brink *et al.* (1995) are marked in bold and underlined cases.

### 2.3.2 Characterisation of the *CPR* sequence

There was no conventional TATA-box found in the 5' region. Yet, this is not unusual for yeast genes since only 20 % of the promoters are supposed to be regulated by TATA-box elements (Basehoar *et al.*, 2004). At position -358 however, the TATA-like structure TATAGTTT was observed. He and Chen (2005) reported a similar sequence 332 bp upstream the *CPR-a* gene of *Candida tropicalis*. Furthermore, a CAAT-box was found at position -37 and the sequence just preceding the translation start site is extremely A-rich, revealing a possible role in transcription initiation. Kozak *et al.* (1991) described two positions most critical for the function of the ATG initiator codon: a purine (usually A) in position -3 and a G in position +4. Both nucleotides are present in the *C. bombicola CPR* gene.

The complement of the more specific pentanucleotide sequence CACAT is present at position -163. This sequence or its complement ATGTG also occurs in the 5' untranslated regions of the *CPR* genes of *Candida maltosa* Ohkuma *et al.*, 1995) and *Cunninghamella elegans* (Yadav and Loper, 2000), and in the promoter region of cytochrome P450 monooxygenase genes of yeasts (Yadav and Loper, 1999), which was also observed for CYP52E3 and CYP52N1 from *C. bombicola* (Section 1.3.2). The consensus sequence of the polyadenylation signal found in most eukaryotic genes is AATAAA, although for yeasts several variations are possible (Guo and Sherman, 1996). In the 3' region of the *CPR* gene of *C. bombicola*, a putative polyadenylation signal sequence (AAAATA) was observed 58 bp downstream of the TAG stop codon.

There were no indications for the presence of an intron. The *CPR* open reading frame is translated into a protein of 687 amino acids with a calculated size of 76.22 kDa and an estimated pI value of 5.6.

The Kyte-Doolittle hydropathy plot (Kyte and Doolittle, 1982) of the CPR amino acid sequence suggests the presence of an N-terminal transmembrane region (Figure VI-9). CPR and its partner enzyme cytochrome P450 monooxygenase are indeed both considered to be N-terminally anchored to the endoplasmatic reticulum with the remainder of the enzyme facing the cytoplasmatic side of the membrane (Sanglard *et al.* 1993). The hydrophobic fragments are in addition thought to be uncleavable signal sequences targeting the CPR towards the endoplasmatic reticulum (Black, 1992). Furthermore, Kärgel *et al.* (1996) stated that the CPR

of *Candida maltosa* triggers a strong proliferation of the membrane system and that the N-terminal signal-anchor sequence is involved in this process.

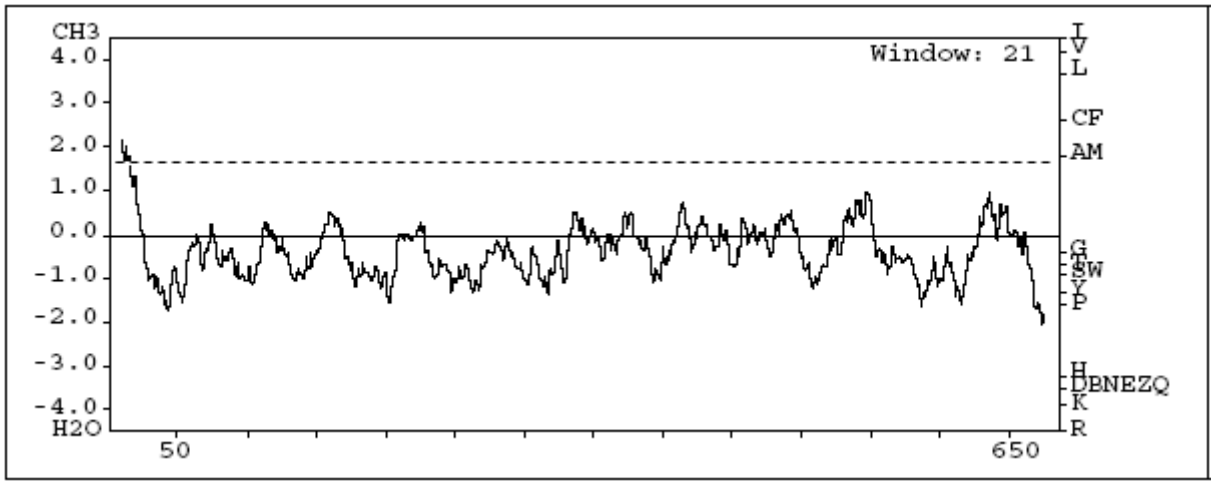


**Figure VI-9:** Kyte-Doolittle hydropathy plot of CPR of *C. bombicola*. When window size of 21 amino acids is used, regions with a score higher than 1.6 are believed to be transmembrane. This is clearly the case for the N-terminal region.

The protein shows a 40 to 49 % amino acid sequence identity with all known and putative NADPH cytochrome P450 reductases of the *Ascomycota*. Furthermore, the reductase of *C. bombicola* turns out to be more homologous to those of vertebrates (37 - 39 %) and insects (36 - 37 %) than to the NADPH cytochrome P450 reductases of higher plants (31 - 34 %). The same trend was also observed by Yadav and Looper (2000) for the filamentous fungi *Cunninghamella elegans* and *Cunninghamella echinulata* and is confirmed when looking at the phylogenetic tree for the eukaryotic CPR's (Figure VI-10). The tree was constructed based on the protein maximum likelihood principle and comprises all the 65 NADPH cytochrome P450 reductases found in the GenBank database. For all species, only one protein was incorporated and putative or hypothetical proteins were excluded. The sequence of *Plasmodium falciparum* was used as outgroup and the tree was rooted against it. The CPR of *Candida bombicola* is retrieved on a separate branch in the yeast group. It does not cluster together with the other *Candida* CPR's. This result agrees with a previous phylogenetic analysis on *Candida* species (Suzuki *et al.* 1999), in which *C. tropicalis, C. albicans* and *C. maltosa* clustered together and were found in the same group as *S. cerevisiae*, whereas

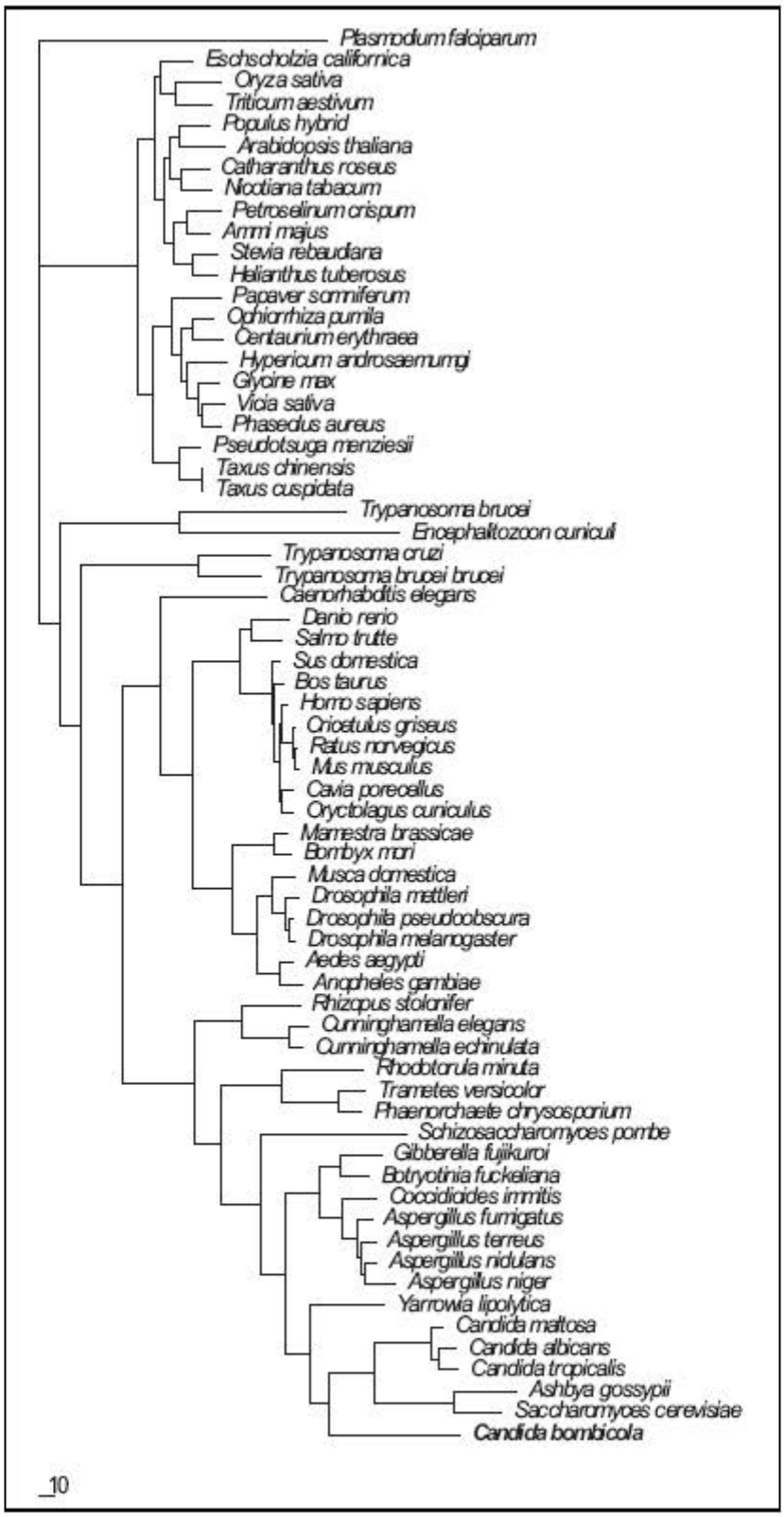


**Figure VI-10:** Phylogenetic tree for eukaryotic NADPH cytochrome P450 reductase rooted against the *Plasmodium falciparum* CPR. The Ln Likelihood of the tree is -49566 and all knobs and branches were statistically significant ($p < 0.01$). The marker bar below denotes the integer branch length.

*C. bombicola* belonged to a different group.

Despite overall amino acid sequence similarities of less than 50 % between the CPR's of *C. bombicola* and those of other organisms, the domains believed to be involved in the binding of NADPH, FAD and FMN exhibit a very high similarity (Vandenbrink *et al.*, 1995). Those highly conserved regions are indicated in Figure VI-8.

### 2.3.3 Putative cytochrome c oxidase assembly factor gene flanking *CPR* (partial)

In the 1837 bp region downstream the *CPR* coding sequence, an open reading frame of 1456 bp preceded by a possible TATA-box was found. The hypothetical gene is almost complete and encodes a putative partial protein of 485 amino acids that shows homology with the cytochrome c oxidase assembly proteins (COX) of several species (54 % amino acid identity match with COX15 of *Saccharomyces cerevisiae*). Those proteins take part in synthesis of heme A for cytochrome c oxidases. They display oxidoreductase activity, acting on NADH or NADPH, and are required for the hydroxylation of heme O to form heme A, which is an essential prosthetic group for cytochrome c oxidase.

Thus like CPR, the putative protein should display oxidoreductase activity and its putative function is also related to heme containing proteins. It is not clear if the co-localisation of both genes is a coincidence or an indication for the clustering of related genes. However, in the genomes of other yeasts and higher eukaryotes, such co-localisation is not observed.

### 2.3.4 Heterologous expression of CPR and activity testing

In order to verify functionality of the proposed *CPR* gene product, the pTopoCPR expression vector harbouring the coding region of the *CPR* sequence was constructed as described in the Materials and Methods section. pTopoCPR was used to express the protein in *E. coli* and heterologous expression was verified by SDS-PAGE (data not shown). The activity was verified by cytochrome c reduction in the cell free extract (Figure VI-11). Time course measurements indeed revealed a clear reductase activity in the extracts of IPTG induced cells and the obtained absorption spectra corresponded to that for cytochrome c in the reduced state. These results confirm the activity of the protein product of the isolated *CPR* gene of *Candida bombicola*.

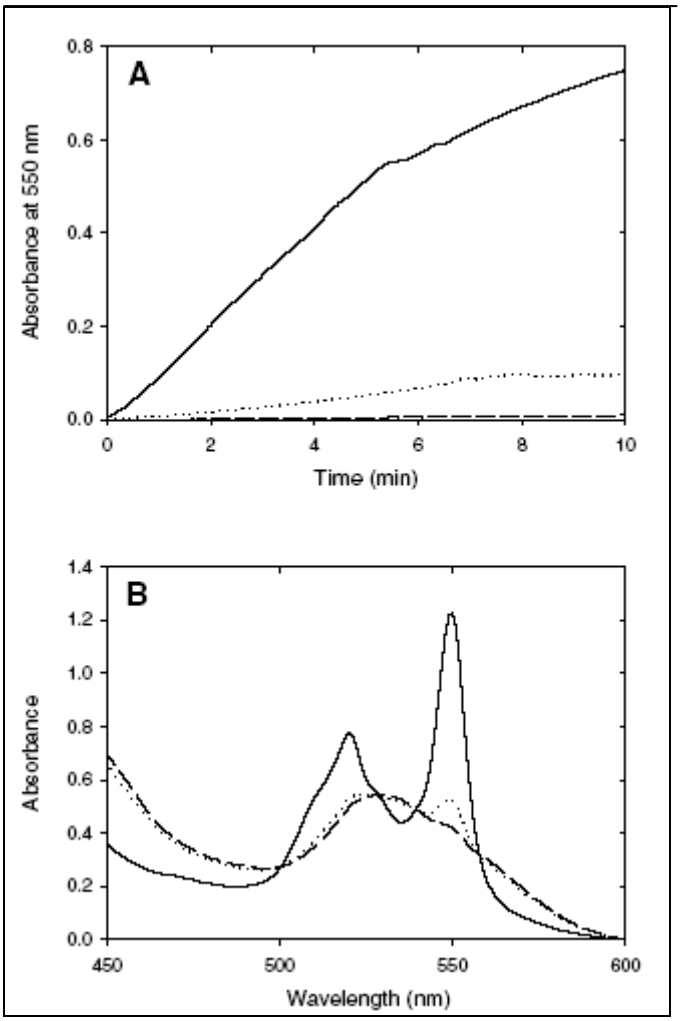


**Figure VI-11:** Activity measurement of the CPR from *Candida bombicola* by cytochrome c reduction to verify the functionality of the isolated gene. — cell free extracts of *E. coli* expressing CPR 6 hours after IPTG induction; ….. blank: cell free extracts of *E. coli* expressing a fragment of the *lacZ* gene 6 hours after IPTG induction; ---- blank: reaction mixture without cell free extract. (A) Time course of cytochrome c reduction monitored at 550 nm. (B) Absorption spectra obtained after saturation of the reaction.

### 2.3.5 Expression profile of the isolated *CPR* from *C. bombicola*

The expression profile of the *CPR* gene was determined by a two-step reverse transcriptase real-time PCR with Sybr® Green detection as described in the Materials and Methods section 2.2.9. The same culture conditions as investigated for the *CYP52* genes were evaluated for the *CPR* gene (Figure VI-12).

Although some variations among the samples could be observed, no big differences as found for the CYP52 genes were observed (for the CYP52-M1, -E3 and –N1 genes, the strongest

signal was respectively 584, 57 and 28 fold higher as compared to the lowest one). When looking at sophorolipid related samples, there is hardly any increase of expression level upon the synthesis of sophorolipids. Yet, for the stationary sample of sophorolipid synthesis on alkanes a slightly higher value is obtained. The *CPR* expression values of the alkane grown samples are 1.9 to 4.6 times higher than the ones obtained in the sophorolipid related samples. It therefore can be concluded that *CPR* is certainly not up-regulated upon sophorolipid production. Since CPR's support several other enzymes, this is no unexpected result.

However, it stays unclear whether the number of available CPR enzymes is sufficient to support the P450 enzymes. Because of differences in mRNA and protein stability, it is quite tricky to compare the expression level of different genes and based thereon draw conclusions about their relative enzyme activity. Experiments at the protein level should be conducted to verify this.

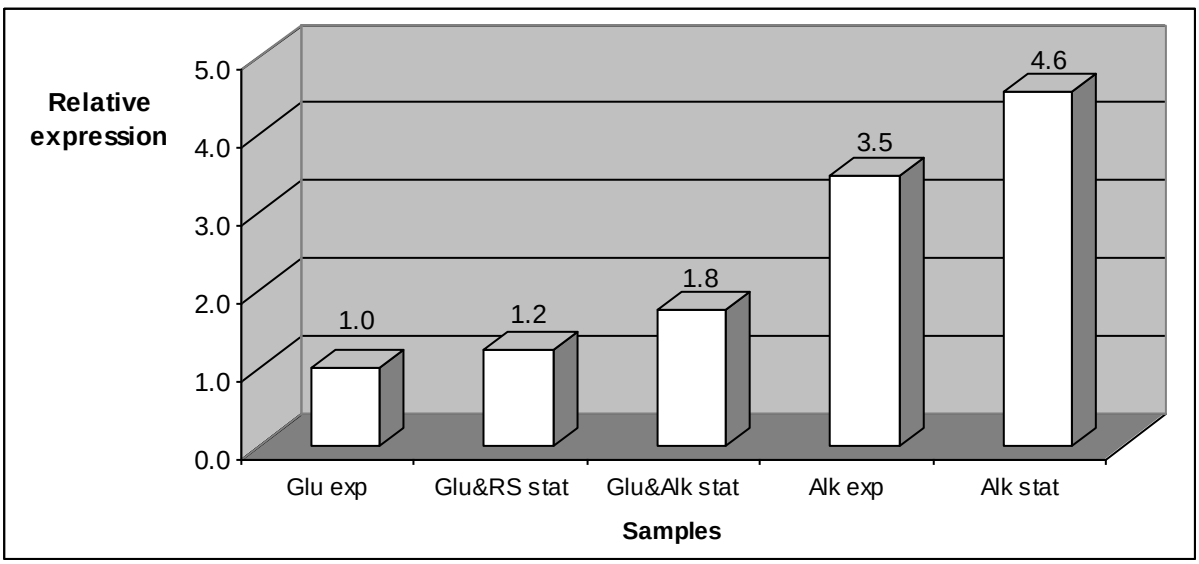


**Figure VI-12:** Relative expression of CPR in different experimental conditions: exponential and stationary phase sample for sophorolipid production on glucose and rapeseed oil (**Glu exp** and **Glu&RS stat** respectively), stationary phase sample for sophorolipid production on glucose and hexadecane (**Glu&Alk stat**) and exponential and stationary phase sample for growth on alkanes (**Alk exp** and **Alk stat** respectively).

# 3 Conclusion

Since *C. bombicola* is a yeast capable of growing on alkanes it was postulated that this species possessed cytochrome P450 monooxygenases belonging to the CYP52 family, which were probably also responsible for terminal or subterminal hydroxylation of fatty acids for sophorolipid synthesis. Three such genes were completely sequenced: one gene product was 92 % identical to CY52E2 from *C. apicola* - a species closely related to *C. bombicola* which also produces the same type of sophorolipids – and was therefore classified into this subfamily and named CYP52E3 (GenBank accession number EU552420). The other proteins did not show any significant match with an existing subfamily and are considered to be the first members of two new groups CYP52M and CYP52N (GenBank accession numbers EU552419 and EU552421). All relevant conserved regions and amino acids were found in the three new proteins, and CYP52E3 and -N1 even possessed the specific pentanucleotide sequence CACAT/ATGTG. Furthermore, the enzymes possessed an N-terminal transmembrane region for N-terminal anchoring in the endoplasmatic reticulum, as observed in P450's and CPR's.

In order to elucidate the possible function of the isolated CYP52's an expression experiment was conducted. CYP52M1 takes no part in alkane assimilation and was clearly induced upon sophorolipid synthesis, indicating a possible role in sophorolipid synthesis. CYP52E3 and CYP52N1 on the other hand are probably of no importance in sophorolipid production, but show a clear up-regulation when the yeast cells are grown on alkanes as the sole carbon source. CYP52E3 is the major expression product in the exponential growth phase, but stays more or less stable in the stationary phase, whereas the expression of CYP52N1 increases and becomes the dominant one there.

Those expression profiles already give an indication on the function of the isolated *CYP52* genes. However, it stays unclear whether they are just induced in the investigated situation, or if they are actually taking part in the biochemical process of sophorolipid synthesis or alkane degradation. For this purpose it is recommended to conduct activity experiments with the enzymes: upon expression of the *C. bombicola* genes in *S. cerevisiae*, microsomal isolates can be used to test the activity towards fatty acids and alkanes with a different chain length. Final confirmation of the physiological function of the CYP52's can be obtained by the creation of

*C. bombicola* knock-out strains. As it is possible that several (iso)enzymes are responsible for one and the same biochemical reaction, it will perhaps be necessary to perform multiple gene deletions in order to observe a clear difference in the yeast phenotype.

One issue one should keep in mind is the presence of at least five other very similar CYP52 genes in the *C. bombicola* genome. The isolation by means of a primer based strategy failed. Ideally, all CYP52 genes can be found upon sequencing of the whole *C. bombicola* genome. Another solution is the creation of a genomic DNA library in *E. coli*. Vectors containing *P450* fragments can be identified by colony hybridisation with a suitable probe designed on the conserved regions. The size of the inserts combined with the results of other positive colonies will make it possible to pick up complete genes. By altering the stringency of hybridisation and the sequence of the probe itself, multiple *CYP52* genes should be picked up.

Most eukaryotic P450 enzymes are not self-sufficient. An electron donor enzyme is required and in the case of the CYP52 family, this is done by the NADPH cytochrome P450 oxidoreductase. The CPR from *C. bombicola* was cloned (Genebank accession number EF050789) and its functionality was proven by heterologous expression in *E. coli*. Eventhough multiple P450's can be present in one organism, only one CPR supports all those enzymes.

The CYP52 cytochrome P450 monooxygenases play a key role in the synthesis of sophorolipids and, by their specific preference for C16 and C18 fatty acids, they determine the molecular structure of the native sophorolipids. Insight into the CYP52 family of *C. bombicola* can teach us to understand this mechanism and maybe alter the fatty acid tail of the sophorolipid molecule by homologous (over)expression of suitable *C. bombicola* genes which are now down-regulated or by heterologous expression of other CYP52 genes with the desired activity. Upon high expression of P450's, CPR enzymes can be the rate-limiting factor. One study of an *in vitro* system suggests a molar ratio of 3:1 for CPR to CYP52 (Scheller *et al.*, 1998). However, it remains unclear if these ratio's can be translated to the *in vivo* system. In this context, the obtained *CPR* sequence of *C. bombicola* can be useful for the improvement of sophorolipid production by overexpressing this gene.

# Chapter VII: Conclusions and perspectives

In terms of production volume, surfactants belong to the most important classes of industrial chemicals with a current total world production of about 10 million ton per year. The large majority of them are petroleum-based and are produced by chemical means. These compounds are often toxic to the environment, and their use may lead to significant ecological problems, particularly in washing applications as these surfactants inevitably end up in the environment after use. In this context, the ecotoxicity, bioaccumulation, and biodegradability of surfactants have become issues of increasing concern.

Glycolipid surfactants are composed of a carbohydrate head and a lipid tail. Their market share has significantly increased during the last 10 years as they offer a vastly improved environmental compatibility as compared to traditional surfactants, combined with excellent functional properties. Whereas the first generation of glycolipids was produced from renewable resources through chemical means *e.g.* alkylpolyglucosides (APG), the second generation of glycolipids is obtained from renewable resources through biotechnological means; indeed, glycolipids produced by fermentation are now entering the market, particularly rhamnolipids and sophorolipids. These latter molecules consist of the dimeric sugar sophorose linked to a long-chain hydroxy fatty acid. They possess good surface active properties and show excellent skin compatibility, a property that is very important for cosmetic and personal care applications. Furthermore, they can be used in various other sectors due either to their emulsifying, antimicrobial, or to other beneficial properties.

In contrast to rhamnolipids, which are synthesized by *Pseudomonas aeruginosa* strains, the sophorolipid producing yeast strains are nonpathogenic. Very high production yields can be achieved (over 400 g/L), and this based on renewable resources or even waste streams. This fact makes sophorolipids particularly attractive for further commercial production and widespread use.

Sophorolipids are a diverse group of molecules differing in their degree of acetylation, presence or absence of lactonization, position of the fatty acid hydroxyl group, fatty acid chain length and saturation. Despite their overall heterogeneity, sophorolipids possess little variation in the length of their lipid tail. The range is limited to C16-C18 saturated or unsaturated fatty acids and this is governed by the specificity of one or more cytochrome P450 monooxygenases (P450). These enzymes hydroxylate the fatty acids at the terminal (ω)

or subterminal (ω-1) position and transform them in this way into suitable substrates for glucosyltransferase I, which will form an ester bond between the introduced hydroxyl group and the first glucose molecule. Stearic acid (C18:0) turns out to be incorporated into sophorolipids with the highest efficiency, but also vegetable oils - which are rich in C16-18 fatty acids - are very suitable as hydrophobic carbon source for sophorolipid synthesis.

However, the incorporation of fatty acids differing from the conventional C16-C18 range could broaden up the application potential of sophorolipids. The incorporation of medium-chain fatty acids should render the molecules more hydrophilic and consequently improve their water solubility and foam forming capacities, making them even better detergents. In order to achieve this, the substrate specificity of the cytochrome P450 monooxygenase must be circumvented.

This can be achieved in two ways. The first method comprises the addition of already hydroxylated substrates, in this way skipping the controlling effect of the P450 enzyme. The other method comprises the incorporation of unconventional stearic acid-like substrates which can undergo post-fermentative modification to give rise to shorter-chained sophorolipids. Substrates with internal ester bonds are particularly useful for this purpose as the bond can be cleaved by alkaline hydrolysis. Stearic acid-like substrates with a carboxylic end are preferred over substrates with alkyl ends since the carboxyl group only allows incorporation into one direction, in this way controlling the chemical nature of the hydrolysed sophorolipids. Furthermore, better yields are obtained with carboxylated substrates (Table VII-1). When performing fermentations under the same conditions as applied for the carboxylated esters with well performing substrates such as rapeseed oil, similar results are obtained: yields of about 0.70 mol/mol and sophorolipid concentrations of 80 to 100 g/L. Therefore, it can be expected that upon further optimization of the fermentation parameters, especially the feeding rate of the esters, higher concentrations of sophorolipids can be reached. Meanwhile, the use of stearic-acid like substrates is patented by the company Ecover Belgium NV (Develter and Fleurackers, 2007).

It is thought that the hydroxylated medium-chain substrates and the esters with alkyl ends are to a certain extent metabolized in the β-oxidation pathway, resulting for such substrates in lower yields. This problem is further discussed and tackled in Chapter V.

**Table VII-1:** Maximal yields for the unconventional substrates described in Chapter II.

| Substrate | Feature | Production (g/L) | Yield (mol/mol) |
|---|---|---|---|
| 12-Hydroxydodecanoic acid* | hydroxylated | 22.0 | 0.29 |
| 1,12-Dodecanediol | hydroxylated | 84.4 | 0.47 |
| Dodecyl glutarate | carboxylated ester | 79.7 | 0.70 |
| Dodecyl and myristyl malonate | carboxylated ester | 55.4$^{+}$ | 0.70 |
| Pentenyl dodecanoate | alkyl ester | 82.6 | 0.25 |
| Dodecyl pentanoate | alkyl ester | 45.7 | 0.20 |

* In contrast to the other results, obtained in shake flask cultures
$^{+}$ Fermentation process was preliminary stopped

The experiments described in Chapter II demonstrate that *Candida bombicola* can produce medium-chain sophorolipids by using unconventional hydrophobic carbon sources. However, these substrates are quite expensive. Substrates with an internal ester bond are not commercially available; synthesis costs about 1.5 €/kg substrate. The higher substrate cost and the expenses for the required post-fermentative modification makes the use of those medium-chain sophorolipids in the cleaning and detergent industry less economical feasible. However, if those sophorolipids could be used for speciality applications in the cosmetic or pharmacy sectors, the higher production price becomes less important. The medium-chain hydroxy fatty acids derived from those sophorolipids, for example, can find applications in the perfume industry. The medium-chain sophorolipids should be further screened for beneficial biological properties such as antibacterial, antifungal, anticancer and anti-HIV activities. Since native sophorolipids already show such behaviour, it is not unlikely that medium-chain sophorolipids possess equal or better activity.

As the production of medium-chain sophorolipids starting from unconventional substrates is not (yet) economically feasable, it would be beneficial if *C. bombicola* could in some way be forced to synthesise sophorolipids starting from the conventional and cheaper medium-chain fatty acids, oils or even waste streams. This would also omit the need for post-fermentative modification. One possible way for achieving this, is the (over)expression of cytochrome P450 monooxygenases with a specificity towards medium-chain fatty acids in the *C. bombicola* genome. To reach this goal one needs molecular tools for this unconventional

yeast species: a transformation and selection system, an expression system, knock-out system, gene sequences *etc*. As yet, none of this is available or reported for *C. bombicola* despite the large efforts spent to optimize the fermentation process to improve yields. Furthermore, genetic engineering of *C. bombicola* is a useful tool for the study of the sophorolipid synthesis pathway and opens up perspectives for improved or altered production. Besides the overexpression of heterologous P450 enzymes, certain pathways such as β-oxidation and *de novo* fatty acid synthesis can be blocked in order to achieve improved synthesis of medium-chain sophorolipids (this is discussed in Chapter V).

The first step in the genetic engineering of *C. bombicola* is the development of a suitable transformation and selection method (Chapter III). *C. bombicola* can be transformed by the slightly modified Lithium-Acetate method used for *Saccharomyces cerevisiae*. The sophorolipid producing yeast turned out to be quite resistant to several antibiotics. In addition, selection systems based on dominant drug resistance genes are expensive and not very suitable for industrial fermentations and/or applications. Therefore, it was decided to develop a transformation system based on the auxotrophic markers orotidine-5'-phosphate decarboxylase (URA3). In order to develop such a system, the *URA3* gene of *Candida bombicola* was cloned (GenBank accession number DQ916828).
The next step is the isolation of a stable ura3-auxotrophic mutant which can be transformed back to prototrophy with the species-own *URA3* gene. Such mutants were found by selection on media containing 5-fluoroorotic acid (5-FOA). Several mutants were tested and evaluated on true phenotype, viability and revertation frequency. The mutant G9, missing 7 bp in the coding sequence of the *URA3* gene, was used to perform the transformation with the wild type *URA3* gene in order to restore the ura3 prototrophy. Successful transformation was confirmed by a PCR-based method discriminating between the wild type and mutated *URA3* gene. In 19 % of the transformants, double cross-over resulting in gene replacement was observed.

One of the logical next steps in the genetic engineering of *C. bombicola* is the construction of an expression system. It was demonstrated in Chapter III that the *TEF1α* promoter of *Eremothecium gossypii* and the *TK* promoter of the *Herpes simplex* virus also show activity in

*C. bombicola*. On the other hand, it is believed that homologous promoters give rise to better and higher expression levels.

Glyceraldehyde-3-phosphate dehydrogenase is one of the key enzymes taking part in the glycolysis, an essential pathway present in every living cell. The constitutively and highly active promoter of native *GAPD* genes has been successfully applied for the expression of heterologous genes in several yeasts and filamentous fungi. The *C. bombicola GAPD* gene and 1613 bp of the upstream region was cloned (GenBank accession number EU315245).

It was demonstrated that even short *GAPD* promoter fragments of 190 bp could still be used for the expression of heterologous genes in *C. bombicola*. However, promoter fragments bigger than 488 bp tend to recombine with the genomic *GAPD* promoter, in this way knocking out the *GAPD* gene which is essential for cell viability. This event can be avoided in future expression experiments by flanking the gene with a functional *URA3* gene and transforming the ura3-negative G9 strain described in Chapter III. This would lead to recombination at the dysfunctional *URA3* gene instead of the *GAPD* promoter.

The efficacy of a short *GAPD* promoter can be a convenient characteristic for the construction of compact expression cassettes or vectors for *C. bombicola*. However, the exact expression level of the different *GAPD* promoter fragments should first be investigated by *e.g.* expressing β-galactosidase and quantifying its activity.

As seen forward from Table VII-1 and the results described in Chapter II, certain substrates used for the production of medium-sophorolipids are to a certain degree metabolized in the β-oxidation pathway. In order to redirect these substrates towards sophorolipid synthesis, the competing β-oxidation pathway was blocked at the genome level of the yeast by targetting the multifunctional enzyme 2 gene. In contrast to other enzymes contributing to the metabolization of fatty acids, *MFE2* is believed to occur as a single copy in the genome making the construction of a *C. bombicola* strain with a blocked β-oxidation quite feasible.

The total gene sequence of the *C. bombicola MFE-2* was cloned (GenBank accession number EU315245) and the obtained nucleotide sequence was used to construct a knock-out cassette. Homologous recombination turned out to be most efficient with large fragments of 1000 bp of the target gene flanking the selection marker.

Several deletion strains were tested in a fermentation run with 1-dodecanol and all of them showed a 1.7 to 2.9 times higher production of sophorolipids, indicating that in strains with a blocked β-oxidation pathway, the substrate is redirected towards sophorolipid synthesis.
This experiment illustrates that the mutants can be used to improve the production and yield on medium-chain substrates, in this way lowering the production costs. In future experiments, other substrates which are subjected to metabolization by the yeast will be evaluated. The best mutants will be selected and with these, larger scale fermentations will be performed in order to optimize the culture conditions and feeding of the substrate.
Somewhat unexpected results were obtained when running fermentations with the mutants on rapeseed oil: instead of equal or slightly better sophorolipid yields, inferior results were obtained. It is possible that the cells are in some way inhibited by rapeseed fatty acid derived products, which accumulate due to the blocked β-oxidation. This learns that for production of sophorolipid on conventional substrates, the wild type strain should be preferred above the mutant strains.

Another pathway interfering with the formation of medium-chain sophorolipids is the synthesis of native *de novo* sophorolipids starting from *de novo* formed fatty acids. In this way, the obtained sophorolipids are in fact always a mixture of medium- and long-chain ones.
The *de novo* synthesis of fatty acids by the fatty acid synthase complex is essential for cell viability and consequently can not be blocked by knocking out one of the involved genes. For this reason, it was decided to work with cerulenin, an inhibitor of the fatty acid synthase complex. Cerulenin could indeed block synthesis of *de novo* sophorolipids in glucose grown cultures. Also for fermentations run on 1,12-dodecanediol, all detected sophorolipids originated from the hydrophobic substrate. These results clearly indicate that when blocking the *de novo* fatty acid synthesis, a more uniform mixture of medium-chain sophorolipids can be obtained.
However, the hydrophobicity, costs and potential toxicity for humans, renders cerulenin not suitable for industrial applications. Blocking the *de novo* fatty acid synthesis on the genome level is no suitable alternative, as fatty acid synthesis is essential for cell growth. Another way of eliminating the *de novo* sophorolipid synthesis is knocking out the cytochrome P450

monooxygenase(s) which hydroxylate(s) the fatty acids. However, those mutants must be supplied with already hydroxylated substrates if sophorolipid production is wanted.
Blocking the *de novo* sophorolipid synthesis results in a sophorolipid mixture with a reduced complexity. This simplifies further evaluation of the physico-chemical characteristics of medium-chain sophorolipids and makes additional purification steps unnecessary or less complex. Furthermore, a disabled *de novo* synthesis will result in the redirection of precursors and co-factors towards the medium-chain sophorolipids, giving rise to improved yields.

As already mentioned before, the length of the fatty acid tail of native sophorolipid molecules is governed by the selectivity of cytochrome P450 monooxygenases. The enzymes turn out to have a strict specificity towards fatty acids with a chain length of 16 to 18 carbon atoms. In order to better understand the controlling role of the P450's and to be able to change their specificity, the P450 system of *C. bombicola* was studied.
It was thought that the *C. bombicola* genome harboured several P450 genes belonging to the CYP52 family and indeed, three of such genes were isolated: *CYP52M1*, *CYP52E3* and *CYP52N1* (Genbank accession numbers EU552419, EU552420 and EU552421). P450 enzymes belonging to the CYP52 family are all derived from alkane assimilating yeasts, just like *C. bombicola*, and hydroxylate fatty acids and/or alkanes. Consequently, it remains to be solved whether the CYP52 enzymes of *C. bombicola* intervene in sophorolipid synthesis, alkane assimilation or are just not expressed at all. In order to get some information on this issue, an expression analysis was performed. CYP52M1 was clearly induced upon sophorolipid synthesis and takes no part in alkane assimilation. CYP52E3 and CYP52N1 on the other hand are probably of no importance in sophorolipid production, but show a clear up-regulation when the yeast cells are grown on alkanes as the sole carbon source.
However, it remains unclear if the genes are just induced under the conditions studied, or if they are actually taking part in the biochemical processes of sophorolipid synthesis or alkane degradation. In order to investigate this further, it is recommended to test the enzyme activities towards fatty acids and alkanes with a different chain length. Final confirmation about the physiological role of the CYP52's can be obtained by the construction of *C. bombicola* knock-out strains. As it is possible that several (iso)enzymes are responsible for

one and the same biochemical reaction, it will perhaps be necessary to perform multiple gene deletions in order to observe a clear difference in the yeast phenotype.
One issue one should also keep in mind is the presence of at least five other very similar *CYP52* genes in the *C. bombicola* genome with an unknown function. In order to obtain a total and clear view on the CYP52 family of *C. bombicola*, it would be recommendable to know the full gene sequences of all these members. Ideally, this can be achieved upon sequencing of the whole *C. bombicola* genome. Another solution is the use of a genomic DNA library in *E. coli*.

Most eukaryotic P450 enzymes are not self-sufficient. In the case of members of the CYP52 family, electrons must be supplied by the NADPH cytochrome P450 oxidoreductase (CPR) and upon high expression of P450's, CPR enzymes can be the rate-limiting factor.
The *CPR* gene from *C. bombicola* was cloned (Genbank accession number EF050789) and its expression level was verified. *CPR* expression is not up-regulated upon sophorolipid synthesis, but it still is possible that the number of initially present CPR molecules is sufficient anyway to support the P450's involved in sophorolipid synthesis. This issue could be further investigated by overexpressing the species' own *CPR* gene and comparing the activity of the isolated microsomes to those of the wild type. If indeed the CPR enzyme turns out to be a rate limiting factor, overexpression could lead to increased and/or faster sophorolipid synthesis.

This PhD-thesis reports the first steps towards the genetic engineering of the sophorolipid producing yeast *C. bombicola*: a transformation- and selection system was developed, a constitutive and strong promoter for heterologous expression was evaluated and knock-outs were created. Furthermore, 7 complete genes of the formerly completely unknown *C. bombicola* genome were isolated. These molecular tools can be used to gain further insights into the sophorolipid biosynthesis and to consequently direct or modify the synthesis towards new-to-nature sophorolipid molecules. One future perspective is the synthesis of medium-chain sophorolipids by expressing a heterologous P450 with a strong specificity and activity towards medium-chain fatty acids. Several candidate genes can be found in the CYP52

family, but also other families harbour enzymes with hydroxylating activity towards fatty acids.

The presented molecular tools developed for *C. bombicola* can of course be used for other purposes instead of sophorolipid related research. For example, *C. bombicola* is an oleaginous yeast and this feature makes the species quite attractive for the production of β-hydroxyalkanoates or dicarboxylic acids. One of the strategies to achieve these molecules involves the use of a yeast blocked in its β-oxidation; such strain was created in the presented PhD-thesis.

One important bottleneck in the study and modification of sophorolipids or the use of *C. bombicola* for other purposes remains the unsolved genome sequence. It is quite laborious to pick up genes by means of degenerate PCR and genome walking and this strategy is only successful if the gene of interest is more or less conserved among different species. For example, the isolation of the glucosyltranferases, acetyltransferases and esterases involved in sophorolipid synthesis is a very challenging task. Furthermore, sequencing of the *C. bombicola* genome would make the study of the CYP52 family much easier and promoter regions of the sequenced genes could be useful for the expression of heterologous or homologous genes in *C. bombicola*.

# References


Abe Y, Inoue S, Ishida A (1981) Cosmetic composition of skin and hair treatment. US patent 427340

Akari S, Akari Y (1987) Method of modifying quality of wheat flour product. Japanese patent 61205449

Aki T, Yanagisawa S, Akanuma H (1997) Identification and characterization of positive regulatory elements in the human glyceraldehyde 3-phosphate dehydrogenase gene promoter. JBiochem 122, 271-278

Alani E, Cao L, Kleckner N (1987) A method for gene disruption that allows repeated use of ura3 selection in the construction of multiply disrupted yeast strains. Genetics 116, 541-545

Albrecht A, Rau U, Wagner F (1996) Initial steps of sophoroselipid biosynthesis by *Candida bombicola* ATCC 22214 grown on glucose. Appl Microbiol Biotechnol 46:67–73

Alderton JJ, Burr I, Mühlschlegel FA, Tuite MF (2006) Zeocin resistance as a dominant selective marker for transformation and targeted gene deletions in *Candida glabrata*. Mycoses 49, 445-451

Allingham RP (1971) Sophoroside esters in prepared food products. US patent 3622344

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

Arima K, Kakinuma A, Tamura G (1968) Surfactin, a crystalline peptide surfactant produced by *Bacillus subtilis*: isolation, characterisation and its inhibition of fibrin clot fomation. Biochem Biophys Res Commun 31, 488-494

Ashby RD, Nunez A, Solaiman DKY, Foglia TA (2005) Sophorolipid biosynthesis from a biodiesel co-product stream. J Am Oil Chem Soc 82, 625–630

Ashby R, Solaiman DKY, Foglia TA (2006) The use of fatty acids to enhance free acid sophorolipid synthesis. Biotechnol Lett 28, 253–260

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 biological active functionalized glycolipids. Bioconjug Chem 17, 1523–1529

Banat IM (1995) Biosurfactants Production and possible uses in microbial enhanced oil-recovery and oil pollution remediation - A review. Bioresour Technol 51, 1-12

Barnun S, Shneyour Y, Beckmann JS (1983) G-418, An Elongation Inhibitor Of 80-S Ribosomes. Biochim Biophys Acta 741, 123-127

Basehoar AD, Zanton SJ, Pugh BF (2004) Identification and distinct regulation of yeast TATA box-containing genes. Cell 116, 699-709

Baviere M, Degouy D, Lecourtier J (1994) Process for washing solid particles comprising a sophoroside solution. US patent 5326407

Beggs JD (1978) Transformation of yeast by a replicating hybrid plasmid. Nature 275, 104-109

Bergsson G, Arnfinnsson J, Steingrimsson O, Thormar H (2001) In vitro killing of *Candida albicans* by fatty acids and monoglycerides. Antimicrob Agents Chemother 45, 3209-3212

Bisht KS, Gross RA, Kaplan DL (1999) Enzyme-mediated regioselective acylations of sophorolipids. J Org Chem 64, 780–789

Bitter GA, Egan KM (1984) Expressioin of heterologous genes in S*accharomyces cerevisiae* from vectors utilizing the glyceraldehyde-3-phorsphate dehydrogenase gene promoter. Gene 32, 263-274

Black SD (1992) Membrane topology of the mammalian P450-cytochromes. Faseb J 6, 680-685

Bluth MH, Kandil E, Mueller CM, Shah V, Lin YY, Zhang H, Dresner L, Lempert L, Nowakowski M, Gross R, Schulze R, Zenilman ME (2006) Sophorolipids block lethal effects of septic shock in rats in a cecal ligation and puncture model of experimental sepsis. Crit Care Med 34, 188–195

Boeke JD, Lacroute F, Fink GR. (1984) A positive selection for mutants lacking orotidine-5'-phosphate decarboxylase activity in yeast - 5-fluoro-orotic acid resistance. Mol Gen Genet 197, 345-346

Borzeix CF (1999) Use of sophorolipids comprising diacetyl lactones as agent for stimulating skin fibroblast metabolism. World patent 99/62479

Brachmann CB, Davies A, Cost GJ, Caputo E, Li JC, Hieter P, Boeke JD (1998) Designer deletion strains derived from *Saccharomyces cerevisiae* S288C: a useful set of strains and plasmids for PCR-mediated gene disruption and other applications. Yeast 14, 115-132

Bradford MM (1976) Rapid and sensitive method for quantitation of microgram quantities of protein utilizing principle of protein-dye binding. Anal Biochem 72, 248-254

Brakemeier A, Lang S, Wullbrandt D, Merschel L, Benninghoven A, Buschmann N, Wagner F (1995) Novel sophorose lipids from microbial conversion of 2-alkanols. Biotechnol Lett 17, 1183–1188

Brakemeier A, Wullbrandt D, Lang S (1998a) *Candida bombicola*: production of novel alkyl glycosides based on glucose/2-dodecanol. Appl Microbiol Biotechnol 50,161–166

Brakemeier A, Wullbrandt D, Lang S (1998b) Microbial alkylsophorosides based on 1-dodecanol or 2-, 3-or 4-dodecanones. Biotechnol Lett 20, 215–218

Breithaupt TB, Light RJ (1982) Affinity–chromatography and further characterization of the glucosyltransferases involved in hydroxy-docosanoic acid sophoroside production in *Candida bogoriensis*. J Biol Chem 257, 9622–9628

Brett D, Howling D, Moris LJ, James AT (1971) Specificity of the fatty acid desaturases. The conversion of saturated to monoenoic acids. Arch Biochem Biophys 143, 535–547

Bucholtz ML, Light RJ (1976) Acetylation of 13-sophorosyloxydocosanoic acid by an acetyltransferase purified from *Candida bogoriensis*. J Biol Chem 251, 424–430

Carr JA, Bisht KS (2003) Enzyme-catalyzed regioselective trans-esterification of peracylated sophorolipids. Tetrahedron 59, 7713– 7724

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

Cavalero DA, Cooper DG (2003) The effect of medium composition on the structure and physical state of sophorolipids produced by *Candida bombicola* ATCC 22214. J Biotechnol 103, 31–41

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

Chen J, Song X, Zhang H, Qu YB, Miao JY (2006b) Sophorolipid produced from the new yeast strain *Wickerhamiella domercqiae* induces apoptosis in H7402 human liver cancer cells. Appl Microbiol Biotechnol 72, 52–59

Cirigliano MC, Carman GM (1984) Isolation of a bioemulsifier from *Candida lipolytica*.

Appl Environ Microbiol 48, 747-750
Cooper DG, Paddock DA (1984) Production of a biosurfactant from *Torulopsis bombicola*. Appl Environ Microbiol 47, 173–176
Craft DL, Madduri KM, Eshoo M, Wilson CR (2003) Identification and characterization of the CYP52 family of *Candida tropicalis* ATCC 20336, important for the conversion of fatty acids and alkanes to alpha, omega-dicarboxylic acids. Appl Environ Microbiol 69, 5983-5991
Daniel HJ, Otto RT, Reuss M, Syldatk C (1998a) Sophorolipid production with high yields on whey concentrate and rapeseed oil without consumption of lactose. Biotechnol Lett 20, 805– 807
Daniel HJ, Reuss M, Syldatk C (1998b) Production of sophorolipids in high concentration from deproteinized whey and rapeseed oil in a two stage fed batch process using *Candida bombicola* ATCC 22214 and *Cryptococcus curvatus* ATCC 20509. Biotechnol Lett 20, 1153–1156
Davies JT (1957) A quantitative kinetic theory of emulsion type, I. Physical chemistry of the emulsifying agent. Gas/Liquid and Liquid/Liquid Interface Proceedings of the International Congress of Surface Activity, 426-438
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, Monin N, Vandecasteele JP (1993) Identification and determination of individual sophorolipids in fermentation products by gradient elution high-performance liquid chromatography with evaporative light-scattering detection. J Chromatogr 648, 139–149
Davila AM, Marchal R, Vandecasteele JP (1994) Sophorose lipid production from lipidic precursors-Predictive evaluation of industrial substrates. J Ind 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
De Jonghe K, De Dobbelaere I, Sarrazyn R, Hofte M (2005) Control of *Phytophthora cryptogea* in the hydroponic forcing of witloof chicory with the rhamnolipid-based biosurfactant formulation PRO1. Plant Pathol 54, 219-226
De Koster C, Heerma W, Pepermans HAM, Groenewegen A, Peters H, Haverkamp J (1995) Tandem mass spectrometry and nuclear magnetic resonance spectroscopy studies of *Candida bombicola* sophorolipids and product formed on hydrolysis by cutinase. Anal Biochem 230, 135–148
De Maeseneire SL, De Groeve MRM, Dauvrin T, De Mey M, Soetaert W, Vandamme EJ (2006) Cloning, sequence analysis and heterologous expression of the *Myrothecium gramineum* orotidine-5'-monophosphate decarboxylase gene. FEMS Microbiol Lett 261, 262-271
Desai JD, Banat IM. (1997) Microbial production of surfactants and their commercial potential. Microbiol Mol Biol Rev 61, 47-59
Develter D, Fleurackers S (2007) A method for the production of short chained glycolipids. EP07101581.2
Develter D, Renkin M, Jacobs I (2007) Detergent compositions. EP1445302

Dmochowska A, Dignard D, Maleszka R, Thomas DY (1990) Structure and transcriptional control of the *Saccharomyces cerevisiae POX1* gene encoding acyl-coenzyme A oxidase. Gene 88, 247-252

Doring F, Klapper M, Theis S, Daniel H (1998) Use of the glyceraldehyde-3-phosphate dehydrogenase promotor for production of functional mammalian membrane transport proteins in the yeast Pichia pastoris. Biochem Biophys Res Commun 250, 531-535

Ducreux J, Ballerini D, Baviere M, Bocard C, Monin N (1997) Composition containing a surface active compound and glycolipids and decontamination process for a porous medium polluted by hydrocarbons. US patent 5654192

Edwards RJ, Murray BP, Singleton AM & Boobis AR (1991) Orientation of Cytochromes-P450 in the Endoplasmic-Reticulum. *Biochemistry* 30: 71-76

Eirich LD, Craft DL, Steinberg L, Asif A, Eschenfeldt WH, Stols L, Donnelly MI, Wilson CR (2004) Cloning and characterization of three fatty alcohol oxidase genes from *Candida tropicalis* strain ATCC 20336. Appl Environ Microbiol 70, 4872-4879

Enoch HG, Strittmatter P (1979) Cytochrome-b5 reduction by NADPH-cytochrome P-450 reductase. J Biol Chem 254, 8976-8981

Esders TW, Light RJ (1972) Glucosyl-and acetyltransferases involved in the biosynthesis of glycolipids from *Candida bogoriensis*.J Biol Chem 247, 1375–1386

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

Fickers P, Benetti PH, Wache Y, Marty A, Mauersberger S, Smit MS, Nicaud JM (2005) Hydrophobic substrate utilisation by the yeast *Yarrowia lipolytica*, and its potential applications. FEMS Yeast Res 5, 527-543

Fierro F, Laich F, Garcia-Rico RO, Martin JF (2004) High efficiency transformation of *Penicillium nalgiovense* with integrative and autonomously replicating plasmids. Int J Food Microbiol 90, 237-248

Finlayson SD, Fleming C, Berry DR, Johnston JR (1991) An improved lithium-acetate method for yeast transformation. Biotechnol Tech, 13-18.

Fleurackers SJJ (2006) On the use of waste frying oil in the synthesis of sophorolipids. Eur J Lipid Sci Technol 108, 5–12

Francois F, Chapeland-Leclerc F, Villard J, Noel T (2004) Development of an integrative transformation system for the opportunistic pathogenic yeast *Candida lusitaniae* using *URA3* as a selection marker. Yeast 21, 95-106

Futura T, Igarashi K, Hirata Y (2002) Low-foaming detergent compositions. World patent 03/002700

Garcia-Ochoa F, Casas JA (1997) Process for the production of sophorose by *Candida bombicola*. Spanish patent 2103688

Gi KE (2004) Red tide inhibiting agent comprising yellow earth and biodegradable sophorolipid. Korean patent 20040083614

Gietz RD, Schiestl RH (1995) Transforming yeast with DNA. Methods Mol Cell Biol 5, 255-269

Gobbert U, Lang S, Wagner F (1984) Sophorose lipid formation by resting cells of *Torulopsis bombicola*. Biotechnol Lett 6, 225–230

Gonzalez A, Jimenez A, Vazquez D, Davies JE, Schindler D (1978) Studies on the mode of action of hygromycin-B, an inhibitor of translocation in eukaryotes. Biochim Biophys Acta 521, 459-469

Goosen T, Bloemheuvel G, Gysler C, Debie DA, Vandenbroek HJW, Swart K (1987) Transformation of *Aspergillus niger* using the homologous orotidine-5'-phosphate-decarboxylase gene. Curr Genet 11, 499-503

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

Graber JH, Cantor CR, Mohr SC, Smith TF (1999) Genomic detection of new yeast pre-mRNA 3 '-end-processing signals. Nucleic Acids Res 27, 888-894

Griffin WC (1954) Calculation of HLB values of non-ionic surfactants. J Soc Cosmet Chem 5, 249-256

Guilmanov V, Ballistreri A, Impallomeni G, Gross RA (2002) Oxygen transfer rate and sophorose lipid production by *Candida bombicola*. Biotechnol Bioeng 77, 489–494

Gunther NW, Solaiman DKY, Fett FW (2007) Processes for the production of rhamnolipids. US patent US7202063

Guo ZJ, Sherman F (1996) 3'-end-forming signals of yeast mRNA. Trends Biochem Sci 21, 477-481

Hall PJ, Haverkamp J, Van Kralingen CG, Schmidt M (1996) Laundry detergent composition containing synergistic combination of sophorose lipid and nonionic surfactant. US patent 5520839

Hamilton R, Watanabe CK, Deboer HA (1987) Compliation and comparison of the sequence context around the aug startcodons in *Saccharomyces cerevisiae* messenger-RNAs. Nucleic Acids Res 15, 3581-3593

Hara A, Arie M, Kanai T, Matsui T, Matsuda H, Furuhashi K, Ueda M, Tanaka A (2001) Novel and convenient methods for *Candida tropicalis* gene disruption using a mutated hygromycin B resistance gene. Arch Microbiol 176, 364-369

Harris JI, Waters M (1976) Glyceraldehyde-3-phosphate dehydrogenase, p1-49. In Boyer PD (eds) The enzymes. Academic Press Inc., New York

Hasemann CA, Kurumbail RG, Boddupalli SS, Peterson JA, Deisenhofer J (1995) Structure and function of cytochromes P450: a comparative analysis of three crystal structures. Structure 2, 41-62

He F, Chen YT (2005) Cloning and heterologous expression of the NADPH cytochrome P450 oxidoreductase genes from an industrial dicarboxylic acid-producing *Candida tropicalis*. Yeast 22, 481-491

Higgins DG, Bleasby AJ, Fuchs R (1992) Clustal-V - improved software for multiple sequence alignment. Comput Appl Biosci 8, 189-191

Hildebrand PD (1989) Surfactant-like characteristics and identity of bacteria associated with broccoli head rot in Atlantic Canada. Can J Plant Pathol 11, 205-214

Hillion G, Marchal R, Stoltz C, Borzeix CF (1998) Use of a sophorolipid to provide free radical formation inhibiting activity or elastase inhibiting activity. US patent 5756471

Hirano T, Sato T, Yaegashi K, Enei H (2000) Efficient transformation of the edible basidiomycete *Lentinus edodes* with a vector using a glyceraldehyde-3-phosphate dehydrogenase promoter to hygromycin B resistance. Mol General Genet 263, 1047-1052

Hommel R, Stuwer O, Stuber W, Haferburg D, Kleber HP (1987) Production of water-soluble surface-active exolipids by *Torulopsis apicola*. Appl Microbiol Biotechnol 26, 199–205

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 (2001) Purification of lactonic sophorolipids by crystallization. J Biotechnol

87, 263–272
Hu YM, Ju LK (2003) Lipase-mediated deacetylation and oligomerization of lactonic sophorolipids. Biotechnol Prog 19, 303–311
Ilan Z, Ilan R, Cinti DL (1981) Evidence for a new physiological-role of hepatic NADPH-ferricytochrome (P-450) oxidoreductase - direct electron input to the microsomal fatty-acid chain elongation system. J Biol Chem 256, 66-72
Im JH, Yanagishita H, Ikegami T, Takeyama Y, Idemoto Y, Koura N, Kitamoto D (2003) Mannosylerythritol lipids, yeast glycolipid biosurfactants, are potential affinity ligand materials for human immunoglobulin. J Biomed Mater Res Part A 65, 379-385
Inoue S, Ito S (1982) Sophorolipids from *Torulopsis bombicola* as microbial surfactants in alkane fermentations. Biotechnol Lett 4, 3-8
Inoue S, Miyamoto N (1980) Process for producing a hydroxyfatty acid ester. US patent 4201844
Inoue S, Kiruma Y, Kinta M (1980) Dehydrating purification process for a fermentation product. US patent 4197166
Isoda H, Kitamoto D, Shinmoto H, Matsumura M, Nakahara T (1997) Microbial extracellular glycolipid induction of differentiation and inhibition of the protein kinase C activity of human promyelocytic leukemia cell line HL60. Biosci Biotechnol Biochem 61, 609–614
Ito S, Kinta M, Inoue S (1980) Growth of yeasts on normal-alkanes-Inhibition by a lactonic sophorolipid produced by *Torulopsis bombicola*. Agric Biol Chem 44, 2221–2223
Jarvis FG, Johnson MJ (1949) A glyco-lipid produced by *Pseudomonas aeruginosa*. Am Chem Soc 71, 4124-4126
Jones DF (1968) Microbiological oxidation of long-chain aliphatic compounds. Part V. Mechanism of hydroxylation. J Chem Soc (C), 2827–2833
Jones DF, Howe R (1968) Microbiological oxidation of long-chain aliphatic compounds. Part I. Alkanes and Alk-1-enes. J Chem Soc (C) 2801–2808
Kanayama N, Ueda M, Atomi H, Tanaka A (1998) Genetic evaluation of physiological functions of thiolase isoenzymes in the n-alkalane-assimilating yeast *Candida tropicalis*. J Bacteriol 180, 690-698
Kappeli O, Finnerty WR (1979) Partition of alkane by an extracellular vesicle derived from hexadecane-grown *Acinetobacter*. J Bacteriol 140, 707-712
Kargel E, Menzel R, Honeck H, Vogel F, Bohmer A, Schunck WH (1996) *Candida maltosa* NADPH-cytochrome P450 reductase: Cloning of a full-length cDNA, heterologous expression in *Saccharomyces cerevisiae* and function of the N-terminal region for membrane anchoring and proliferation of the endoplasmic reticulum. Yeast 12, 333-348
Kawano J, Suzuki T, Inoue S, Hayashi S (1981a) Powdered compressed cosmetic material. US patent 4305931
Kawano J, Suzuki T, Inoue S, Hayashi S (1981b) Stick-shaped cosmetic material. US patent 4305929
Kim SY, Oh DK, Lee KH, Kim JH (1997) Effect of soybean oil and glucose on sophorose lipid fermentation by *Torulopsis bombicola* in continuous culture. Appl Microbiol Biotechnol 48, 23–26
Kim HS, Kim YB, Lee BS, Kim EK (2005) Sophorolipid production by *Candida bombicola* ATCC 22214 from a corn-oil processing byproduct. J Microbiol Biotechnol 15, 55–58
Kimsey HH, Kaiser D (1992) The orotidine-5'-monophosphate decarboxylase gene of *Myxococcus xanthus* - Comparison to the OMP decarboxylase gene family. J Biol Chem 267, 819-824

Kitamoto D, Haneishi K, Nakahara T, Tabuchi T (1990) Production of mannosylerythritol lipids by *Candida antarctica* from vegetable-oils. Agric Biol Chem 54, 37-40.

Kitamoto D, Isoda H, Nakahara T (2002) Functions and potential applications of glycolipid biosurfactants - from energy-saving materials to gene delivery carriers. J Biosci Bioeng 94, 187-201

Kitamoto D, Yanagishita H, Haraya K, Kitamoto HK (1995) Effect of cerulenin on the production of mannosyl-erythritol lipids as biosurfactants by *Candida antarctica*. Biotechnol Lett 17, 25-30

Klekner V, Kosaric N, Zhou QH (1991) Sophorose lipids produced from sucrose. Biotechnol Lett 13, 345–348

Konishi M, Imura T, Fukuoka T, Morita T, Kitamoto D (2007) A yeast glycolipid biosurfactant, mannosylerythritol lipid, shows high binding affinity towards lectins on a self-assembled monolayer system. Biotechnol Lett 29, 473-480

Kozak M (1991) An analysis of vertebrate messenger-RNA sequences - Intimations of translational control. J Cell Biol 115, 887-903

Kuo CY, Chou SY, Huang CT (2004) Cloning of glyceraldehyde-3-phosphate dehydrogenase gene and use of the gpd promoter for transformation in Flammulina velutipes. Appl Microbiol Biotechnol 65, 593-599

Kyte J, Doolittle RF (1982) A simple method for displaying the hydropathic character of a protein. J Mol Biol 157,105-132

Lang S (2002) Biological amphiphiles (microbial biosurfactants). Curr Opin Colloid Interface Sci 7, 12-20

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

Lang S, Brakemeier A, Wullbrandt D, Seiffert-Stoericko A (2002) Sophoroselipids, method for their production and use. US patent 6433152

Lang S, Katsiwela E, Wagner F (1989) Antimicrobial effects of biosurfactants. Fett Wiss Technol Fat Sci Technol 91, 363–366

Lang S, Philp JC (1998) Surface-active lipids in *Rhodococc*i. Antonie Van Leeuwenhoek 74, 59-70

Lang S, Wagner G (1987) Structure and properties of biosurfctants, p 21-47. In Kosaric N, Cairns WL, Gray, NCC (eds.) Biosurfactants and biotechnology. Marcel Dekker, Inc., New York

Lee KH, Kim JH (1993) Distribution of substrates carbon in sophorose lipid production by *Torulopsis bombicola*. Biotechnol Lett 15, 263–266

Lin SC (1996) Biosurfactants: recent advances. J Chem Technol Biotechnol 66, 109–120

Lottermoser K, Schunck WH, Asperger O (1996) Cytochromes P450 of the sophorose lipid-producing yeast *Candida apicola*: Heterogeneity and polymerase chain reaction-mediated cloning of two genes. Yeast 12, 565–575

Mager H, Röthlisberger R, Wagner F (1987) Use of sophorose-lipid lactone for the treatment of dandruffs and body odeur. European patent 0209783

Maier RM, Soberon-Chavez G (2000) *Pseudomonas aeruginosa* rhamnolipids: biosynthesis and potential applications. Appl Microbiol Biotechnol 54, 625-633

Maingault M (1999) Utilization of sophorolipids as therapeutically active substances or cosmetic products, in particular for the treatment of the skin. US patent 5981497

Mann RM, Bidwell JR (2001) The acute toxicity of agricultural surfactants to the tadpoles of four Australian and, two exotic frogs. Environ Pollut 114, 195–205

Mann RM, Boddy MR (2000) Biodegradation of a nonylphenol ethoxylate by the autochthonous microflora in lake water with observations on the influence of light. Chemosphere 41, 1361– 1369

Marchal R, Lemal J, Sulzer C, Davila AM (1999) Production of sophorolipid acetate acids from oils or esters. US patent 5900366

Masaru K, Takashi N, Yoji A, Kazuo N, Tatsu N, Sumiko T, Jotaro N (2001) Composition for high-density cold storage transportation. Japanese patent 2001131538

Masuda Y, Park SM, Ohta A, Takagi M (1995) Cloning and characterization of the *POX2* gene in *Candida maltosa*. Gene 167, 157-161

Matsuyama T, Sogawa M, Yano I (1987) Direct colony thin-layer chromatography and rapid characterization of *Serratia marcescens* mutants defective in production of wetting agents. Appl Environ Microbiol 53, 1186-1188

McAlister L, Holland MJ (1985) Differential expression of the three yeast glyceraldehyde-3-phosphate dehydrogenase genes. J Biol Chem 260, 15019-15027

Mulligan CN, Gibbs BF (1990) Recovery of biosurfactants by ultrafiltration. J Chem Technol Biotechnol 47, 23–29

Mulligan CN, Yong RN, Gibbs BF (2001) Heavy metal removal from sediments by biosurfactants. J Hazard Mater 85, 111–125

Napolitano LM (2006) Sophorolipids in sepsis: Antiinflammatory or antibacterial? Crit Care Med 34, 258–259

Nebert DW, Gonzalez FJ (1987) P450 genes: structure, evolution, and regulation. Annu Rev Biochem 56, 945–993

Nelson DR (1998) Cytochrome P450 nomenclature. Methods Mol Biol 107, 15–24

Nelson DR (1999) Cytochrome P450 and the individuality of species. Arch Biochem Biophys 369, 1–10

Nelson DR, Strobel HW (1989) Secondary structure prediction of 52 membrane bound cytochromes P450 shows a strong structural similarity to P450cam. Biochem 28,656-660

Neuberger G, Maurer-Stroh S, Eisenhaber B, Hartig A, Eisenhaber F (2003) Prediction of peroxisomal targeting signal 1 containing proteins from amino acid sequence. J Mol Biol 328, 581-592

Ngan WY, Nga BH, Pridmore D, Mollet B, Tan HM (2000) Transformation of *Endomyces fibuliger* based on its gene for orotidine-5 '-phosphate decarboxylase. Gene 254, 97-103

Nunez A, Foglia TA, Ashby R (2003) Enzymatic synthesis of a galactopyranose sophorolipid fatty acid-ester. Biotechnol Lett 25, 1291–1297

Ogawa S, Ota Y (2000) Influence of exogenous natural oils on the omega-1 and omega-2 hydroxy fatty acid moiety of sophorose lipid produced by *Candida bombicola*. Biosci Biotechnol Biochem 64, 2466-2468

Ohama T, Suzuki T, Mori M, Osawa S, Ueda T, Watanabe K, Nakase T (1993) Nonuniversal decoding of the leucine codon cug in several *Candida* species. Nucleic Acids Res 21, 4039-4045

Ohkawa H, H. I, Shiota N, Yamada T, Inui H, Ohkawa Y (1998) Molecular mechanisms of herbicide resistance with special emphasis on cytochrome P450 monooxygenases. Plant Biotechnol 15, 173-176

Ohkuma M, Masuda Y, Park SM, Ohtomo R, Ohta A, Takagi M (1995) Evidence that the expression of the gene for nadph-cytochrome p-450 reductase is n-alkane-inducible in *Candida maltosa*. Biosci Biotechnol Biochem 59, 1328-1330

Okazaki K, Tan H, Fukui S, Kubota I, Kamiryo T (1987) Peroxisomal acyl-coenzyme A oxidase multigene family of the yeast *Candida tropicalis*; nucleotide sequence of a third gene and its protein product. Gene 58, 37-44

Omura S (1976) The antibiotic cerulenin, a novel tool for biochemistry as an inhibitor of fatty acid synthesis. Bacteriol Rev 40, 681-697

Ono T, Ozasa S, Hasegawa F, Imai Y (1977) Involvement of NADPH-cytochrome c-reductase in rat-liver squalene epoxidase system. Biochim Biophys Acta 486: 401-407.

Parra JL, Guinea J, Manresa MA, Robert M, Mercade ME, Comelles F, Bosch MP (1989) Chemical characterization and physicochemical behavior of biosurfactants. J Am Oil Chem Soc 66, 141-145

Pekin G, Vardar-Sukan F, Kosaric N (2005) Production of sophorolipids from *Candida bombicola* ATCC 22214 using Turkish corn oil and honey. Eng Life Sci 5, 357–362

Pellecier F, André P (2004) Cosmetic use of sophorolipids as subcutaneous adipose cushion regulation agents and slimming application. World patent 2004/108063

Pesce L (2002) A biotechnological method for the regeneration of hydrocarbons from dregs and muds, on the base of biosurfactants. World patent 02/062495

Pesole G, Lotti M, Alberghina L, Saccone C (1995) Evolutionary origin of nonuniversal cug(ser) codon in some *Candida* Species as inferred from a molecular phylogeny. Genetics 141, 903-907

Phadtare S, Parekh P, Shah S, Tambe A, Joshi R, Sainkar SR, Prabhune A, Sastry M (2003) *Candida bombicola* cells immobilized on patterned lipid films as enzyme sources for the transformation of arachidonic acid to 20-HETE. Biotechnol Prog 19, 1659-1663

Phadtare S, Shah S, Prabhune A, Wadgaonkar PP, Sastry M (2004) Immobilization of *Candida bombicola* cells on free-standing organic-gold nanoparticle membranes and their use as enzyme sources in biotransformations. Biotechnol Prog 20, 1817-1824

Piechaczyk M, Blanchard JM, Marty L, Dani C, Panabieres F, El Sabouty S, Fort P, Jeanteur P (1984) Post-transcriptional regulation of glyceraldehyde-3-phosphate dehydrogenase gene expression in rat tissues. Nucleic Acids Res 12, 6951-6963

Pierce D, Heilman TJ (1998) Germicidal composition. World patent 98/16192

Pompon D, Louerat B, Bronine A, Urban P (1996) Yeast expression of animal and plant P450s in optimized redox environments. Method Enzymol 272, 51-64

Punt PJ, Dingemanse MA, Kuyvenhoven A, Soede RD, Pouwels PH, Van den Hondel CA. (1990) Functional elements in the promoter region of the *Aspergillus nidulans gpdA* gene encoding glyceraldehyde-3-phosphate dehydrogenase. Gene 93, 101-109

Qin YM, Marttila MS, Haapalainen AM, Siivari KM, Glumoff T, Hiltunen JK (1999) Yeast peroxisomal multifunctional enzyme: (3R)-hydroxyacyl-CoA dehydrogenase domains A and B are required for optimal growth on oleic acid. J Biol Chem 274, 28619-28625

Rapp P, Bock H, Wray V, Wagner F (1979) Formation, isolation and characterization of trehalose dimycolates from *Rhodococcus erythropolis* grown on n-alkanes. J Gen Microbiol 115, 491-503

Rau U, Manzke C, Wagner F (1996) Influence of substrate supply on the production of sophorose lipids by *Candida bombicola* ATCC 22214. Biotechnol Lett 18, 149–154

Rau U, Heckmann R, Wray V, Lang S (1999) Enzymatic conversion of a sophorolipid into a glucose lipid. Biotechnol Lett 21, 973–977

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

Renkin M (2003) Environmental profile of sophorolipid and rhamnolipid biosurfactants. La

rivista Italiana delle sostanze grasse 58, 249-252.
Rodrigues L, Banat IM, Teixeira J, Oliveira R (2006) Biosurfactants: potential applications in medicine. J Antimicrob Chemother 57, 609-618
Rodriguez L, Chavez FP, Gonzalez ME, Basabe L, Rivero T (1998) Isolation and sequence analysis of the orotidine-5 '-phosphate decarboxylase gene (*URA3*) of *Candida utilis*. Comparison with the OMP decarboxylase gene family. Yeast 14, 1399-1406
Rosa C, Lachance M (1998) The yeast genus *Starmerella* gen. nov. and *Starmerella bombicola* sp. nov., the teleomorph of *Candida bombicola* (Spencer, Gorin & Tullock) Meyer & Yarrow. Int J Syst Bacteriol 48, 1413–1417
Rosa CA, Lachance MA, Silva JOC, Teixeira ACP, Marini MM, Antonini Y, Martins RP (2003) Yeast communities associated with stingless bees. Fems Yeast Res 4, 271-275
Rose M, Grisaf P, Botstein D (1984) Structure and function of the yeast *URA3* gene - Expression in *Escherichia coli*. Gene 29, 113-124
Rosenberg E, Rubinovitz C, Legmann R, Ron EZ (1988) Purification and chemical-properties of *Acinetobacter calcoaceticus* A2 biodispersan. Appl Environ Microbiol 54, 323-326
Ruiz-Dìez B (2002) Strategies for the transformation of filamentous fungi. J Appl Microbiol 92,189-195
Sakai Y, Kazarimoto T, Tani Y (1991) Transformation system for an asporogenous methylotrophic yeast, *Candida boidinii* - Cloning of the orotidine-5'-phosphate decarboxylase gene (*URA3*), isolation of uracil auxotrophic mutants, and use of the mutants for integrative transformation. J Bacteriol 173, 7458-7463
Sambrook J, Russell DW (2001) Molecular Cloning: A Laboratory Manual. 3rd edition. Cold Spring Harbour Laboratory Press: Cold Spring Harbour, New York
Sanchez M, Teruel JA, Espuny MJ, Marques A, Aranda FJ, Manresa A, Ortiz A (2006) Modulation of the physical properties of dielaidoylphosphatidylethanolamine membranes by a dirhamnolipid biosurfactant produced by *Pseudomonas aeruginosa*. Chem Phys Lipids, 118-127
Sanglard D, Sengstag C, Seghezzi W (1993) Probing the membrane topology of *Candida tropicalis* cytochrome-P450. Eur J Biochem 216, 477-485
Schacter BA, Nelson EB, Masters BSS, Marver HS (1972) Immunochemical evidence for an association of heme oxygenase with microsomal electron-transport system. J Biol Chem 247, 3601-3607
Scheller U, Zimmer T, Becher D, Schauer F, Schunck WH (1998) Oxygenation cascade in conversion of n-alkanes to alpha,omega-dioic acids catalyzed by cytochrome P450 52A3. J Biol Chem 273, 32528-32534
Seghezzi W, Sanglard D, Fiechter A (1991) Characterization of a 2nd alkane-inducible cytochrome-P450-encoding gene, CYP52a2, from *Candida tropicalis*. Gene 106, 51-60
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
Shulga AN, Karpenko EV, Eliseev SA, Turovsky AA (1993) The method for determination of aniogenic bacterial surfade-active peptidolipids. Microbiol J 55, 85-88
Singh P, Cameotra SS (2004) Potential applications of microbial surfactants in biomedical sciences. Trends Biotechnol 22, 142-146
Singh SK, Felse AP, Nunez A, Foglia TA, Gross RA (2003) Regioselective enzyme-catalyzed synthesis of sophorolipid esters, amides, and multifunctional monomers. J Org Chem 68, 5466–5477

Solaiman DKY, Ashby RD, Nunez A, Foglia TA (2004) Production of sophorolipids by *Candida bombicola* grown on soy molasses as substrate. Biotechnol Lett 26, 1241–1245

Spencer JFT, Gorin PAJ, Tulloch AP (1970) *Torulopsis bombicola* sp.n. Antonie Van Leeuwenhoek 36, 129–133

Stanghellini ME, Miller RM (1997) Biosurfactants: their identity and potential efficacy in the biological control of zoosporic plant pathogens. Plant Disease 81, 4-12

Suzuki M, Suh SO, Sugita T, Nakase T (1999) A phylogenetic study on galactose-containing *Candida* species based on 18S ribosomal DNA sequences. J Gen Appl Microbiol 45(5):229-23

Syldatk C, Lang S, Wagner F, Wray V, Witte L (1985) Chemical and physical characterization of 4 interfacial-active rhamnolipids from *Pseudomonas spec* Dsm 2874 grown on normal-alkanes. Zeitschrift fur Naturforschung C-A Journal of Biosciences 40, 51-60

Tuite MF, Santos MAS (1996) Codon reassignment in *Candida* species: an evolutionary conundrum. Biochimie 78, 993-999

Tulloch AP, Spencer JFT (1968) Fermentation of long-chain compounds by *Torulopsis apicola*. IV. Products from esters and hydrocarbons with 14 and 15 carbon atoms and from methyl palmitoleate. Can J Chem 46, 1523–1528

Tulloch AP, Spencer JFT, Gorin PAJ (1962) The fermentation of long-chain compounds by *Torulopsis magnoliae*. Can J Chem 40, 1326–1338

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

Tulloch AP, Hill A, Spencer JFT (1968b) Structure and reactions of lactonic and acidic sophorosides of 17-hydroxyoctadecanoic acid. Can J Chem 46, 3337–3351

Van Bogaert INA, De Maeseneire SL, De Schamphelaire W, Develter D, Soetaert W, Vandamme EJ (2007a) Cloning, characterisation and functionality of the orotidine-5′-phosphate decarboxylase gene (URA3) of the glycolipid producing yeast *Candida bombicola*. Yeast 24, 201–208

Van Bogaert INA, Develter D, Soetaert W, Vandamme EJ (2007b) Cloning and characterisation of the NADPH cytochroom P450 redutase gene (*CPR*) from *Candida bombicola*. FEMS Yeast Res 7, 922-928

Vandenbrink HJM, Vanzeijl CMJ, Brons JF, Vandenhondel C, Vangorcom RFM (1995) Cloning and characterization of the NADPH cytochrome-P450 oxidoreductase gene from the filamentous fungus *Aspergillus niger*. Dna Cell Biol 14, 719-729

Van der Vegt W, Vandermei HC, Noordmans J, Busscher HJ (1991) Assessment of bacterial biosurfactant production through axisymmetrical drop shape-analysis by profile. Appl Microbiol Biotechnol 35, 766-770

Van Wert SL, Yoder OC (1992) Structure of the *Cochliobolus heterostrophus* glyceraldehyde-3-phospate dehydrogenase gene. Curr Genet 22, 29-35

Yadav JS, Loper JC (1999) Multiple P450alk (cytochrome P450 alkane hydroxylase) genes from the halotolerant yeast *Debaryomyces hansenii*. Gene 226, 139-146

Yadav JS, Loper JC (2000) Cloning and characterization of the cytochrome P450 oxidoreductase gene from the zygomycete fungus *Cunninghamella*. Biochem Biophys Res Commun 268, 345-353

Yang VW, Marks JA, Davis BP, Jeffries TW (1994) High-efficiency transformation of *Pichia stipitis* based on its *URA3* gene and a homologous autonomous replication sequence, ARS2. Appl Environ Microbiol 60, 4245-4254

Yoo DS, Lee BS, Kim EK (2005) Characteristics of microbial biosurfactant as an antifungal agent against plant pathogenic fungus. J Microbiol Biotechnol 15, 1164–1169

Wakamatsu Y, Zhao XX, Jin CY, Day N, Shibahara M, Nomura N, Nakahara T, Murata T, Yokoyama KK (2001) Mannosylerythritol lipid induces characteristics of neuronal differentiation in PC12 cells through an ERK-related signal cascade. Eur J Biochem 268, 374-383

Wakil SJ (1989) Fatty acid synthase, a proficient multifunctional enzyme. Biochem 28, 4523−4530

Wang HJJ, Le Dall MT, Wache Y, Laroche C, Belin JM, Gaillardin C, Nicaud JM (1999a) Evaluation of acyl coenzyme A oxidase (Aox) isozyme function in the n-alkane-assimilating yeast *Yarrowia lipolytica*. J Bacteriol 181, 5140-5148

Wang HJ, Le Dall MT, Wache Y, Laroche C, Belin JM, Nicaud JM (1999b) Cloning, sequencing, and characterization of five genes coding for Acyl-CoA oxidase isozymes in the yeast *Yarrowia lipolytica*. Cell Biochem Biophys 31,165-174

Wang XX, Li G, Deng YT, Yu XW, Chen F (2006) A site-directed integration system for the nonuniversal CUG(Ser) codon usage species *Pichia farinosa* by electroporation. Arch Microbiol 184, 419-424

Wesolowski-Louvel M, Tanguy-Rougeau C, Fukuhara H (1988) A nuclear gene required for the expression of the linear DNA-associated killer system in the yeast *Kluyveromyces lactis*. Yeast 4, 71-81

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

Wullbrandt D, Ciani C, Brakemeier A, Lang S, Wagner F (1996) Glucose- and sophorose-lipids, a process for their preparation and their use. US5767255

Zerkowski JA, Solaiman DKY (2006) Synthesis of polyfunctional fatty amines from sophorolipid-derived 17-hydroxy oleic acid. J Am Oil Chem Soc 83, 621–628

Zerkowski JA, Solaiman DKY, Ashby R, Foglia TA (2006) Head group-modified sophorolipids: Synthesis of new cationic, zwitterionic, and anionid surfactants. J Surfactants Deterg 9, 57–62

Zhang B, Carlson R, Srienc F (2006) Engineering the monomer composition of polyhydroxyalkanoates synthesized in *Saccharomyces cerevisiae*. Appl Environ Microbiol 72, 536-543

Zhang L, Somasundaran P, Singh SK, Felse AP, Gross R (2004) Synthesis and interfacial properties of sophorolipid derivatives. Colloid Surf A Physicochem Eng Asp 240, 75–82

Zhou QH, Kosaric N (1993) Effect of lactose and olive oil on intracellular and extracellular lipids of *Torulopsis bombicola*. Biotechnol Lett 15, 477–482

Zhou QH, Kosaric N (1995) Utilization of canola oil and lactose to produce biosurfactant with *Candida bombicola*. J Am Oil Chem Soc 72, 67–71

Zhou QH, Kleckner N, Kosaric N (1992) Production of sophorose lipids by *Torulopsis bombicola* from safflower oil and glucose. J Am Oil Chem Soc 69, 89–91

Zimmer T, Iida T, Schunck WH, Yoshida Y, Ohta A, Takagi M (1998) Relation between evolutionary distance and enzymatic properties among the members of the CYP52A subfamily of *Candida maltosa*. Biochem Biophys Res Commun 251, 244-247

# Summary

Surfactants are widely used in the food, pharmaceutical, cosmetic and cleaning industries. Throughout the years, surfactants have been produced from petrochemical raw materials, but during the last decades the environmental awareness has become a more important issue in the study, development and application of surfactants. In this respect, microorganisms are a potential source of (novel) biodegradable surfactants. With the advantages of biodegradability, low ecotoxicity and the production based on renewable-resource substrates, biosurfactants may eventually replace their chemically synthesised counterparts.

A promising group of biosurfactants are sophorolipids. They are produced by yeasts such as *Candida bombicola* and consist of the dissacharide sophorose linked to a long-chain hydroxy fatty acid.

Despite the fact that sophorolipids are actually a mixture of slightly different molecules, they posses little variation in the length of the lipid tail. The range is limited to C16-C18 fatty acids and is governed by the specificity of one or more cytochrome P450 monooxygenases (P450). These enzymes hydroxylate the fatty acids at the terminal (ω) or subterminal (ω-1) position and transform them into suitable substrates for glucosyltransferase I, which will form an ester bond between the introduced hydroxyl group and the first glucose molecule. However, the incorporation of fatty acids differing from the conventional C16-C18 range (more in particular medium-chain fatty acids) could broaden up the application potential of sophorolipids. In order to achieve this, the substrate specificity of the cytochrome P450 monooxygenase must be circumvented.

This can be achieved on two ways. The first method comprises the addition of already hydroxylated substrates, in this way skipping the controlling effect of the P450 enzyme. The other method comprises the incorporation of unconventional stearic acid-like substrates which can undergo post-fermentative modification to give rise to shorter-chained sophorolipids.

Both methods resulted in the synthesis of medium-chain sophorolipids, although with different yields. It is believed that certain substrates are partially metabolized in the β-oxidation pathway.

Another way to eventually obtain medium-chain sophorolipids is the genetic modification of the producing yeast strain *C. bombicola*. Nevertheless, no molecular tools were available to achieve this objective. Therefore, a transformation and selection system based on the auxotrophic ura3 marker was developed and the promoter sequence of the *C. bombicola* GAPD gene was obtained in order to allow constitutive and high expression of heterologous genes.

The created transformation and selection system was used to construct a *C. bombicola* knock-out strain with a blocked β-oxidation route. The obtained strains were evaluated for sophorolipid synthesis on 1-dodecanol and rapeseed oil. Better yields were indeed achieved for production on 1-dodecanol, but unexpectedly, the yields for the rapeseed fermentations were inferior.

Furthermore, another pathway interfering with medium-chain sophorolipid synthesis - *de novo* sophorolipid synthesis - was inhibited on purpose. Hereto, the antibiotic cerulenin was used and this indeed resulted in synthesis of medium-chain sophorolipids without the presence of long-chain *de novo* sophorolipids.

Finally, the enzymes controlling the sophorolipid fatty acid tail, the cytochrome P450 monooxygenases, were investigated. Three *P450* genes of *C. bombicola* could completely be isolated, among them one which is possibly involved in sophorolipid synthesis. The exact function of the gene must however be further verified by enzyme activity tests and the construction of a deletion strain. However, there are at least five other very similar *P450* genes in the *C. bombicola* genome with an unknown function. In order to gain a total and clear view on the P450 family of *C. bombicola*, it would be recommendable to know the full gene sequences of all these members. Ideally, they can be achieved upon sequencing of the whole *C. bombicola* genome.

# Samenvatting

Surfactantia of oppervakte-actieve stoffen worden op grote schaal gebruikt in de voedings-, cosmetica-, pharma- en reinigingsindustrie. Momenteel worden deze moleculen hoofdzakelijk geproduceerd via chemische processen uitgaande van petrochemische grondstoffen. In het kader van de huidige milieuproblematiek komen de bio-surfactantia echter versterkt in de belangstelling wegens hun lage toxiciteit en goede biodegradeerbaarheid. Eén van de meestbelovende bio-surfactantia zijn de sophorolipiden. Ze worden geproduceerd door de niet-pathogene gist *Candida bombicola* en bestaan uit het disaccharide sophorose waaraan een langketen hydroxy-vetzuur gekoppeld is.

Ondanks het feit dan sophorolipiden in feite bestaan uit een mengsel van sterk gelijkaardige moleculen, vertonen ze weinig variatie in de lengte van hun vetzuurstaart. Deze is gelimiteerd tot C16-C18 vetzuren en wordt gecontroleerd door de specificiteit van één of meerdere cytochroom P450 mono-oxygenases (P450). Deze enzymen hydroxyleren het vetzuur op de terminale (ω) of subterminale (ω-1) positie en zetten het op deze manier om in een geschikt substraat voor glucosyltransferase I. Dit enzym vormt op zijn beurt een ester binding tussen de geïntroduceerde hydroxylgroep en de eerste glucose molecule. De incorporatie van vetzuren met een andere ketenlengte (in het bijzonder middelketen vetzuren) zou echter het toepassingsgebied van de sophorolipiden kunnen verruimen. Om dit te bereiken dient men de substraatspecificiteit van het cytochroom P450 mono-oxygenase te omzeilen.

Dit kan op twee manieren. De eerste methode bestaat uit de toediening van reeds gehydroxyleerde substraten waardoor de controlerende actie van het P450 enzym wordt overgeslaan. De andere methode behelst de integratie van speciale stearinezuur-achtige substraten, welke na post-fermentatieve modificatie aanleiding geven tot sophorolipiden met kortere vetzuurstaarten. Beide methoden werden met succes toegepast, alhoewel er verschillen waren in de behaalde opbrengsten; waarschijnlijk worden bepaalde substraten gedeeltelijk gemetaboliseerd in de β-oxidatie pathway.

Een andere manier om middelketen sophorolipiden te bekomen is de genetische modificatie van de producerende giststam *C. bombicola*, maar spijtig genoeg waren hiervoor nog geen moleculaire technieken voorhanden. Daarom werd een transformatie- en selectiesysteem

ontwikkeld, gebaseerd op de auxotrofe ura3 merker en werd de promotersequentie van het *C. bombicola GAPD* gen gecloneerd zodat constitutieve en hoge expressie van heterologe genen mogelijk wordt.
Het ontwikkelde transformatie- en selectiesysteem werd gebruikt voor de constructie van een *C. bombicola* knock-out stam met een geblokkeerde β-oxidatie route. De verkregen stammen werden getest op hun sophorolipidesynthese op 1-dodecanol en raapzaadolie. De knock-out stammen presteerden inderdaad beter op 1-dodecanol, maar vreemd genoeg werden voor de fermentaties op raapzaadolie slechtere resultaten behaald dan voor de wild type stam.
Vervolgens werd ook een andere pathway die interfereert met middelketen sophorolipidesynthese geïnhibeerd: de *de novo* sophorolipide synthese. Hiervoor werd het antibioticum cerulenine gebruikt, wat inderdaad resulteerde in de synthese van middelketen sophorolipiden zonder de aanwezigheid van de *de novo* langketen sophorolipiden.

Ten slotte werden ook de enzymen onderzocht die door hun specificiteit de ketenlengte van de sophorolipide vetzuurstaart controleren: de cytochroom P450 mono-oxygenasen. Drie dergelijke genen werden geïsoleerd uit het *C. bombicola* genoom en één van deze speelt mogelijks een rol in de sophorolipidesynthese. De exacte functie van dit gen dient echter nog verder te worden onderzocht door middel van activiteitstesten en de constructie van een deletie-stam.
Er zijn echter nog minstens vijf andere, zeer sterk op elkaar gelijkende *P450* genen aanwezig in het *C. bombicola* genoom, waarvan de functie niet gekend is. Om een duidelijk en totaal overzicht te verkrijgen over de P450 familie van *C. bombicola*, is het aangewezen te beschikken over de volledige gensequenties van al deze genen. Deze kunnen idealerwijze bekomen worden door sequenering van het volledige *C. bombicola* genoom.

# Curriculum Vitae

## 1 PERSONALIA


| | |
|---|---|
| Name: | Van Bogaert Inge Noëlle Adriënne |
| Professional address: | Laboratory of Industrial Microbiology and Biocatalysis,<br>Department of Biochemical and Microbial Technology,<br>Faculty of Bioscience Engineering, Ghent University,<br>Coupure Links 653<br>9000 Ghent<br>Belgium |
| Telephone: | +32(0)9/264.60.34 |
| Fax: | +32(0)9/264.62.31 |
| e-mail: | Inge.VanBogaert@UGent.be |


## 2 EDUCATION

1993 – 1999: Latin-Mathematics (8h)
Sint-Lodewijks College, Lokeren, Belgium

1999 – 2004: Studies at the Faculty of Bio-Science Engineering, Ghent University, Belgium

- 1999 – 2000: First year Bioscience Engineering, greatest distinction (83.4 %).
- 2000 – 2001: Second year Bioscience Engineering, greatest distinction (82.3 %).
- 2001 – 2002: First year Bioscience Engineering, option Cell & Gene biotechnology greatest distinction (82.5 %).
- 2002 – 2003: Second year Bioscience Engineering, option Cell & Gene biotechnology The second term was spend at the Swedish University of Agricultural Sciences, Uppsala, Sweden. Great distinction (77.3 %).
- 2003 – 2004: Third year Bioscience Engineering, option Cell & Gene biotechnology Dissertation on "Detection of the minimal residual disease in neuroblastoma patients", at the University Hospital, Ghent (Promoters: Prof. Dr. Patrick Van Oostveldt and Prof Dr. Jan Philippé). Great distinction (81.4 %).

## 3 PROFESSIONAL ACTIVITIES

2004 – Present: Doctoral research, financially supported by Ecover Belgium NV (Malle, Belgium) and the *Bijzonder onderzoeksfonds* (BOF) from Ghent University
Research topic: Microbial synthesis of sophorolipids by the yeast *Candida bombicola*
Laboratory of Industrial Microbiology and Biocatalysis, Department of Biochemical and Microbial Technology, Faculty of Bioscience Engineering, Ghent University, Ghent, Belgium (Promoters: Prof. Dr. ir. Erick Vandamme and Prof. Dr. ir. Wim Soetaert)

Student guidance: laboratory exercises "General Microbiology"

M.Sc. thesis guidance:

- Wouter De Schamphelaire, 2005-2006. Synthesis of short chain sophorolipids. Faculty of Bioscience Engineering, Ghent University.
- Marjolijn Vanzandweghe, 2005-2006. Microbial synthesis of sophorolipids by *Candida bombicola*. Hogeschool Gent
- Katrien Geunes, 2006-2007. Cloning and characteristation of cytochrome P450 monooxygenases of *Candida bombicola*. Faculty of Bioscience Engineering, Ghent University.
- Dries Robberecht, 2006-2007. Development of a genetic engineering system for *Candida bombicola*. Faculty of Bioscience Engineering, Ghent University.
- Ellen Verbeke, 2006-2007. Study and genetic engineering of *Candida bombicola* for the synthesis of short chain sophorolipids. Kaho Sint-Lieven.
- Magdalena Mielczarek, 2007-2008. Microbial synthesis of sophorolipids by *Candida bombicola* mutant strains. Faculty of Bioscience Engineering, Ghent University.

Co-promoter M.Sc. thesis:

- Katrien Geunes, 2006-2007. Cloning and characteristation of cytochrome P450 monooxygenases of *Candida bombicola*. Faculty of Bioscience Engineering, Ghent University.
- Dries Robberecht, 2006-2007. Development of a genetic engineering system for *Candida bombicola*. Faculty of Bioscience Engineering, Ghent University.

Member of the organising committee 'Renewable Resources and Biorefineries', Ghent, September, 19-21, 2005.

Member of the organising committee 'Renewable Resources and Biorefineries', Ghent, June, 4-6, 2007.

Co-worker of collaborative project with an industrial partner on the production of bioethanol with *Saccharomyces cerevisiae*. October-December, 2007.

## 4 PUBLICATIONS

### *4.1 A1 publications*

**Van Bogaert, I.N.A.**, De Maeseneire, S.L., De Schamphelaire, W., Develter, D., Soetaert, W. & Vandamme, E.J. (2007). Cloning, characterisation and functional analysis of the orotidine-5'-phosphate decarboxylase gene (*URA3*) of the glycolipid producing yeast *Candida bombicola*. *Yeast* **24**: 201-208.

(I.F. 1.955)

**Van Bogaert, I.N.A.**, W., Develter, D., Soetaert, W. & Vandamme, E.J. (2007). Cloning and characterisation of the NADPH cytochroom P450 redutase gene (*CPR*) from *Candida bombicola*. *FEMS Yeast Research* **7**: 922-928.

(I.F. 2.274)

**Van Bogaert, I.N.A.**, Saerens, K., De Muynck, C., Develter, D., Soetaert, W. & Vandamme, E.J. (2007). Microbial production and application of sophorolipids. *Applied Microbiology and Biotechnology* **76**: 23-34.

(I.F. 2.441)

**Van Bogaert, I.N.A.,** De Maeseneire, S.L., Develter, D., Soetaert, W. & Vandamme, E.J. (2008). Development of a transformation and selection system for the glycolipid producing yeast *Candida bombicola*. *Yeast* **25:** 273-278.

(I.F. 1.955)

**Van Bogaert, I.N.A.,** De Maeseneire, S.L., Develter, D., Soetaert, W. & Vandamme, E.J. (2008). Cerulenin inhibits the de novo sophorolipid synthesis of *Candida bombicola*. *Biotechnology Letters* (accepted).

(I.F. 1.134)

De Muynck, C., Van der Borght, J., De Mey, M., De Maeseneire, S.L., **Van Bogaert**, **I.N.A.,** Beauprez, J., Soetaert, W.& Vandamme, E.J. (2007).Development of a selection system for the detection of L-ribose isomerase expressing mutants of *Escherichia coli*. *Applied Microbiology and Biotechnology* **76**:1051-1057.

(I.F. 2.441)

De Maeseneire, S.L., **Van Bogaert, I.N.A.,** Dauvrin, T., Soetaert, W.K. & Vandamme, E.J. (2007). Rapid isolation of fungal genomic DNA suitable for long distance PCR. *Biotechnology Letters* **29**:1845-1855.

(I.F. 1.134)

De Maeseneire, S.L., Dauvrin, T., Jonniaux, J-L.**Van Bogaert, I.N.A.,** Soetaert, W.K. & Vandamme, E.J. (2008).Cloning and characterisation of the glyceraldehyde-3-phosphate dehydrogenase gene and the use of its promoter for expression in *Myrothecium gramineum*, a novel expression host. *FEMS Microbiology Letters* **25:** 273-278.

(I.F. 2.068)

### *4.2 Chapters in Books*

**Van Bogaert, I.N.A**., De Maeseneire, S.L., Vandamme, E.J. (2008). Extracellular polysaccharides produced by yeasts and yeast-like fungi. In Diversity and potential biotechnological applications of yeasts (Kunze, G. & Satyanarayana, T.) Elsevier: *in press.*

### *4.3 A2 publications*

**Van Bogaert, I.N.A.,** De Maeseneire, S.L., Vandamme, E.J. & Soetaert, W. (2005). Cloning and partial sequencing of the NADPH-cytochrome P450 reductase gene of *Candida bombicola*. *Yeast* **22**: S41.

### *4.4 Posters and abstracts for conferences*

**Van Bogaert, I.N.A.,** De Maeseneire, S.L., Vandamme, E.J. & Soetaert, W. (2005). Cloning and partial sequencing of the NADPH-cytochrome P450 reductase gene of *Candida bombicola*. XXII$^{nd}$ International Conference on Yeast Genetics and Molecular Biology (August 7-12, 2005). Bratislava, Slovak Republic. Poster and abstract, Yeast 22: S41.

**Van Bogaert, I.N.A.,** De Maeseneire, S.L., Vandamme, E.J. & Soetaert, W. (2005). Cloning and characterization of the NADPH-cytochrome P450 reductase gene of *Candida bombicola*. Renewable Resources and Biorefeneries (September 19-21, 2005). Ghent, Belgium. Poster and abstract, Renewable Resources and Biorefineries Conference – 2005: 40.

**Van Bogaert, I.N.A.** & Vandamme, E.J. (2006) Cloning and characterisation of cytochrome P450 genes of *Candida bombicola*. 2006 Society for Industrial Microbiology Annual Meeting (July 30-August 3, 2006). Baltimore, USA. Poster and abstract, SIM Annual Meeting Program & Abstracts 2006: 154.

**Van Bogaert, I.N.A.** & Vandamme, E.J. (2006) Cloning and characterisation of cytochrome P450 genes of *Candida bombicola*. Greenchem symposium on Efficient biocatalytic routes for selective introduction of oxygen (November 9, 2006), Malmö, Sweden. Poster and abstract, Greenchem symposium on Efficient biocatalytic routes for selective introduction of oxygen – Program and abstracts – 2006: 12.

**Van Bogaert, I.N.A.**, Develter, D. & Vandamme, E.J. (2007). Genetic engineering of the sophorolipid producing yeast *Candida bombicola*. Third International Conference on Renewable Resources and Biorefeneries (June 4-6, 2007) Ghent, Belgium. Poster and abstract, Book of abstracts: 90-91.

**Van Bogaert, I.N.A.**, Develter, D. & Vandamme, E.J. (2007). The P450 population of *Candida bombicola*: novel members of the CYP52 family. 15$^{th}$ International Conference on Cytochromes P450 (June 17-21, 2007) Bled, Slovenia. Poster and abstract, Book of Abstracts: 203.

**Van Bogaert, I.N.A.**, Develter, D. & Vandamme, E.J. (2007). Genetic engineering of the sophorolipid producing yeast *Candida bombicola*. 4$^{th}$ International Symposium Cosm'ing 2007. (June 27-29, 2007) Saint-Malo, France. Poster and abstract, Book of Abstracts: 208.

Vandamme, E.J., De Maeseneire, S.L. & **Van Bogaert, I.N.A.** (2007). Exopolysaccharides from yeasts and (yeast-like) fungi. Asian Mycology Congress - AMC 2007. (December 3-6 2007) Penang, Malaysia. Poster and abstract, Book of Abstracts: 56.

## 5 PARTICIPATION AT CONFERENCES

XXII$^{nd}$ International Conference on Yeast Genetics and Molecular Biology (August 7-12, 2005). Bratislava, Slovak Republic. Poster and abstract.

Renewable Resources and Biorefeneries (September 19-21, 2005). Ghent, Belgium. Poster and abstract.

Greenchem symposium on Efficient biocatalytic routes for selective introduction of oxygen (November 9, 2006), Malmö, Sweden. Poster and abstract.

Third International Conference on Renewable Resources and Biorefeneries (June 4-6, 2007) Ghent Belgium. Poster and abstract.

15$^{th}$ International Conference on Cytochromes P450 (June 17-21, 2007) Bled Slovenia. Poster and abstract.

## 6 REVIEWER OF A1 PUBLICATIONS

- Journal of Industrial Microbiology and Biotechnology – September 2007
- FEMS Yeast Research – October 2007
- Journal of Industrial Microbiology and Biotechnology – November 2007
- Current Microbiology – January 2008
- Process Biochemistry – February 2008
- Journal of Basic Microbiology – April 2008